Assessment Report on Consistency Validation of the CDRs/ECVs

Deliverable De3.33

Issue 0.1

Due date of deliverable: 01 December 2014
Actual submission date: 01 December 2014

CORE-CLIMAX

Coordinating Earth observation data validation for RE-analysis for CLIMAte ServiceS

FP7 Grant agreement No 313085

Document History

Date	Update	Contributor	Organization
01/08/2014	Preliminary Structure Drafted	Y. Zeng, Z. Su, W. Timmermans	ITC
17/09/2014	Cross Cutting Validation	J.-C. Calvet	MTF
05/10/2014	Reanalysis help identify in-situ bugs	A. Kaiser-Weiss, F. Kaspar	DWD
10/10/2014	Albedo consistency validation input	T. Manninen, A. Riihelä	FMI
19/10/2014	Limitation of in-situ gridded field	F. Kaspar, A. Kaiser-Weiss	DWD
24/10/2014	Reanalysis help identify in-situ bugs with histograms	A. Kaiser-Weiss, F. Kaspar, V. Heene	DWD
25/10/2014	First draft	Y. Zeng, Z. Su, W. Timmermans, J. Timmermans	ITC
14/11/2014	Comments	T. Manninen, A. Riihelä	FMI
16/11/2014	Histogram Figure	A. Kaiser-Weiss, F. Kaspar, V. Heene	DWD
1/12/2014	Comments	Z. Su, Y. Zeng	ITC

Document Release Sheet

Book captain:	Y. Zeng (ITC)	Sign Date
Approval (lead contractor for WP3):	Bob Su (ITC)	Sign Date
Endorsement:	Bob Su, Co-ordinator (ITC)	Sign Date

Dissemination level	
Public	**x**
Restricted to a group (including the Commission Services) specified by the consortium	
Confidential, only for members of the consortium and for the Commission Services	

Table of Contents

1. Introduction

1.1. Rationalization

The climate data record is a time series of measurements of sufficient length, consistency and continuity to determine climate variability and change. The generation of ECVs (Essential Climate Variables)/ CDRs (Climate Data Records) needs to put strong emphasis on the generation of fully described, error-characterized and consistent satellite-based ECV products. For example, generation of many ECVs (such as in the ESA CCI projects) requires ancillary information about the state of the atmosphere (e.g. cloud screening for SST, atmospheric correction for space-borne altimeters). As such, the consistency between the various ECV products (e.g. cloud flags in one ECV and non-flag in another one) extends to ensuring consistency in the approaches of CDR generation. The in-situ datasets also need to be continuously characterized in terms of their long-term accuracy, stability and homogeneity. Reanalysis results, as an alternative source of ECV, requires similar endeavors to investigate its consistency.

1.2. Scope of this document

In this report, the current practices on consistency validation will be analyzed, based on the consistency validation requirements, the validation capacities, and the current practice examples. The essentials of consistency validation will be summarized. Based on the essentials, the generic strategy for consistency validation will be proposed and discussed.

1.3. Who are addressed

At the European level there are major initiatives, which provide and intend to expand the provision of ECV products and services from satellites, including EU FP7 Climate projects (e.g. ERA-Clim, EURO4M, MONARCH), relevant Copernicus projects (e.g. MACC-II, MyOcean-II, Global Land Service), ESA CCI projects, EEA led Projects (e.g. "Climate Change Impact and Vulnerability Indicator Report" & "Adaptation Report"), the EUMETSAT Central Facility and distributed Satellite Application Facilities such as that for Climate Monitoring (CM-SAF), and in-situ climate data records projects (e.g. E-OBS (KNMI), GPCC (DWD), HADSST (UKMO)). At the international level and national level, there are major coordination mechanisms and CDR programs existing, such as: Coordination Group for Meteorological Satellites (CGMS), CEOS Working Group on Climate, CEOS Working Group on Calibration/Validation (WGCV), CEOS WGCV on land products validation (CEOS WGCV-LPV), the WMO GSICS, SCOPE-CM programs, and NOAA CDR Programme. Except for the above mentioned international, national and European initiatives/services, stakeholders including WMO, GCOS, DG Climate and CEOS Working Group on Climate should be addressed as well.

2. Consistency Validation Requirements and Capacities

2.1. Consistency Validation Requirements

2.1.1. In-situ products

For in-situ products, the scope of quality control may include [*WMO* 2013]: a) data validation; b) Data cleaning or remedial actions and c) quality control monitoring. How all these aspects are in consistency shall be checked to ensure the climate quality of in-situ products.

Based on the in-situ network observations, the gridded datasets can be produced with the internationally accepted estimation methods, which includes: a) mathematical estimation methods (e.g. inverse distance weighting, spline functions); b) estimation based on physical relationships (e.g. regression, discriminant and principal component analysis); c) spatial estimation methods (e.g. Kriging). The combination of different methods (e.g. the use of a regression model and interpolation of residuals) is a common practice. Often, the production of gridded datasets follows a step-wise approach incorporating different estimation procedures.

The estimation method for producing gridded datasets may contain errors. This is because the spatial interpolation assumed that the climatological patterns between widely spaced stations are known and can be modeled, while in reality many factors (e.g. topography, local peculiarities or the existence of water bodies) influence the climate of a region. It is therefore essential to validate the gridded datasets to estimate such errors. The validation of gridded datasets may include: a) split validation (testing the methodology using a smaller subset excluded in the estimation procedure); and b) cross-validation (repeated removals of observations from the sample and analysis of residuals between observed and estimated values).

2.1.2. Satellite Products

The major challenge for climate observation is to have a consistent architecture for observations that is independent of a climate variable's origin and observing method and principles. This requires that each key climate variable shall be measured using independent observations and examined with independent analysis. This highlights the continuing importance of single-source long-term climate datasets for climate variability and trend analysis, the uncertainty of which shall be quantified by (inter) comparing with other independent datasets.

The independent analysis is to verify algorithms that are used for generating climate data (e.g. intercomparison between different retrieval algorithms). It is especially crucial for satellite-based observation data where analysis systems may involve different sets of combination between algorithm theoretical basis documents (ATBDs) and instruments. For example, the independent product can be generated with 3 scenarios by: 1) using the same instrument with different sets of ATBDs; 2) using the same ATBDs on different instruments; or 3) using different ATBDs and different instruments. It is therefore important to verify the accuracy and stability of various outputs for specific variables by thorough inter-comparison, providing insights into errors and help product users to be aware of product differences.

In addition to single-source independent datasets, integrated products are needed as well to ensure reliable conclusions on the detected variability and trends. The integrated data is produced by blending data from different sources, or by integrating observations of variables related to one another (e.g. using data from different instruments or using other data or products to warn of potential issues[*GCOS* 2010b]). For example, the estimation of soil moisture from microwave emission can benefit from analyses of precipitation (e.g. either from in-situ or satellite observation). This requires endeavor to check the physical consistency among different physical variables, apart from the inter-comparison among different independent integrated products.

Although the satellite observations plays a vital role in monitoring global climate, to contribute fully and effectively to the detection of climate variability and change (thus long-term consistent climate records), the satellite observing system shall be implemented and operated in a manner to ensure that these data are sufficiently homogeneous, stable and accurate for climate purposes. To address these technical and resource challenges, the GCOS Climate Monitoring Principles (GCMPs) were proposed to and extended to assist space agencies in addressing the key operational issues: 1) Continuity, consistency and overlap; 2)Orbit stability; 3)Sensor calibration; 4)Data interpretation, sustained data products and archiving.

To follow the GCMPs and be able to address the consistency requirement, there are few key issues need to be taken into account: 1) GAPS in FCDRs must be avoided; 2) Different instruments should be well (inter) calibrated; 3) The generation of long-term ECV products shall be sustained including regular reprocessing; and 4) The optimum use of satellite data (e.g. integrated with in-situ data and model results) requires the organization of data service systems that ensure an on-going accessibility to the data into the future.

2.1.3. Reanalysis Products

Within the current WMO Global Framework for Climate Services (GFCS), reanalysis contributes to both components of "observation and Monitoring" and "Research, Modelling and Applications". The reanalysis can be referred to "reanalysis products" or "reanalysis process".

The reanalysis products (datasets) contain possibly the gridded fields of physical variables from NWP model (e.g. including land surface and sea-state components), ocean model (e.g. including dynamic sea ice and biogeochemistry components), or atmospheric composition model (e.g. GEMS and MACC). The reanalysis products can be considered as a scientific and numerical blending of model data and observational data.

The reanalysis process refers to the activities in integrating an invariant, modern version of a data assimilation system and numerical weather prediction model, over a long time period, by assimilating a selection of observations. The reanalysis process shall also include evaluation, monitoring and quality control of reanalysis products and of observations.

It is well acknowledged that the reanalysis products are consistent with the meteorological parameters, in the sense that they are constrained by the coupled model. On the other hand, in general, there are still significant disconnections between the various Earth system elements in the assimilation, although the models of the various elements can be as far as coupled or fully integrated.

For example, analysis updates from observations (so-called increments) are typically computed separately for each Earth system element's data assimilation, and it is only during the model integration that all states are made physically consistent between one another. This point indicates that such reanalysis results is produced by a weakly coupled data assimilation scheme, because it uses the coupled model only for generating background estimates for each analysis cycle while the analysis itself is uncoupled. This has consequences for the consistency validation of reanalysis production, which entails considering not only validation of the overall system/process but also the validation of the individual ECV datasets. For the consistency validation of ECV datasets, the inter-comparison of reanalysis products or with other independent datasets will be sufficient. As for the validation of reanalysis systems, it requires information for the components listed as below [*Zeng et al.* 2014]: 1) the observations input; 2) the forcing or boundary datasets input; 3) the model configuration (for the various Earth system elements); 4) the data assimilation system (in the various Earth system elements).

2.2. Consistency Validation Capacities

The vision of GCOS is to enable all users having access to the climate observations, data records and information which they require to address climate variability and change. GCOS strives to provide sustainably reliable physical, chemical and biological observations and data records for the total climate system – across the atmospheric, oceanic and terrestrial domains, including hydrological and carbon cycles and the cryosphere. To achieve this vision, both the in-situ and satellite observing systems are indispensable components for GCOS, to provide observations over the breadth of environments from ocean bottom to the upper atmosphere. The existing in-situ and satellite observing systems represent the validation capacities, which will enable the inter-comparison among independent observations and the independent analysis.

2.2.1. *In Situ*

GCOS has coordinated three types of in-situ networks for different purposes:

- To produce stable long-term series and for calibration/validation purpose, the Global Reference observing networks were built/coordinated to provide highly-detailed and accurate observations at a few locations. This includes the most advanced of the Reference networks, the GCOS Reference Upper Air Network (GRUAN). It is to note that GRUAN is not a set of identical stations, but all stations will make a core set of first priority observations. There are guidelines for setting up sites, characterizing instrument error, data quality control, and manual for the data management [*Seidel et al.* 2009].
- The Global Baseline observing networks were built/coordinated to provide long-term high-quality data records of key global climate variables and enable calibration for the comprehensive and designated networks. It involves a limited number of selected locations that are globally distributed. For example, the GCOS Surface Network (GSN), the GCOS Upper Air Network (GUAN), the WMO Global Atmosphere Watch (GAW) and the Baseline Surface Radiation Network (BSRN) [*GCOS* 2010b; *Houghton et al.* 2012].
- The Comprehensive Observing networks were built/coordinated to provide observations at the detailed space and time scales required to fully capture the nature, variability and change of a specific climate variable. It includes regional and national networks, for example, the GCOS-affiliated WMO GAW global Atmospheric CO_2 and CH_4 Monitoring Networks. These networks are operated primarily for non-climate monitoring but also provide important observations for climate purposes [*GCOS* 2010b; *Houghton et al.* 2012].

The above mentioned networks are used to produce reference data for different purposes, the detailed procedures/guidelines/manuals are documented to make such selection of networks as transparent as possible [*GCOS* 2003; *GCOS* 2010a]. It serves for documenting an traceable validation process.

Table 1 Overview of all GCOS-Relevant network components and systems

Atmosphere			**Oceans**	**Terrestrial**
Surface	**Upper-air**	**Composition**		
GCOS Surface Network (GSN)	GCOS Reference Upper-air Network (GRUAN)	GCOS-affiliated WMO/GAW Global Atmospheric N2O, CO2 and CH4 monitoring networks	Global Surface drifting buoy array on 5*5 degree resolution	GCOS/GTOS Baseline Global Terrestrial Network Rivers (GTN-R)
Baseline Surface Radiation Network (BSRN)	GCOS Upper-air network (GUAN)	WMO/GAW GCOS Global Baseline Total Ozone Network	Global tropical moored buoy network	GCOS/GTOS Baseline Global Lake Network
Full WWW & GSN	Full WWW & GUAN	WMO/GAW GCOS Global Baseline Profile Ozone Network	Voluntary Observing Ships	WWW/GOS Synoptic Network
Global tropical moored buoy network	Aircraft (ASDAR etc.)	WMO/GAW Aerosol Network	Global Reference Mooring Network	GCOS/GTOS Baseline Global Terrestrial Network-Glaciers (GTN-G)
Voluntary Observing Ships	Profiler (radar) network		GLOSS Core Sea-Level Network	GCOS/GTOS Baseline Global Terrestrial Network-Permafrost (GTN-P)
Global reference mooring network	Ground-based GPS receiver network		Argo Network	Global Terrestrial Network Hydrology (GTN-H)
GLOSS core sea-level network				
Argo network				

The difference among these three types of in-situ networks can be demonstrated by using GUAN and GRUAN. The GUAN is designed to provide evenly distributed radiosonde network over the globe, measuring temperature, pressure (geopotential height), wind, and humidity (at least to the troposphere) at least 25 days each month, although the target is the collection of twice daily radiosonde observations at the 0000 and 1200 UTC synoptic hours. The lowered observation requirement can be attributed to: 1) instrumentations used in GUAN varies from country to country; 2) data collection rate is better in some areas than others; 3) lapse/inability in the acquisition of replacement radiosondes at some locations may lead to temporal gaps in the climate record; 4) lack of sustaining resources to support existing networks, which leads to the change of instrumentation and the change in location, which will affect homogeneity of GUAN stations. It is clear that these shortfalls of GUAN can also cause inconsistency in the climate record. It is therefore needed for upper air observations meeting reference standards, which led to the establishment of GRUAN in 2004 [*Lawrimore* 2014].

GURAN network served as reference standards for the larger GUAN network, aiming to quantify and reduce measurement uncertainties and to make measurements in a stable way over multi-decadal time scales to achieve data homogeneity in time and spatially between measurement sites. In addition, GRUAN will provide a traceable reference standard for global satellite-based measurements of atmospheric essential climate variables and will ensure that potential gaps in satellite measurements programs do not invalidate the long-term climate record. However, it is to note that GRUAN network has its own limitations to measure all parameters with minimum systematic error [*Lawrimore* 2014].

Thematically, the in-situ network support different domains of ECVs, which includes 5 categories: atmosphere surface, atmosphere upper-air, atmosphere composition, oceans, and terrestrial. It is noticed that except for Terrestrial and Atmosphere-Composition categories, the observing network in the rest of domains contains both routine/baseline networks and reference networks, which indicate the relatively high capacity of the in-situ observing system for doing consistency validation. On the other hand, this indicates the relatively low capacity of the in-situ observing networks of Terrestrial and Atmosphere-Composition domains, to do consistency validation for the domain-relevant ECVs (See Table 1), which requires the reference in-situ observations.

2.2.2. Satellite

There are 50 GCOS Essential Climate Variables (ECVs) (2010) required to support the work of the UNFCCC and the IPCC. Table 2 indicates 29 of the 50 ECVs can be observed from space (noticed that the observing principles for carbon dioxide, methane and other GHGs are similar and considered as one ECV product). The space-observable ECVs are in **bold** and those cannot be measured are <u>underlined</u>. Please find appendix A the overview of products for different domains that can be monitored from space [*GCOS* 2011].

To meet the requirements indicated in 2.1.1, space agencies working through the CEOS and the Coordination Group on Meteorological Satellites (CGMS) have established mechanisms to ensure coordination in agencies' operation and exploitation, for the optimum use of satellite data, by establishing Virtual Constellations on Atmospheric Composition, Precipitation, Land Surface Imaging, Ocean Surface Topography, Ocean Colour Radiometry, and Ocean Surface Vector Winds. The similar coordination mechanism was taken to avoid gaps in FCDRs, which due to missing observations or instrument changes can introduce errors in trend analyses.

The potential future gaps in the satellite ECVs have been conducted through WMO and CEOS, which has been regarded necessary to be routinely updated and acted upon [*Wilson et al.* 2010; *CEOS* 2012]. Developed by WMO in support of Earth Observation applications, studies and global coordination, OSCAR provides the status and the planning of global observing system as well as instrument specifications at platform level (http://www.wmo-sat.info/oscar/).

The inter-comparison of sensors and the (inter)calibration of instruments between satellites has received more and more attentions, which leads to the development of the Global Space-based Inter-calibration System (GSICS) jointly between the WMO Space programme, CGMS and the CEOS Working Group on Calibration and Validation (WGCV). GSICS is to ensure the generation of well-calibrated FCDRs. To focus on the sustained generation of long-term ECV products, the SCOPE-CM (The sustained coordinated processing of environmental satellite data for climate monitoring) initiative has been established with contribution from various space agencies

Table 2 Overview of ECVs capable of being monitored from space

Domain	GCOS Essential Climate Variables
Atmospheric (over land, sea and ice)	**Surface:**[1]Air temperature**, Wind speed and direction, Water vapour,** Pressure**, Precipitation, Surface radiation budget.** **Upper-air:**[2]**Temperature, Wind speed and direction, Water vapour, Cloud properties, Earth radiation budget (including solar irradiance).** **Composition: Carbon dioxide, Methane, and other long-lived greenhouse gases**[3]**, Ozone and Aerosol, supported by their precursors**[4].
Oceanic	**Surface:**[5] **Sea-surface temperature, Sea-surface salinity, Sea level, Sea state, Sea ice**, Surface current, **Ocean colour**, Carbon dioxide partial pressure, Ocean acidity, Phytoplankton. **Sub-surface:** Temperature, Salinity, Current, Nutrients, Carbon dioxide partial pressure, Ocean acidity, Oxygen, Tracers.
Terrestrial	River discharge**,** Water use**,** Groundwater**, Lakes, Snow cover, Glaciers and ice caps, Ice sheets,** Permafrost**, Albedo, Land cover (including vegetation type), Fraction of absorbed photosynthetically active radiation (FAPAR), Leaf area index (LAI), Above-ground biomass,** Soil carbon**, Fire disturbance, Soil moisture.**

[1] Including measurements at standardized, but globally varying heights in close proximity to the surface.

[2] Up to the stratopause.

[3] Including nitrous oxide (N_2O), chlorofluorocarbons (CFCs), hydrochlorofluorocarbons (HCFCs), hydrofluorocarbons (HFCs), sulphur hexafluoride (SF_6), and perfluorocarbons (PFCs).

[4] In particular nitrogen dioxide (NO_2), sulphur dioxide (SO_2), formaldehyde (HCHO) and carbon monoxide (CO).

[5] Including measurements within the surface mixed layer, usually within the upper 15m.

The above indicates the current efforts to address the requirements for generating ECV climate data record from space, and more or less represent the current capacity of satellite remote sensing in contributing to GCOS ECVs. There are specific gap analysis implemented by JRC [*Wilson et al.* 2010] and UKEOF [*Parker and Stott* 2013], to identify how better to coordinate among the existing players to the maximum use of current resources for space-based ECVs, for Europe and UK, respectively.

The generation of many ECVs requires ancillary information about the state of the atmosphere and others. The various ECV products may use different sets of ancillary information (e.g. cloud flags in one ECV and non-flag in another one). This will certainly cause inconsistency among different ECV products. On the other hand, for the generation of a certain ECV, it needs different sources of data. For example, for sea surface temperature, there are different sensors observing the temperature of different “surfaces” [*GCOS* 2010b]: the traditional in-situ SST is taking measurement from well underneath the sea surface (e.g. near-surface and mixed-layer “bulk temperatures”); the infrared (IR) radiometer measures the ‘skin’ SST; and, the passive microwave (MW) radiometer measures the ‘sub-skin’ SST. How to harmonize these three different observations of SST is complicated, due to different measurement depths and diurnal thermal stratification.

The similar inconsistency issue may exist in other ECV products. On the other hand, to be able to identify such inconsistency, the satellite observing system for certain ECV and the corresponding in-situ observing networks shall be maintained and collocated temporally and spatially. In the appendix, the current existing “satellite-in-situ” observing pairs are overviewed (Appendix B) [*GCOS* 2010b]. For example: in the atmospheric domain – surface ECVs, the water vapor and air pressure are recorded with in-situ observing networks (e.g. GSN, WWW/GOS Surface synoptic network), but there is no corresponding space-based observation for them; In the oceanic domain, the surface ECVs can be monitored from both the space and the in-situ. However, for the subsurface ECVs, the current existing satellite observation network is still not capable to observe; For terrestrial domain, the soil carbon cannot be directly measured from the satellite, but the in-situ networks exists. On the other hand, for water use, albedo, FAPAR and LAI, there are no dedicated in-situ baseline network, but some contributing in-situ networks exist.

Although the co-existence of in-situ and satellite observing systems for ECVs indicates a higher capacity in implementing consistency validation than the singular existence of either one, we should be aware of the uncertainty in observational records from both sources. From the appendix Table B1-5, we can see there are vast sources of observations, among which the vast majority of historical/modern weather observations were not made explicitly for climate monitoring purposes.

These datasets (both satellite and in-situ) may be affected by the changes in demands on the data, observing practices and technologies. These changes can alter the characteristics of observational records (e.g. change in mean and/or variability). This therefore requires a certain procedure to be followed to process the raw measurements before they can be used to detect climate variability and climate change. In this sense, for satellite observing systems, the mentioned GSICS and SCOPE-CM can be helpful in establishing an unbroken chain for the satellite measurements to be used/produced in a way to meet internationally recognized measurement standards, which consequently means consistent, homogeneous observational records.

2.2.3. Reanalysis

With a sufficiently realistic global circulation model, by assimilating observational data from multiple sources into a dynamically coherent dataset, reanalysis can produce multi-decadal, gridded datasets that estimate a large variety of atmospheric, sea-state, and land surface parameters, including many that are not directly observed. Reanalysis data can help improve the medium-range forecasts, by using them to assess the performance of operational forecast system and to evaluate the effect of new model developments and other changes [*Simmons and Hollingsworth* 2002]. This is actually corresponding to a strong feedback loop between improvements in the global observing system, advances in data assimilation methodology, and development of better forecast models through analysis [*Dee et al.* 2013].

The currently well recognized need for the development of reanalysis is a more explicit representation of atmosphere-ocean interaction, which will improve surface fluxes of heat and momentum, tropical precipitation, and surface wave fields. The current reanalysis estimate of these parameters suffers from bias and drifts, due to the lack of model feedback between ocean and atmosphere and therefore the poor representation of processes that govern the near surface temperature and wind conditions [*Dee et al.* 2013].

A similar concern holds for the representation for land surface. For example, in the current configuration of ECMWF, the 4D-Var analysis of upper-air prognostic variables is performed separately from simpler analyses of screen-level parameters (temperature, humidity) and land surface parameters (soil moisture, soil temperature, snow depth). Consequently, there is a lack of dynamic feedback in the analysis between the land surface and the atmospheric boundary layer [*Dee et al.* 2013].

For atmospheric composition, the coupled modelling of meteorological, chemical, and aerosol variables and combined use of observations of trace species and meteorology in the 4D-Var analysis is desired. It is because these are the basic elements needed for a fully coupled data assimilation system, in which observations of atmospheric constituents lead to physically consistent adjustments to the meteorological variables, and conversely, meteorological observations can have an immediate impact on estimates of the constituent concentrations [*Dee et al.* 2013]. However, due to the not-yet-adequate observations (e.g. the model background is not well constrained by observations), there are large and unrealistic changes in the upper-stratospheric circulation. As a result, the current practice for atmospheric composition reanalysis does not yet allow direct adjustments to the meteorological parameters based on trace-gas observations, but implement variational bias corrections for most of the constituent observations before being assimilated into the coupled 4D-Var analysis.

From the above, on one hand, there are limits in the current reanalysis in achieving consistency among different physical parameters across domains. On the other hand, the analysis provides complete description of physically plausible atmosphere, ocean and land parameters consistent with instrumental observations. The reanalysis adds value to the instrumental record.

For example, by constraining the model background with the observations, reanalysis produces useful estimates for model variables that are not well observed, such as stratospheric winds, radiative fluxes, root-zone soil moisture, etc. However, due to the absence of direct observations, it is difficult to quantify the uncertainty of those model generated variables, which depend on errors in the model as well as on the observations. Nevertheless, the reanalysis permits the budget diagnostics [*Berrisford et al.* 2011; *Trenberth et al.* 2011; *Lorenz and Kunstmann* 2012; *Mapes and Bacmeister* 2012], which are useful for demonstrating shortcomings as well as progress in climate reanalysis, and examination of the increments can be highly informative about shortcomings in the assimilating model [*Dee et al.* 2013]. It is to note that, depending on the diagnostic, the results can be different due to differences either in the observation data, the assimilation scheme or forecast model, or any combination of these.

It therefore requires uncertainty characterization and consistency validation for reanalysis (e.g. including both products and processes), to understand uncertainty that may come from insufficient observation coverage, insufficient data quality, unknown observation uncertainties or assimilating model deficiencies [*Gregow et al.* 2014b].

There are currently international efforts to tackle this issue through intercomparing reanalysis, for example: the SPARC reanalysis inter-comparison project was proposed in 2012 [*Fujiwara et al.* 2013], and the Reanalysis.org was established to provide researchers with help to obtain, read and analyze reanalysis datasets created by different organizations. The CORE-CLIMAX project has also proposed a procedure for comparing reanalysis, and comparing reanalysis to assimilated observations and CDRs, through work package 5 (inter-comparing reanalysis results). Furthermore, there are recently emerging many reanalysis inter-comparison tools, for example: web-based reanalysis inter-comparison tools (https://reanalyses.org/atmosphere/writ), KNMI Climate Explorer (http://climexp.knmi.nl/start.cgi? someone@somewhere), MERRA Atlas (http://gmao.gsfc.nasa.gov/ref/merra/atlas/), and the Climate Reanalyzer (http://cci-reanalyzer.org/).

3. Current Practice Examples

3.1. Copernicus

3.1.1. GIO Global Land Component (Cross-Cutting Validation)

The direct validation of land CDRs is not easy, as in-situ observations are limited in space and time. Therefore, indirect validation has a key role to play. It consists in comparing the products with similar preexisting products derived from satellite observations or from land surface model (LSM) simulations. State-of-the-art LSMs are able to represent the diurnal cycle of the surface fluxes together with the seasonal, inter-annual and decadal variability of the vegetation biomass. They are able to diagnose a number of land ECVs such as LAI, fAPAR, surface albedo, land surface temperature (LST), and soil moisture.

The most advanced indirect validation technique consists in integrating satellite-derived ECV products into a LSM using a data assimilation scheme. The LSM and the data assimilation scheme (e.g. an Extended Kalman Filter or an Ensemble Kalman filter) are embedded into a modelling platform forming a Land Data Assimilation System (LDAS). The reanalysis provided by the LDAS accounts for the synergies of the various upstream products and provides statistics which can be used to monitor the quality of the assimilated observations.

The current version of the LDAS used for the cross-cutting validation in the Copernicus Global Land Service is operated by MTF and assimilates SPOT-VGT LAI and ASCAT surface soil moisture (SSM) products over France (at 8km x 8km resolution). A passive monitoring of albedo, FAPAR and LST is performed (i.e., the simulated values are compared with the satellite products). The system is used to monitor the quality of upstream products. The LDAS generates statistics whose trends can be analyzed in order to detect possible drifts in the quality of the products: (1) for LAI and SSM, metrics derived from the active monitoring (i.e. assimilation) such as innovations (observations vs. model forecast), residuals (observations vs. analysis), and increments (analysis vs. model forecast) ; (2) for albedo, LST, and FAPAR, metrics derived from the passive monitoring such as the Pearson correlation coefficient, z-score, RMSD, SDD, mean bias.

Figure 1 shows that the analyzed FAPAR from January to May 2014 presents score values similar to those for the 2007-2013 period. In June, the analyzed FAPAR presents a positive bias (i.e. the simulated FAPAR is higher than the observations, by about 0.5 on average) while it is generally unbiased in June. This may denote a problem caused by the transition from SPOT-VGT to PROBA-V, and the FAPAR bias needs to be monitored further in near real time. The similar comparison can be implemented to monitor quality of surface albedo and land surface temperature.

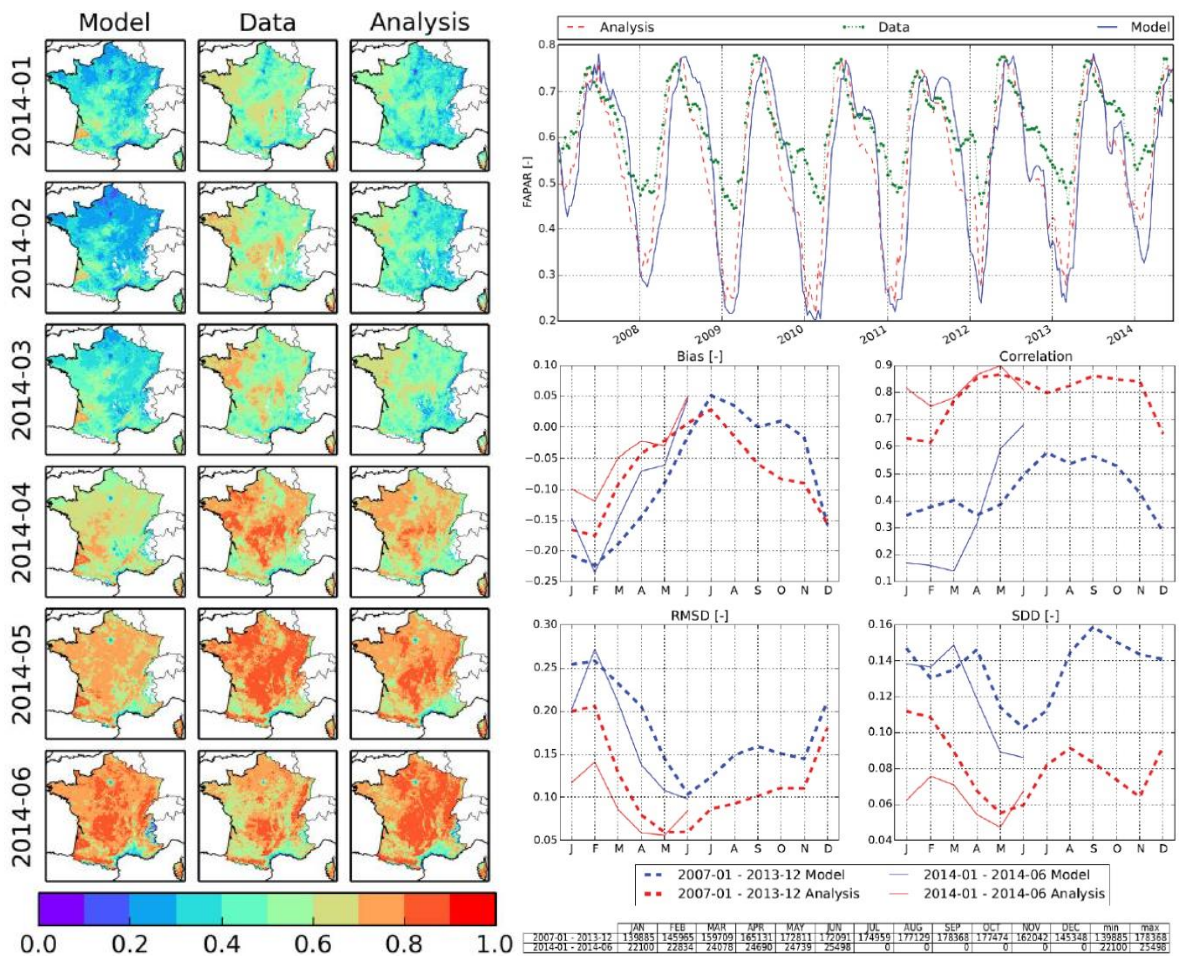

*Figure 1 [Left Panel] Monthly average values of FAPAR over France at 8km*8km spatial resolution from January 2014 (top) to June 2014 (bottom). Model-analysis differences show the impact of assimilating LAI and Surface Soil Moisture on the simulated FAPAR; [Right Panel] the top row shows the monthly average values of FAPAR over France from 1 January 2007 to 30 June 2014, the middle and bottom rows show monthly FAPAR scores of analysis over France at 8km*8km spatial resolution: for thick lines, all data pairs (e.g. model vs. data, analysis vs. data) from 2007to 2013, with N ranging from 139885 in January to 178368 in September, and for fine lines data pairs in 2014, with N ranging from 22100 in January to 25498 in June.*

The importance of the cross-cutting validation, i.e. check the consistency between distinct ECV products (e.g. LAI and surface albedo over land), even across domains (e.g. soil moisture and precipitation) has to be emphasized. The data assimilation systems used to produce reanalysis can be used to monitor the consistency of ECV products of a given domain (e.g. this case study). It is important to have users involved as the impact of the using the products in applications (e.g. river discharge simulations) is a key information. Also, the users are often able to provide relevant in-situ observations. At the same time, generic global models able to ingest the satellite products (e.g. in reanalysis) are needed to assess the consistency of the products at a continental or global scale.

3.1.2. MACC Project

In the Copernicus Atmospheric Service, the different products and services share serveral common characteristics. This is corresponding to the harmonisation, traceability and quality assurance requirements raised by GEOSS, in developing interoperability of systems and the synergistic use of derived information. MACC-I, II & III therefore define the definition and implementation of generalised quality assurance/validation principles (i.e. validation protocols), aplicable to the whole MACC prtfolio. The MACC validation protocol is essentially divided into two main sections addressing respectively generic validaiton principles and specific guidelines for MACC projects. It addresses questions related to data prodcut validation (e.g. rules to ensure unbiased, independent and traceable validation) and to documentation and service operation (e.g. metadata; terminology; service validation; delivery monitoring etc.), allocating a central role to user interaction in these two areas [*Lambert* 2013; *Rudder* 2014].

The overarching principles of the Quality Assurance Framework for Earth Observation (QA4EO) was applied directly to MACC [*QA4EO*], which establishes the data quality strategy for GEOSS as well. With the overarching principles, the quality assessment of a data product in MACC includes:

1. The assessment of uncertainties associated with the way the data product is measured or calculated, including the uncertainties on individual components and where applicable the characterization of the information content;
2. The confrontation with independent reference measurements of documented quality;
3. In the context of services with identified users, quality must be evaluated through critical analysis of the service line against specifications; and
4. of the suitability of data products for the targeted applications.

It highlights responsibilities of research partners in validating algorithm and data product development, of developers and operators in validating service specifications, of VAL sub-project in validating against user requirements and of service provider in implementing quality assessment and quality control. All the validation results wll be feedbacked to the each appointment of responsibility for futher improvement or development. This can be depicted as an end-to-end validation of the service line of MACC –I, II & III [*Lambert* 2013].

The validation of individual service components enable the implementation of quality assessments (or, quality controls) of every component of the production chain, for which validation of only the final data product is not sufficient. This enables the fully traceable information, and avoids that the apparently good behavior of the final data product hides large compensating errors affecting intermediate components of the systems [*Lambert* 2013]. The validation of individual atmospheric components involve confirming that software development output meets its input requirements, various static and dynamic analyses, code and document inspections, and walkthroughs. The validation of service specifications and user requirements focus on service specifications having clear links with user requirements, which were expressed officially through Service Level Agreements (SLAs), WMO Rolling Requirements [*WMO*] and IGOS/IGACO requirements [*IGACO* 2004]. A specific user requirement analysis was implemented by MACC and reported to express user requirements of direct relevance to validation and quality assessment of all services.

Validation of MACC atmospheric data products requires reference measurements, the requirements of which indicated that the reference measurements shall include:

1. Accurate and well documented observations from ground-, aircraft, balloon- and satellite-based system;
2. Equivalent data products in the MACC service portfolio, already validated or produced by an independent method;
3. The output of operational atmospheric models with documentary traceability.

The above requirements indicate the validation strategy for MACC: the observation is the preferred source of validation data, superseding the use of modelling results and climatological values as validation data. However, when suitable observation data are not available, the validation of MACC products may involve comparison with “reference” model data sets. The reference correlative data need to fulfil various requirements (e.g. independence, well calibrated, characterized and documented, representativeness, long term, availability etc.).

As indicated, due to the limitation of in-situ observation, a comprehensive validation of atmospheric satellite observations and model results cannot be achieved always. On the other hand, the inter-comparison with equivalent products and of satellite gridded results with model output provides additional statistical information on the consistency of products. Such equivalent products shall be produced as either interim or final products in more than one processing system, which enable the inter-comparison of equivalent data sets from different sensors and algorithms.

The error budget of a data comparison is highlighted to be implemented, in order to understand uncertainties associated with the selection of data and the methodology of comparison. For example, it may include comparison uncertainties associated with the difference in sampling and smoothing of atmospheric variability and structures [*Lambert* 2013]. Note that the retrieval of geophysical quantities from remote sounding measurements need to use a set of a priori constraints (e.g. climatology), which may mix somehow in the retrieved quantities with the information actually contributed by the measurement. It is therefore also addressed the importance to document the characterization of sensitivity and information content of a measurement, which can help to understand what, in the final product, comes from the climatology, and what comes really from the measurement.

This end-to-end validation of the service line has a very practical application for end users to easily identify problems and make informed decisions. For example, MACC project recently supported

French authorities to understand why there are observations of elevated SO2 values (https://www.gmes-atmosphere.eu/news/ bardarbunga/).

In the middle of September 2014 several European countries were surprised by measurements of high sulphur dioxide (SO2) concentrations at ground level. Nowadays, it is rare to find high concentrations of SO2 in Western Europe -except in specific areas affected by industrial or shipping emissions. It is speculated that these high values could be linked to ship emissions trapped in the lower atmosphere. However, it is unlikely because they were exceptionally high and observations in the United Kingdom and the Netherlands also showed high concentrations between 21 and 25 September.

Thanks to the use of satellite observations to constrain the model forecasts, the MACC near-real-time forecasting system explains the above situation. The SO2 emitted by the Icelandic Bardarbunga volcano was observed by OMI satellite instrument. And, these observations were assimilated by the MACC forecasting system. The subsequent 5-day forecast captured the transport of this plume of volcanic SO2 southward reaching the Channel on 23 September (Figure 2, left). A parallel forecast (Figure 2, right), for which no OMI data were used, further identified the volcanic nature of the plume. No elevated SO2 values were forecasted without assimilating the OMI data, which means that “normal” emissions of SO2 (including shipping and industrial activities) could not explain the observed situation (https://www.gmes-atmosphere.eu/news/ bardarbunga/).

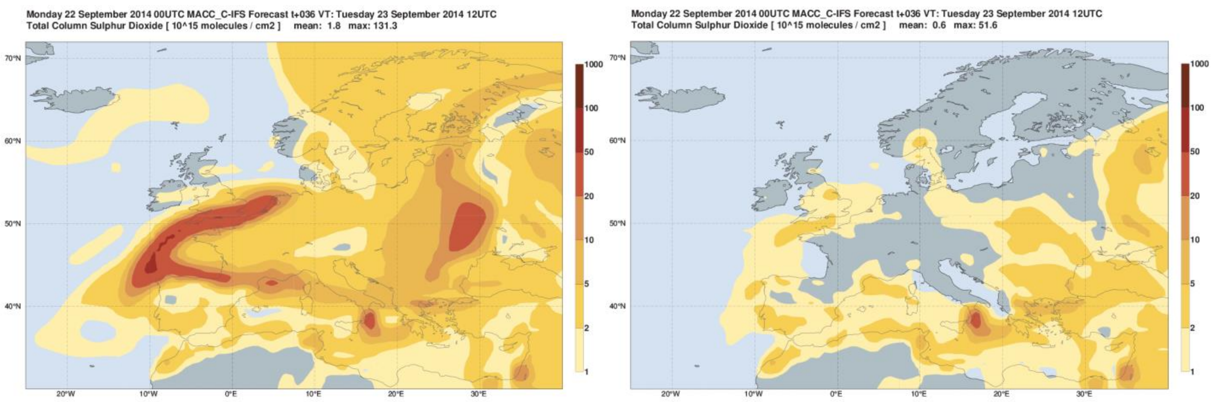

Figure 2. Total Column Sulphur Dioxide: (left)Monday 22 Steptember MACC_C-IFS forecast t+036: Tuesday 23 September 2014, with OMI observation; (right) Monday 22 Steptember MACC_C-IFS forecast t+036: Tuesday 23 September 2014, without OMI observation (From https://www.gmes-atmosphere.eu/ news/bardarbunga/).

3.1.3. MYOCEAN Project

MyOcean project has a dedicated “Product Quality and Calibration/Validation Group”, which produced a large number of diagnostics and proposed complementary methodologies based on the well-defined calibration and validation metrics [*Lellouche et al.* 2013]. All these metrics were synthesized and homogenized, and divided into four main categories, based on Crosnier and Le Provost [2007].

The CLASS 1 metrics facilitate the consistency check between different forecast systems’ solutions or between a forecast system and observations, by implementing ‘eyeball’ verification that consists in comparing subjectively two instantaneous or time mean spatial maps of a given parameter (e.g. the coherent spatial structures or oceanic processes such as main currents, fronts and eddies are evaluated by this CLASS 1 metrics). The metrics for the comparisons of mooring time series and statistics between time series was referred as CLASS2, which facilitate the check on consistency over time. For space and/or time integrated values such as volume and heat transport, heat content and eddy kinetic energy are referred to as CLASS3, where their values are generally compared with literature values or values obtained with climatologies or reanalysis. The CLASS4 metrics are used to measure the real-time accuracy of forecast systems, by calculating various statistics of the differences between all available oceanic observations (in-situ or satellite) and their model equivalent at the time and location of the observation.

The quality check of real-time reanalysis and forecasts is currently being automated with indicators based on distribution (percentiles) thresholds computed from the scientific assessment stage. MyOcean has been publishing the Quarterly Ocean Validation Bulletin "Quo Va Dis?" since July 2010. From the validation information given in such reports, one can find observation minus analysis (called as "residual") and observation minus forecast (called "innovation") statistics for temperature and salinity vertical profiles, sea surface temperature and sea level anomalies observations that are assimilated. The currents at 15m derived from drifting buoys, sea ice concentration and drift, or tide gauges are used as independent validation datasets to do comparison.

Due to the limitations of current monitoring network (e.g. monitoring density is not enough to have a full coverage of the globe), the well-established high resolution reanalysis products (e.g. SST OSTIA [*Donlon et al.* 2012] & GLORYS2V1 [*Ferry et al.* 2012]) are used as "reference". In addition, the integrated parameters such as sea ice extent and global mean SST are monitored, and process studies focusing on one process or region, or short research and development validation studies also complement the bulletins [*Lellouche et al.* 2013].

MyOcean project also implement quality control on in-situ observations and feedback to input data providers (e.g. for temperature and salinity), to minimize the risk of erroneous observations being assimilated in the model [*Lellouche et al.* 2013]. This is known as "background quality control". The basic hypothesis of the data assimilation system is that innovations are normally distributed. Observations for which the innovation is in the tail of the distribution are thus considered to be questionable. The seasonally and spatially variable statistics (mean, standard deviations) from GLORYS2V1 reanalysis (1993-2009) were used to define a space and season dependent threshold value.

For example, an observation is considered to be suspect if the associated innovation is abnormally large than the preset threshold and the associated difference between the observation and the climatology is larger than the half of the innovation. The first condition is a test on the innovation, an abnormally large value of which would be due to an erroneous observation. The second condition avoids rejection "good" observations (i.e. an observation close to the climatology), even if the innovation is high due to the model background being biased [*Lellouche et al.* 2013].

Figure 3 shows an example of an erroneous temperature profile detected by the QC procedures. It is obvious that the innovation of the observed temperature profile is larger than the threshold (Figure

3a left panel), and the difference between the observation and the climatology is much larger than the half of innovation (i.e. the climatology is close to the model already, Figure 3b left panel). And, when this 'bad' temperature observation is assimilated, an abnormal value of salinity can be seen at the temporal and spatial positons of this profile (Figure 3b).

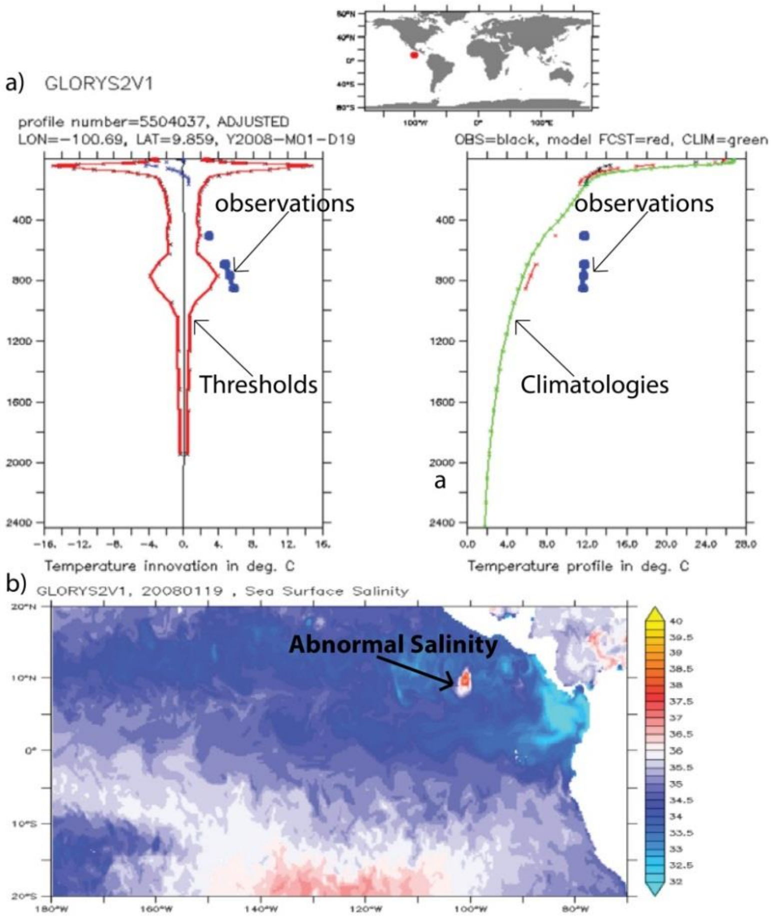

Figure 3 Example of a suspicious temperature vertical profile highlighted by the QC procedure in MyOcean project. (a) left panel shows temperature innovation profile in blue and the innovation threshold in red. Right panel shows the absolute vertical temperature profile, observation in black, climatology in green and model in red. Large blue dots correspond to 'bad' observations. (b) when this profile is assimilated, an abnormal value of salinity appears at the position of this profile (adopted from Figure 5 in [Lellouche et al. 2013]).

3.2. EUMETSAT CM SAF Example (Albedo Consistency Validation)

Because of the high albedo contrast between snow-covered and snow-free land surfaces, consistency checks between surface albedo retrievals and snow cover extent products are, in principle, feasible and relatively straightforward if the snow cover data contains fractional snow cover estimates. In a favourable scenario, the albedo and snow cover retrievals are from different satellite instruments whose measurement uncertainties are well known. In reality, these uncertainties are usually only partially known so that neither dataset can be considered as a "reference" for the other – yet a comparison between the two provides added value for both through helping to identify areas of inconsistency. Here, we will present a showcase consistency check between the CM SAF CLARA-A1-

SAL surface albedo dataset [*Riihelä et al.* 2013] and the ESA GlobSnow snow cover extent dataset [*Metsämäki et al.* 2012].

There are numerous aspects that need to be carefully considered during such a consistency check to avoid misleading analysis results. Firstly, albedo and snow cover extent retrievals from different satellite instruments may have very different spatial resolutions; in our showcase, the albedo dataset (from AVHRR) has a spatial resolution of 0.25 degrees latitude-longitude, whereas the snow cover dataset (from ATSR-2 and AATSR) is at a spatial resolution of 0.01 degrees. Area topography and the density and type of vegetation may have significant impacts on the consistency if the imaging geometry (including variance in sampling times) is significantly different. The SAL product contains a topography correction, the fractional snow cover product not, as it is not designed for mountains areas. Another significant impact factor is the reliability of cloud masking in both datasets, as it is a prerequisite for both snow cover and surface albedo retrieval. Misclassifications of cloudy and snowy pixels will result in large inconsistencies between the datasets, unless both algorithms make similar classification mistakes.

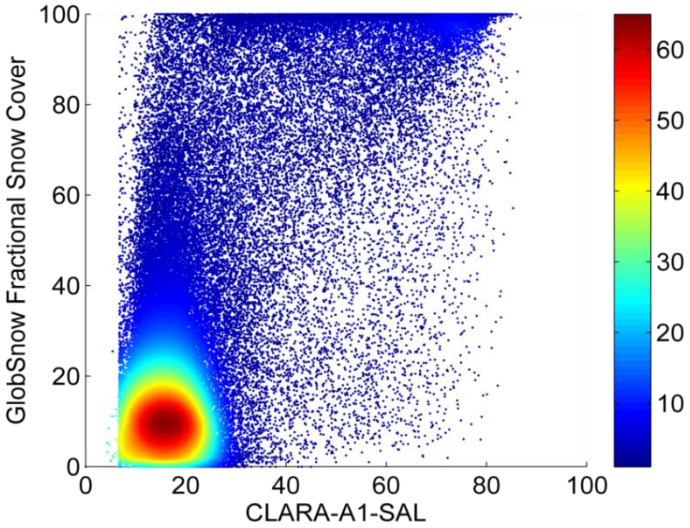

Figure 4: CLARA-A1-SAL versus GlobSnow Fractional Snow Cover for 2009 (April-September). One fifth of all compared data shown.

Example results from the GlobSnow - CLARA-SAL consistency check are shown in Figure 4. The results show that the albedo-snow cover relationship is more complicated than simply a linear correlation. Hence, just having a snow product does not enable the estimation of the albedo. It is notable that while there is considerable scatter in the comparison, low FSC values correspond strongly with low albedo (the comparison is made for all land cover in the Northern Hemisphere). These cases are typically from open areas, where the shrinking snow cover at melting season is replaced by the bare soil or withered ground vegetation, which are darker than the snow. Likewise the case, when both the albedo and the fractional snow cover obtain high values correspond to open areas having full snow cover. In April in the northernmost areas the snow is still in midwinter status having a really high value. In more southern areas the melting has already started, and the albedo of the snow is decreasing, although the snow coverage is still complete. Then, gradually the snow cover starts to be patchy and the dark ground decreases the albedo even more.

In addition, the case of coniferous forests complicates the comparison. Even at full snow coverage, the albedo will not be high. Depending on the leaf area index the broadband albedo can be as low as about 0.2 [*Manninen and Stenberg* 2009]. This is nicely demonstrated also by Figure 4 with the high number of points with 100% snow coverage and albedo in the range 0.2 … 0.85, the latter value being about the maximum albedo of midwinter snow. The few points, for which the albedo is high and the snow cover is small are not realistic. The most probably cause for them could be their location in mountainous areas, where the fractional snow cover estimation is disturbed by pixels shaded either by the sun or the satellite.

The probability of cloudiness masking the highest snow albedo values was tested also in a measurement site in Finland, where SYNOP cloud coverage values are available every three hours. The average cloud fraction at UTC times 6, 9, 12 and 15 was determined from that data during January – March in years 2005 - 2010. As Figure 5 shows, the average cloudiness is high during January – March, when the snow cover is at midwinter status and the albedo is highest. Yet, one has to take into account that only clear sky pixels are used for satellite products. Therefore, also the fraction of cloud coverage classes 0 and 1 (perfectly clear sky and almost clear sky) was also determined. Still, the probability of having a clear sky event at a certain UTC time is about 10%, which would mean 9 images out of 90 possible during January – March. Thus, in Jokioinen at least it is not probably, that the cloudiness would seriously affect the surface albedo retrieval.

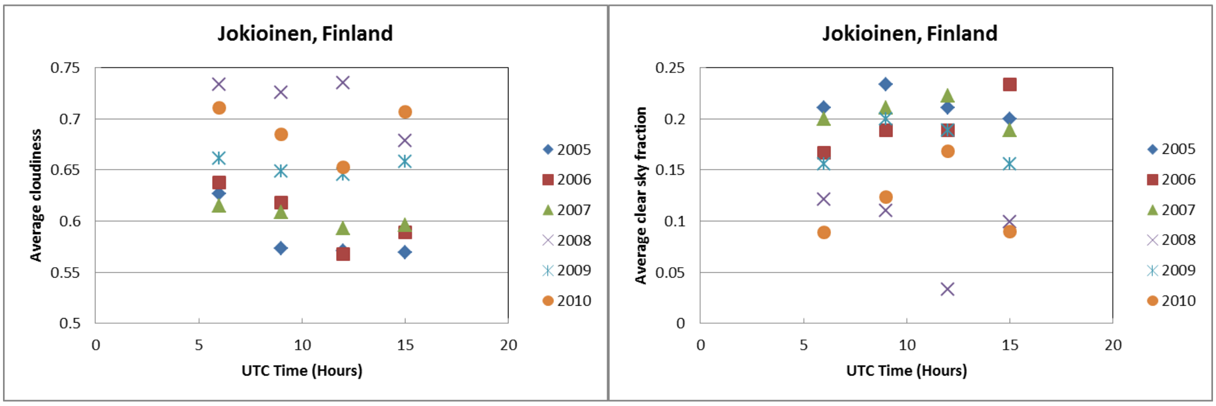

Figure 5. Cloudiness characteristics at Jokioinen weather station at SYNOP UTC times 6, 9, 12 and 15 for years 2005 -2010: left) the average cloudiness and right) the average clear sky fraction (cloud classes 0 and 1).

3.3. ESA CCI (ECV Quality Assessment)

The consistency issues of ECVs are tackled by each CCI science team, and also assessed by Climate Modelling User Group (CMUG). The CMUG, as user of the Climate Data Records (CDRs) delivered by the ESA CCI, provides feedback to CCI project the comments, technical advices on some of the CCI Project Task 4 deliverables, including the “Product Validation Plan”, “Product Validation Protocol”, “Product Validation and inter-comparison Report”, and “climate assessment report”. The comments and technical advices on these validation-relevant deliverables are reported by CMUG as a validation and user assessment report [*Saunders* 2014]. In the report, issues are discussed include: 1) is the validation of the products adequate? 2) are the error characteristics provided by CCI projects adequate? 3) What are the different components of the uncertainty.

The purpose of the validation and user assessment report mentioned above is to prepare for a comprehensive quality assessment on the climate data records (CDRs) delivered by ESA CCI, and understand issues before using them with coupled Earth System Models. In the CMUG ECV quality assessment report, the consistency validation was discussed in details [*CMUG* 2014].

The purpose of the ECV quality assessment is to provide added value for climate modelling activities such as initialization, assimilation, model evaluation and development, trend analysis and monitoring. The CMUG ECV quality assessment of the CCI CDRs will not repeat the validation activities performed by each CCI team, but is rather complimentary to what has been achieved. CMUG provides an independent assessment of the CDRs.

From the perspective of the climate modeling user, there are four key applications of the CCI CDRs: enabling Model-Observation Confrontation, providing boundary conditions, initial conditions and observations capable of assimilation. The model-observation confrontation plays a significant role in the decision process that determines whether a dataset is deemed suitable for the other 3 key applications.

For the model-observation confrontation, the consistency check includes aspects showed as below;

- Consistency of Global Satellite Data Products in time (e.g. stability, uncertainty of bias);
- Consistency with independent observations (e.g. limb view, in-situ, ground-based, remote sensing, air borne etc.);
- Consistency with precursor datasets to understand the differences and assess if the CCI datasets are better representations of the atmospheric/surface state;
- Consistency compared to reanalysis fields;
- Consistency across ECVs
- Ability to capture climate variability and small climate change signals (e.g. observed trends) for their use in climate monitoring and attribution

Accordingly, the data used for assessment of CDR includes: 1) climate model results (Single, Ensemble), 2) reanalysis, 3) precursors datasets, 4) independent satellite or in-situ measurements, and 5) related observations (e.g. surface and Top of Atmosphere fluxes, temperature, water vapor). The various kinds of assessment datasets have their own advantages and disadvantages. Although reanalysis and climate model datasets can provide spatially and temporally complete ECVs, the model and the assimilating system are subjected to various kinds of errors (e.g. model and observation errors), treatment of which comes to the consistency validation of model and assimilating system themselves. Although the independent satellite or in-situ measurement can be used to assess CDRs objectively, they also suffer large uncertainties, erroneous records, representativeness errors, homogeneity issues, which, on the other hand, can be identified by using reanalysis and climate model datasets.

The consistency across ECVs has been drawing attentions increasingly, during the development of CDRs. The climate modelling community approaches consistency from an integrated perspective which includes consistency across ECV product levels, e.g. from Level -1 radiances to Level -2 swath based geophysical products to Level-3 gridded products, and also extends to ancillary data products such as bias corrections and homogenization. The CMUG didn’t tackle this issue in the Phase I, and will explore this point in future studies by the CMUG during phase 2 using the CCI datasets.

3.4. CEOS WGCV (LAI product)

Validation of global LAI products has progressed from producer driven studies to CEOS sponsored efforts to ongoing validation of operational products [*Camacho et al.* 2013]. From the previous practice, there are three fundamental components for a validation protocol for LAI product [*Fernandes et al.* 2014]: 1) direct validation over upscaled in-situ reference datasets; 2) inter-comparison of products over a representative global sample; and 3) statistics related to the temporal completeness of LAI products. On the other hand, most of previous validation studies have performed accuracy assessments using both pooled global reference datasets and comparisons to regional reference datasets. This has several limitations [*Fernandes et al.* 2014]:

1. The sampled in-situ sites cannot be assumed to be representative of the global LAI distribution;
2. LAI products are frequently based on biome specific algorithms that should then be validated at the biome level; and
3. Validation statistics from pooled global reference sites may be biased since differences between reference and products may be systematically related to land cover or climate conditions, while common accuracy statistics usually assume that differences arise from simple (e.g. bivariate normal, unimodal) distributions.

Based on above, the recommended approach for global LAI product validation is proposed as below:

1. The reference datasets shall include or meet following items: a) reference estimates that are traceable to in-situ measurements; b) Heuristic reference estimate (e.g. with the uncertainty for a given land cover class and given land cover map reported); c) Co-location of LAI estimates (e.g. the geolocation uncertainty and binning uncertainty shall be reported);
2. Validation exercises should explicitly define accuracy, precision and completeness applicable to LAI;
3. LAI validation should be performed across a representative sampling of LAI magnitudes within a spatial and temporal stratification, which includes: a) employ a spatial stratification for performance assessments corresponding to continental biomes map; b) employ a temporal stratification like separation of snow free and snow covered conditions; c) sample across a representative range of LAI within a stratification for all performance statistics; and d) evaluate the precision and completeness of spatial and temporal patterns in addition to reporting statistics based on LAI product estimates in a stratum without spatial or temporal consideration;
4. Validation statistics should be reported for each stratum, by using the recommended statistics, including: a) total measurement error and bias; b) precision (e.g. inter-annual and intra-annual precisions); c) completeness (e.g. spatial and temporal completeness); and d) ensemble inter-comparison;
5. The results of validation exercises should be reported publicly after review by the data producers and independent scientific peer review.

In the following, the validation exercises for GEOV1 data product reported by Camacho [2013] will be introduced as a show case. To assess the continuity, consistency and precision as well as the accuracy of the GEOV1 products, **the reference data** are selected from the current existing global products (i.e. for inter-comparison) and in-situ networks (i.e. for direct validation). Although for the production of maps and the computation of the fraction of missing data, the original projection of the products was kept, the spatial sampling was adapted to get similar spatial support across all the products investigated when performing the inter-comparison over site extracts. *A resampling procedure* was implemented to achieve the minimum consistent spatial support (MCSS), to reduce co-registration errors between products and inconsistencies associated to differences in the point spread function of the re-projected products [*Camacho et al.* 2013]. As for the in-situ measurement*, a recommended upscaling procedure* was implemented to achieve spatial matching of the reference LAI and satellite LAI products.

With the MCSS, the fraction and distribution in space and time of the missing data (gaps) can be estimated to assess **the spatial and temporal continuity**. The fraction is mainly due to cloud or snow contamination, poor atmospheric conditions or technical problems during the acquisition of the images. To investigate **the spatial consistency** among different global products, maps of annual mean difference between products and maps of the annual root mean square error were computed over each MCSS for the two year period. To better compare the maps, a common temporal support period (CTSP) was also considered. The LAI, FAPAR and FCOVER variables are expected to change smoothly and seasonally. Therefore, the smoothness and seasonality are taken as the measure of the short time stability, for checking **the temporal consistency**.

For the consistency check mentioned above, only limited spatial domain (e.g. few sites for the temporal consistency) or a temporal period (e.g. few dates for the spatial consistency) will be employed. To have a more comprehensive analysis, **the bulk statistical analysis** was conducted over the whole common time period available (2003-2005) and for a globally representative set of sites. In the bulk statistical analysis, the distribution of products as a function of the biome type and the consistency between GEOV1 and the other products were checked.

The direct validation facilitates the assessment of the accuracy of satellite products using a ground reference data set representative of area meeting MCSS, by an empirical "transfer function" between high spatial resolution radiometric signal and the biophysical measurements established using a representative number of elementary sampling units (ESUs). To improve the consistency of reference ground measurements, several sources of uncertainties (e.g. caused by different instruments, sampling strategy, number of observations, and band combination used in the transfer function etc.) were reported and screened out to provide the best consistency within these reference values.

After all these consistency check, the results suggest that the GEOV1 products have reached stage level of 2 according to CEOS LPV criteria and are ready to be released to the users' community. On the other hand, due to the limitation of the ground data set, the validation over broadleaf evergreen and deciduous forest (e.g. canopy understory was not measured) were not yet implemented comprehensively. It is over such area, where contamination by cloud or snow limits the reliability of the reflectance values used as inputs in the algorithms, the highest discrepancies (systematic and overall) and lowest correlation between products were found [*Camacho et al.* 2013]. It is suggested to expand the current GEOV1 validation exercise by increasing the period of study, and including new multi-temporal ground measurement over new sites around the world.

3.5. In-Situ

In contrast to satellite-based datasets, the regional quality of in-situ based datasets depends on the availability of data provided by national agencies. Data policy differs significantly between European countries. Therefore European and global gridded datasets of ground-based parameters are typically generated based on very inhomogeneous station density. For Europe, a frequently used collection of meteorological in-situ measurements was generated by KNMI within the project 'European Climate Assessment and Data' [*Klok and Klein Tank* 2009]. Its dataset contains series of daily observations of 12 elements at meteorological stations throughout Europe and the Mediterranean (40189 series at 10233 meteorological stations; (ecad.knmi.nl: 17.10.2014)). Part of the dataset is freely available for non-commercial research and education; depending on the national data policies. However, the density of available (free and non-free) stations differs significantly between countries.

This collection is used to generate gridded products for Europe for temperature and precipitation [*Haylock et al.* 2008] and pressure [*van den Besselaar et al.* 2011]: the 'E-Obs datasets'. Gridded products are also produced for smaller regions within national activities, esp. by NMHSs. These are typically based on higher station density and are often produced at higher spatial resolution. A number of studies evaluated the suitability of the E-Obs dataset for specific applications by analysing the impact of station-density or comparison with national datasets that are based on higher station-density [*Hofstra et al.* 2009; *Hofstra et al.* 2010; *Kyselý and Plavcová* 2010; *Maraun et al.* 2012]. They find significant differences in regions with low-station density that restricts the suitability for specific research questions. For example, Maraun et at [2012] conclude based on an analysis for the UK, that the applicability of the E-OBS data set for validation studies on a gridbox level is limited.

All the climate related data, collected from ground stations in Germany over many decades (stretching far beyond last century) had been measured with great care, and were documented with meta-data. However, the measurement practice, the instrumentation, the station location, the total number of stations, the ownership, archives and storage systems, and various forms of quality control changed over time. A reanalysis which is relatively stable over time gives us a chance to compare the whole set of historical measurements against it. It is noted that reanalysis is not the truth, and has biases, drifts. It contains also different kind of representativity of physical parameters other than what is measured from ground stations. There are more efforts needed to be addressed to relate these two worlds (reanalysis/station measurements). On the other hand, the current reanalysis we have now (e.g. ERA-20C, ERA-20CL) is good enough to spot a few serious bugs in the in-situ station system.

For example, it was spotted in the archive of the Nationales Klimadatenzentrum (NKDZ) that in the early eighties there were suspiciously many calms in the 10m winds. If that would be only one station, probably one would have looked for changes in station surrounding (land use, buildings). Puzzlingly, several stations showed the same effect at the same time. Could it be a real phenomenon?

The investigation is implemented to check whether ERA-20C showed the same phenomenon. With the comprehensive atmospheric variables in the reanalysis, it is expected to discover some special effect in wind speeds over Germany for the early eighties, as being seen in the station data. However, it turns out no such calms in the reanalysis products (see Figure. 6). Further checks on the cause of the bug came to the preliminary conclusion (see ftp://ftp-cdc.dwd.de/pub/CDC/Error_log_CDC_ftp.txt): 20140819 ftp://ftp-cdc.dwd.de/pub/CDC/observations_germany/climate/hourly/wind/, and quoted as "In some years, for some stations, there is an unrealistic number of zero winds. Possibly, in earlier years, there was not always properly distinguished between instrument failure and calm. It is advised to inspect thoroughly any historical values of wind = 0 m/ sec before using it."

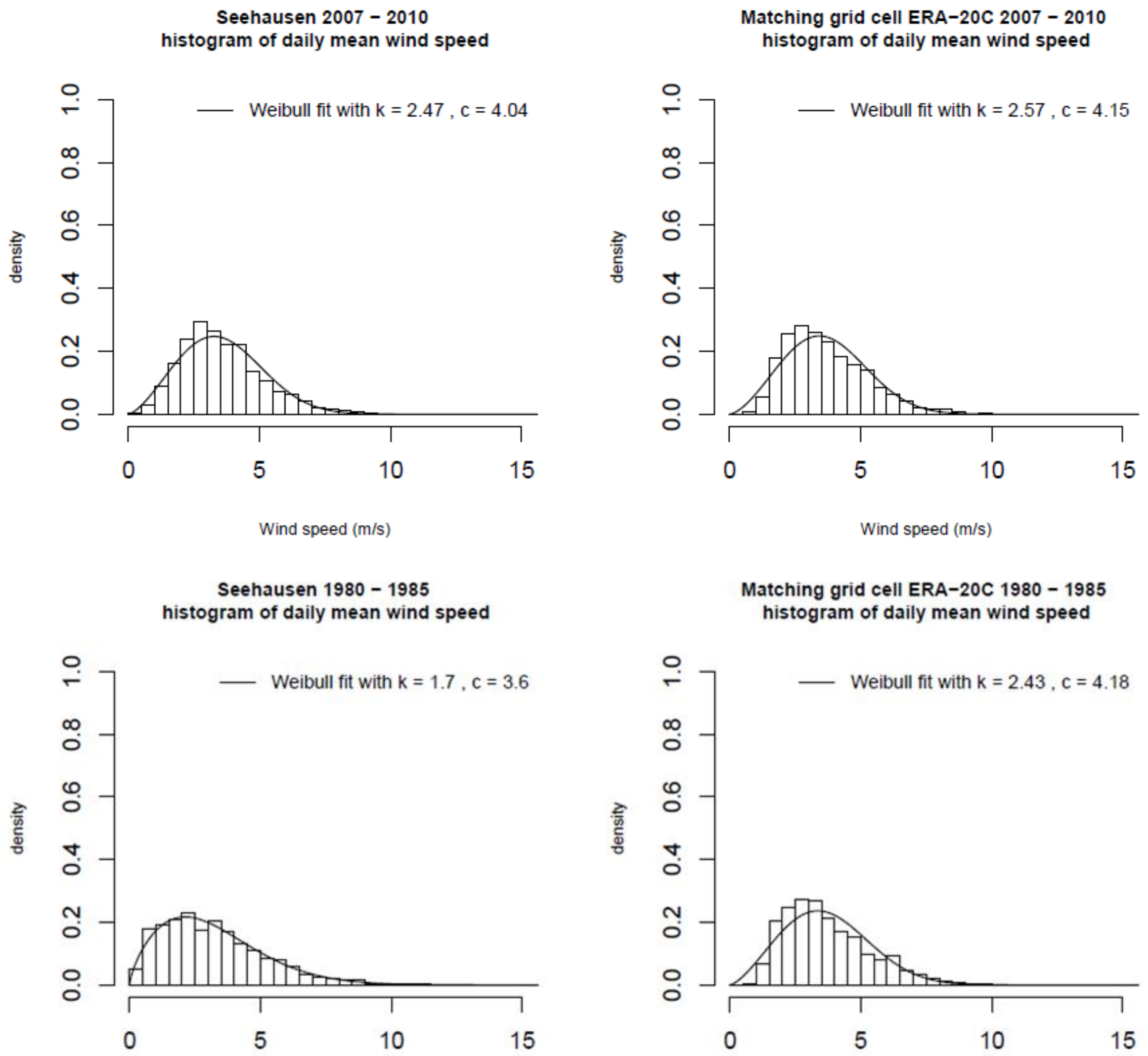

Figure 6: Example for applying reanalysis for historical quality control: observed daily 10m wind speed at the station Seehausen , Germany (left) match well with the frequency distribution from the interpolated ERA 20 C wind fields (right) in recent years (top) whereas 1980-1985 (bottom) the observed station values (bottom left) are found to be suspicious.

The above example shows how to make use of reanalysis as a historical data quality control tool (among other ways of searching for problems in historical data). Although 10 m winds is not the best parameter to use from reanalysis, in this case, it was good enough to spot a bug in the in-situ measurement. This allows to rescue more meta-data, which might still be available in paper archives, but had not made it into the early electronic archives [*Kaspar et al.* 2013].

3.6. Reanalysis

3.6.1. SPARC Reanalysis Inter-comparison Project (S-RIP)

The SPARC community has been using reanalysis datasets to understand atmospheric processes and variability, and to validate chemistry-climate models, and potentially for trend analysis. On the other hand, different reanalysis give different results for the same diagnostic, for example [*Fujiwara et al.* 2013]: global energy budget and hydrological cycle, Brewer-Dobson circulation, stratospheric vortex weakening and intensification events, polar winter lower stratospheric temperature, large-scale wave activity at the tropical tropopause, diurnal tides, temperature trends, climatology of the middle atmosphere. Depending on the diagnostic, the different results may be due to differences either in the observational data assimilated, the assimilation scheme or forecast model, or any combination of these.

The SPARC-Reanalysis Intercomparison Project is therefore aiming to create a communication platform between the SPARC community and reanalysis centres, to understand current reanalysis products, and to contribute to future reanalysis improvements. The project will compare newer reanalysis, i.e., MERRA, ERA-Interim, JRA-25/JCDAS, NCEP-CFSR, and 20CR (the JRA-55, ERA-20C, ERA-SAT, MERRA-2 will be included when available). The inter-comparison period is 1979-2012, i.e., called as “the satellite era”, but some analysis will be implemented over the pre-satellite era before 1978.

In S-RIP’s 4-year workplan and the outline structure of its final report, the inter-comparison will be implemented over different natural process, for example, Chapter 3 to 11 include, respectively: 3: Climatology and Interannual Variability of Dynamical Variables, 4: Climatology and Inter-annual Variability of Ozone and Water Vapor, 5: Brewer-Dobson Circulation, 6: Stratosphere-Troposphere Coupling, 7: Extratropical Upper Troposphere and Lower Stratosphere, 8: Tropical Tropopause Layer, 9: Quasi-Biennial Oscillation and Tropical Variability, 10: Polar Processes, 11: Upper Stratosphere and Lower Mesosphere.

3.6.2. CORE-CLIMAX Reanalysis Inter-Comparison Procedure

The CORE-CLIMAX work package Nr. 5 (WP5) focuses on the inter-comparison of reanalysis, which is the key component of characterizing reanalysis uncertainties. This is playing a paramount important role in assisting users in deciding which reanalysis product to use for their own purposes. The WP5 proposed a set of procedures for comparing reanalysis, and comparing reanalysis to assimilated observations and CDRs.

The users will benefit from all these five categories of reanalysis inter-comparison, according to the analysis of the user questionnaire and literature studies [*Gregow et al.* 2014b]. All these comparisons will help drawing conclusions on the value and on the proper use of the reanalysis products for specific applications. The generic descriptive product comparison will provide the first level of information needed before proceeding further with any other comparison. This category of comparison is simple to conduct and simple to interpret.

Table 3 The five categories of comparisons accompanied with two complexity ratings [*Gregow et al.* 2014a]

Categories of Comparison	**Complexity of Conducting**	**Complexity of Interpreting**
Descriptive Product	Simple	Simple

Comparison		
Comparison with 3rd party observation-based CDRs	Moderate	Moderate
Inter-comparison between different reanalysis	Moderate	Moderate/Difficult
Thematic comparison	Difficult	Difficult
Internal Metrics Comparison	Difficult	Moderate

The comparison with third-party products involve comparison with gridded observation-based CDRs and with in-situ or swath observation-based CDRs, at the collocated times and spaces. For this second category, the interpretation of the results requires more knowledge of how each product was derived (i.e. resolution, representativeness and exact domain area of validity). It is moderate to conduct such comparison and also moderate to interpret the results.

Then it comes to the inter-comparison between different reanalysis products, which includes three kinds of comparisons: global versus global reanalysis, regional versus global reanalysis, and regional versus regional reanalysis. For the global reanalysis inter-comparison, it is noted that good agreement might not necessarily indicate a reduction in uncertainty, but could be related to the sameness of methodology or technical parameters. On the other hand, any disagreement is easier to interpret as a sign to raise alertness that the observations might be imperfect (biased), or the observations do not constrain the model sufficiency, or model errors might play a role. For the second kind, one aspect worth examining concerns the extent to which long-term variability in the regional reanalysis is dictated by features resolved in the global reanalysis, or whether the regional assimilation process modifies this significantly. For the third kind, the most important factor to check is whether they use different global reanalysis as boundary condition. As a whole, this category is rated as moderate to conduct and moderate/difficult to interpret.

The fourth category is thematic comparison, which consists in evaluating how well the reanalysis products can be applied to understand a particular problem. In this sense, the reanalysis product can at least be compared with different sources of datasets by climate service users (e.g. wind energy company, insurance company) or scientific researchers. The climate service user application comparisons can be also called as crowd comparisons. One of the first actions for this group is to compare the reanalysis products with other products already in their possession, to see how well they agree. As for the natural processes representation comparison (e.g. by scientific researchers), many existing such comparisons are essentially bi-lateral in nature, i.e. they compare two datasets. Some are multi-lateral but in a limited sense, in that the number of datasets compared is more than two but does not cover the full range of available datasets. For this category, both kinds of comparisons are difficult to be conducted and interpreted.

The last category is internal metrics comparison. There are three classes of internal metrics: a) Internal metrics based on differences between a prior estimate and new estimate (i.e. systematic analysis increments, as the first measure of temporal discontinuity); b) Internal metrics based on differences between new information (observations) and past information (e.g. from persistence or from a forecast model). These are called as innovations and can be a measure of quality with respect to observations; and c) internal metrics characterizing the error estimates produced by the system.

Such error estimates may include bias correction or adjustments, ensemble spread, random error estimates. This category is difficult to be conducted, but moderate to be interpreted.

3.7. Consistency among Hydrological Cycle Variables

Based on observation data (e.g. both in-situ and satellite), we can quantify the water cycle components in river basins and compare these to the results obtained by using reanalysis data. TWS (terrestrial water storage) are obtained by balancing precipitation, evaporation and river runoff from satellite observations and in-situ observations. The same is also obtained from Interim Reanalysis data (ERA-Interim). Upon comparing these TWS data to the GRACE observations of storage changes, we conclude that a method can be devised to separate the impacts on the water cycle components by climatic and human factors. Demonstration cases are presented for the Yangtze river basin (Figure 7).

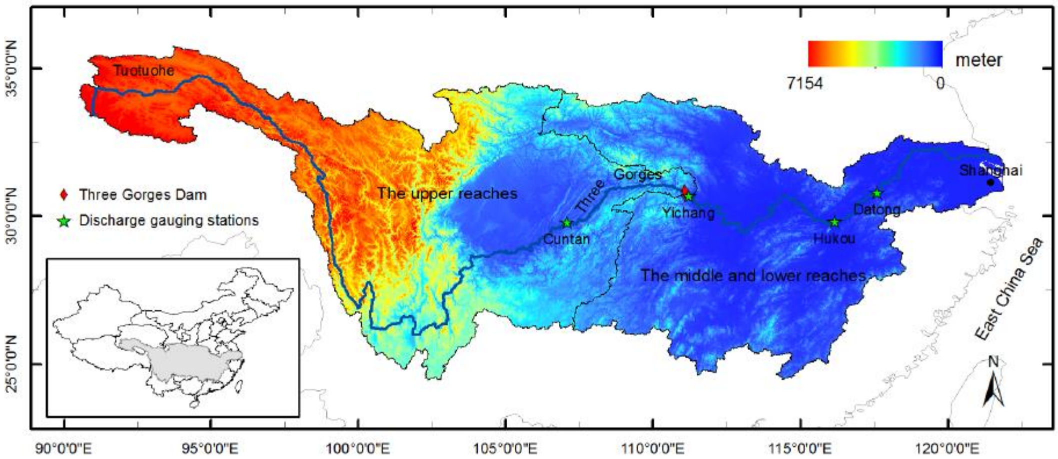

Figure 7 Yangtze river basin (Upper Yangtze reach, from Tuotuohe, to Yichang; Middle reach from Yichang to Hankou; Lower reach from Hankou to the river mouth near Shanghai; Cuntan, Yichang, Hankou, and Datong are four gauging stations located along the mainstream of the Yangtze)

We start with the mass conservation equation for water which takes the form of

$$\frac{\partial S}{\partial t} = P_{GPCP} - E_{SEBS} - R_{Obs} \cdot f(P_{i,j}, E_{i,j}) \qquad (1)$$

Where S is the amount of water stored at surface and subsurface per unit of land surface, P_{GPCP}, the GPCP (Global Precipitation Climatology Project) data is used; E_{SEBS}, the SEBS derived land evapotranspiration is used; R_{Obs}, the in-situ observed river discharges; $f(P_{i,j}, E_{i,j}) = (P_{i,j} - E_{i,j})/(P - E)$, a scaling factor to distribute the observed discharge to each pixel, $P_{i,j}, E_{i,j}$ are GPCP precipitation and SEBS ET for pixel (i, j) and P, E are the mean GPCP precipitation and SEBS ET for the catchment area of interest, all expressed in cm water depth.

The TWS anomaly and cumulative TWS anomaly are estimated from GPCP precipitation, SEBS estimated evapotranspiration (Chen 2014), observed discharge as well as from ERA-interim data which are compared with GRACE TWS for the Upper Yangtze reach and the whole Yangtze river basin. The river discharge measurements from Yichang station for the period of 2001-2010 are used for the

Upper reach study. The discharge measurements from Datong for the period of 2005-2010 are used for the whole Yangtze River basin study. The GPCP precipitation data (precipitation depth cm/month) was obtained at http://jisao.washington.edu/data/gpcp/; GRACE data obtained at: http://grace.jpl.nasa.gov/ (data version: RL05.DSTvSCS1401). The GRACE monthly grid data represents equivalent water thickness deviation to the average over Jan 2004 to DEC 2009).

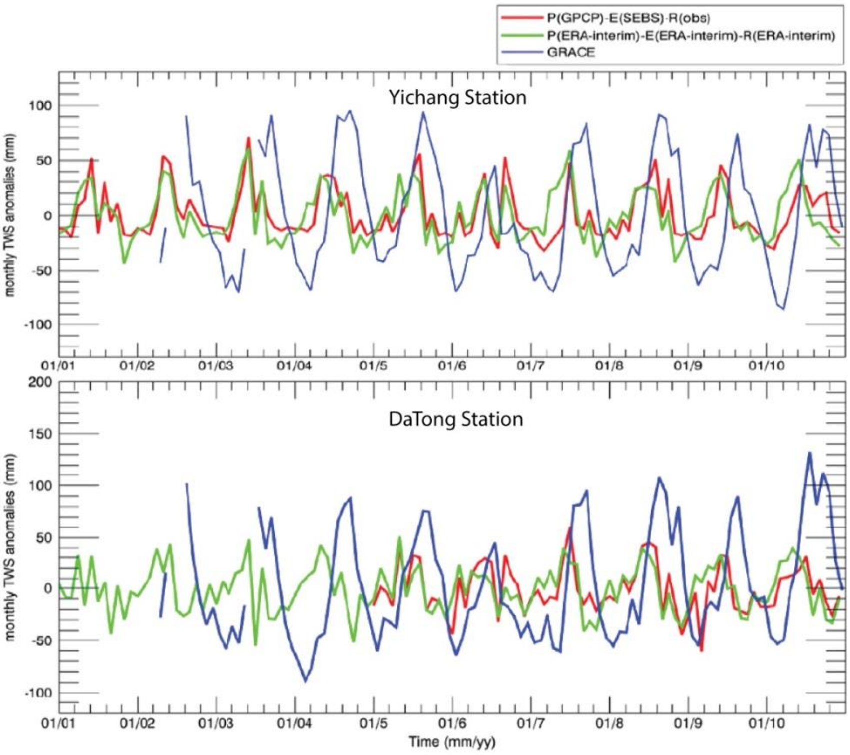

Figure 8 Terrestrial water storage anomaly over (Top) Yichang Station and (bottom) Datong Station.

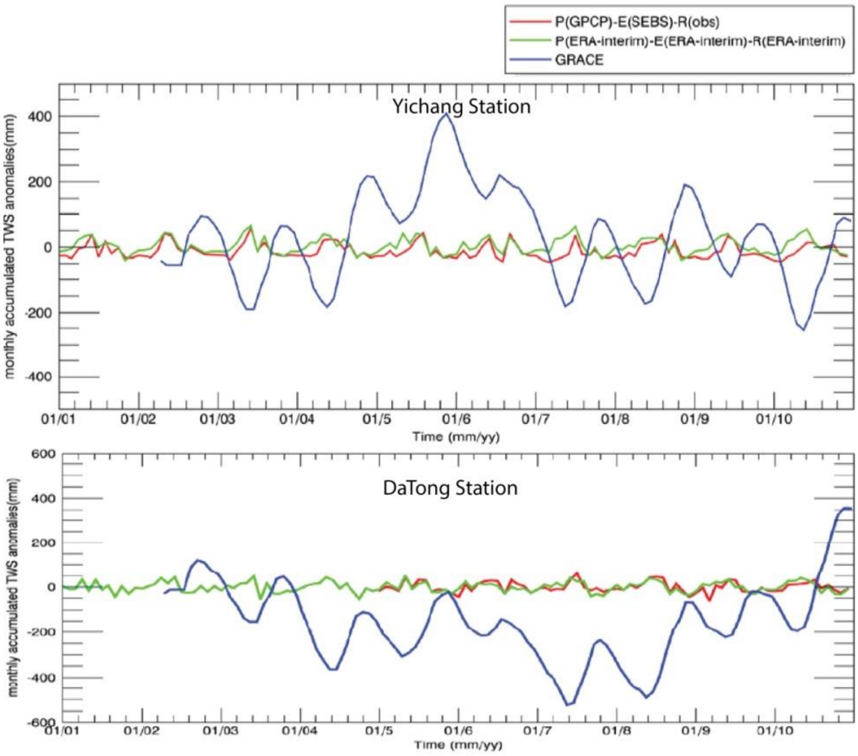

Figure 9 Cumulative terrestrial water storage anomaly over *(Top) Yichang Station and (bottom) Datong Station.*

From the results shown in Figures 8-9, it can be seen that the TWS derived from the observation data and from ERA reanalysis data are consistent with each other, indicating both datasets capture the surface dynamics of the water cycle fluxes. However the TWS derived from GRACE has somewhat larger amplitudes than those from observation data and reanalysis data, indicating deep groundwater contributions. From Figure 9, it may be concluded that after the filling of the Three Gauge Dam reservoir (Yichang station, upper panel in Figure 9) started in 2004, the storage of the upper reach increased in the following years in 2005-2007 period, but returned to the average gradually afterwards. This indicates the dominant climatic control of the Yangtze river system in the upper reach. For the lower reach, it can be seen that the there was a reduced storage from 2004-2008 but it gradually returned to its pre-Three Gauge stage afterwards.

It is noted that the monthly TWS anomaly and cumulative anomaly over Yichang Station (i.e. the upper reach), calculated using an earlier version of GRACE data (RL04 ssv201008), are completely different from the results shown in Figure 8-9. For the TWS anomaly, the earlier version of GRACE data has a phase difference of about 10 days, when compared to the current version (results not shown). For the cumulative TWS anomaly, the earlier version doesn't show the increase of monthly accumulative TWS anomaly after 2004 (results not shown), which was expected due to the filling of the Three Gauge Dam reservoir (upper panel in Figure 9).

The reason for the difference between the two versions of GRACE data can be very complex. Many parameter choices and solution strategies are possible for the complex inversion of relative ranging observations between the two formation-flying GRACE spacecraft, the precise orbit determination via GPS and various corrections for spacecraft accelerations not related to gravity changes. One of the most important parameters for the post-processing of GRACE observations is the land grid scaling coefficients, the use of which enables the representation of surface mass variations at small spatial scales. Without the scaling coefficients, the mass variations at small spatial scales tend to be attenuated. On the other hand, the scaling coefficients are computed by applying the same filters (e.g. destriping, Gaussian, and degree 60 filters) applied to the GRACE data to a land-hydrology model (i.e. NCAR's CLM4), through which the gain factor is derived by minimizing the difference between the model's smoothed and unfiltered monthly water storage variations at any geographic location. With its origin, the gain factors tend to be dominated by the annual cycles of water storage variations. Meanwhile, the inter-annual trends in particular in hydrology models are very uncertain, it is therefore suggested that it may not be suitable to quantify trends. Nevertheless, within the accepted error ranges, GRACE data is still useful for detecting trends [*Wang et al.* 2011; *Wang et al.* 2013].

From the above, it is obvious that the exact explanation for the different results over the upper Yangtze reach, calculated from the two data versions, needs intensive dedicated research. However, the log file about each change made for the production of GRACE data can be accessed easily from ftp://podaac.jpl.nasa.gov/ allData/grace /docs/ , therefore it is the first action to check if the difference can be attributed to what has been changed during the production of new version dataset. Another place to look at such information is: http://www.csr.utexas.edu/grace/RL05.html . From the above two sources, the text quoted below may explain the possible reasons for the different results over the upper Yangtze reach with two versions of data : "..*NOTE-1 (2012-07-17): We have replaced the CSR-GSM products for the following four months: July-2004; October-2004; March-2005 and February-2006. The GSM products for these four months have been updated with the refined data editing. While the GAC/GAD products for these four months are unchanged, they have been updated as well, for consistency. If you downloaded the data products for these four months prior to July 17, 2012, please download the replacement products. No other products previously delivered are so affected…*"

4. Essentials of Consistency Validation for Current Practice Examples

According to the above, the climate monitoring needs to ensure consistency and quality of the products, realization of which requires thorough inter-comparison at two aspects:

a. Thorough (inter)comparison among multiple independent datasets/products for a specific climate variable;
b. Thorough (inter)comparison among different climate variables, which are physically interlinked.

The first type of inter-comparison can be implemented at point or grid scales to check the product differences of a specific climate variable and the reasons for such differences, while the second type of inter-comparison will check the physical consistency from functional point of view of different climate variables. In addition, the cross-cutting validation using data assimilating system is an alternative way to check physical consistency of ECVs across domains:

c. Cross-cutting consistency validation among different climate variables across domains, using data assimilating system.

The above three types of validation activities can be regarded as ***the quality consistency check*** on climate data products.

At the same time, it is important to recognize that different methodologies and verification approaches used for inter-comparison may lead to the same conclusion, but with different reasons behind. Other than this, the generation processes (/production chain) of different datasets and products differing from each other can complicate the analysis of the (inter)comparison results, if all the necessary information are not provided.

To facilitate the assessment on this kind (e.g. production chain), the CORE-CLIMAX project[1] proposed a System Maturity Matrix (SMM), which is adapted from *Bates* and *Privette* [2012; *EUMETSAT* 2014]. The SMM is a tool to assess the system maturity of a CDR. SMM basically assesses whether CDR generation procedures have been compliant with best practices developed and accumulated by the scientific and engineering communities. This can be regarded as ***the process consistency check***.

5. Generic Strategy of Consistency Validation

For the quality consistency check, one of the first step is to understand the validation requirement (see section 2.1). It shall be realized that for different users (therefore about different applications) the requirements will be different.

For example, there are two kinds of reanalysis (e.g. NWP-like reanalysis or reanalysis for climate change assessment), having different requirements in terms of data usage. A traditional NWP-like reanalysis (e.g. for the past 30yrs) tends to assimilate all available observations unless they are known to be unusable for certain reasons. The climate reanalysis, going further back in time (e.g. for the past 100 yrs), on the other hand, only assimilates those observations that are known to be suitable for climate applications. It implies that a climate reanalysis requires extra efforts in validating the input data than the NWP-like reanalysis.

With that in mind, it is important to implement ***user (validation) requirement review*** as the first step, which include (but not limited to):

- definition of user requirement on products (e.g. coverage, vertical resolution, spatial and temporal resolution, data length etc.);
- consistency validation requirement (e.g. only inter-comparison required or the ECV consistency across domains required);
- service specification (e.g. near-real-time monitoring/forecast at European/Global Scale, value added products, satellite retrievals etc.);
- requirement for measures and metrics (e.g. root mean square error, relative frequency histograms, Pearson's correlation coefficient etc.);

[1] http://www.coreclimax.eu

- requirement for independent reference observation data (e.g. is traceable to in-situ measurements?);
- requirement for equivalent products (e.g. is the heuristic reference needed?).

Based on the requirement review, the dedicated ***validation plan*** can be made and may include (not limited to):

- definition of terminology (e.g. consistency, accuracy, stability etc.);
- description of data under evaluation (data processing and archiving center, model/data processor version, instrument, calibration version, log-file, input and initialization data, measured parameter, native data format, file name convention);
- reference data selection (e.g. the same as the above item, plus the information error budget of data comparison, characterization of sensitivity and information content);
- range of comparison;
- co-location criteria, conversion of units, temporal/spatial re-sampling and smoothing;
- performance and validation statistics (e.g. error budge analysis);
- description of the validation protocol;
- description of the validation process (e.g. both internal and external, can be referred to De3.32 [*Zeng et al.* 2014]).

It is noted that for the description of data under evaluation and the reference data selection, the known and relevant uncertainties shall be detailed. This will facilitate proper interpretation of the validation results, and traceability of the validation process.

After implementing the validation plan, all results of validation measures shall be integrated and published as a ***validation report***. This third step needs to coordinate and harmonize all validation activities/results, available quality information and the service endorsement information. It implies that the validation results from geophysical product and algorithm validation, through validation against service specifications and requirements, to the service endorsement by core/key users (e.g. external review) shall be all collected. The endorsement by core/key users facilitate the feedback loops of the validation process, which is the most important element of the whole validation processes. This external review element provides feedbacks from the end-users to the developers, producers and providers the new information for improving the quality of current products.

Arguably, the three steps identified above: 1) User (Validation) Requirement Review; 2) Validation Plan; and 3) Validation Report, are general in a way to validate products corresponding to user requirements, and cannot be referred to consistency validation. On the other hand, it is needed to understand how the consistency is defined before implementing consistency validation, which can be reflected/specified in the user (validation) requirement review. And then, in the validation plan, especially for the description of the data under evaluation and the reference data selection the detailed information shall be investigated, which will help to identify the causes when inconsistency being identified. For example, both show cases presented in section 3.5 (i.e. for in-situ) and section 3.7 (i.e. for hydrological cycle) indicate that the cause of inconsistency can be identified from the log-files of the data production. In addition, in the validation plan, the practical aspects of consistency validation can be identified, for example: co-location criteria, conversion of units, temporal/spatial re-sampling and smoothing. The show cases presented in section 3.4 (i.e. for the LAI product) is a good example on how the practical aspects of consistency validation can be defined.

6. Summaries and Discussions

6.1. The Consistency Validation Principle

The climate data store (CDS) for Copernicus Climate Change Services (CCCS) envisions that it shall include essential climate variables (ECVs), uncertainty estimates, reanalysis, multi-model data (e.g. seasonal forecasts and up-to-date climate projections), and in-situ and satellite data. Furthermore, CDS should contain only 'climate compliant' data, as defined by Evaluation and Quality Control working group of CCCS. This implicates that one may find certain ECV variable from different sources, and there is a need to do consistency check to determine which dataset are fit for certain particular purpose. On the other side, when there is a certain ECV variable under evaluation, one would like to collect as many as possible independent datasets, under the constraints that they are measuring the same thing, to do (inter)comparison, in order to understand comprehensively the associated uncertainty.

For the in-situ datasets, the homogeneity adjustment plays a critical role in producing long-term climate data. There are direct and indirect methodologies to do homogeneity detection and adjustment [*Peterson et al.* 1998]. However, it is still difficult to avoid all the inhomogeneities caused by, for example, changes in instrumentation, station moves, changes in the local environment (e.g. building construction), or the introduction of different observing practices like a new formula for calculating mean daily temperature or different observation times. The existence of independent datasets and data assimilation technique provide unprecedented opportunities to spot "bugs" in the in-situ datasets. The section 3.5 shows the use of reanalysis data to detect erroneous in-situ measurement. In the section 3.1.3, the high-resolution oceanic reanalysis help to identify the suspicious temperature profile observations, the assimilation of which into the system will cause abnormal salinity value. The use of satellite data can also help monitoring in-situ observations, for example, by using AATSR SST data, the error characteristics of the 'bad' buoy for measuring SST can be identified [*CMUG* 2013].

For the satellite, reanalysis, and multi-model datasets, the same principle of consistency validation as discussed above (e.g. collect as many as possible available independent datasets) is applicable. On the other hand, this principle needs to be constrained with the reference data selection procedure as mentioned in Section 5. It is to note that based on the discussion in Section 5 the collection of all kinds of validation information is to identify the missing validation information and processes (e.g. validation gaps), which are necessary to set priorities for future validation report. And, consistency validation is a key for identifying such gaps.

6.2. Identify Observation Gaps

Through consistency validation, it is also helpful to identify gaps for the current capacities of existing networks (see section 2.2). For example, taking water cycle closure as an example, we can thematically identify what needs to be measured. When the needed variable was compared to a dataset, one may find out that in the dataset only land fluxes were observed while soil moisture and soil temperature, or water vapor, or relative humidity were not. From this sense, one may identify the gap, the filling of which requires measurements of other relevant physical variables were needed. This can be useful to help bring different observation networks together, which may be established from different initiatives/projects/programs. In this way, one can suggest a way to gain added-value to current existing observation entities.

The consistency validation has identified that in the MACC assimilation system the direct adjustments to the meteorological parameters based on trace-gas observations is still not feasible, due to the insufficient stratospheric tracer observations [*Dee et al.* 2013]. Similarly, in the MyOcean assimilation system, due to the difficulty to obtain a reliable estimate of the net surface water flux, the surface freshwater budget in the system is manually set in a way to reduce errors in sea level anomalies assimilation [*Lellouche et al.* 2013]. These two examples may suggest the need for the future development of observing system on stratospheric tracer observations, and the continental runoff observations, for improving atmospheric and oceanic reanalysis, respectively.

For the albedo consistency validation, the inter-comparison between the satellite and the in-situ albedo reveals huge difference in albedo maxima, when the in situ albedo is measured over open area and the satellite pixel contains mainly forest (section 3.2). For the LAI example, due to the limitation of the ground data set, the validation of LAI over broadleaf evergreen and deciduous forest (e.g. canopy understory was not measured) were not yet implemented comprehensively. However, it is over such area, where contamination by cloud or snow limits the reliability of the reflectance values used as inputs in the algorithms [*Camacho et al.* 2013]. The above two consistency validation examples help to identify the current limitations in validating albedo and LAI, under certain circumstances. This will suggest the need for future development in the relevant observing systems.

6.3. Coupled Model in Reanalysis

In the hydrological cycle closure example implemented over Yangtze River Basin, it is assumed that the total water storage change at a scale of river basin equals to the input minus output. In this sense, the basic water balance equation (i.e. total water storage change = Precipitation – Evaporation - Runoff) can be used to identify what are needed for observation already. However, for the current case, only runoff data can be collected from in-situ observation. The precipitation data was from GPCP, the evaporation was calculated by using SEBS model. Apart from the difficulty of defining runoff data at grid scale, it is needed a harmonized approach to bring different sources of data together to do consistency check. The harmonized approach can include two parts: a) how to adjust different physical metrics to a common benchmark that enables them to be used to do consistency check; 2) a physically thematic framework shall be defined to do consistency check (e.g. in this case, the water balance equation).

Nevertheless, from the Yangtze river case, the TWS anomaly signal calculated from different sources of data (e.g. GPCP precipitation, SEBS evaporation and In-situ runoff) are consistent with that calculated from reanalysis data. It seems GPCP, SEBS evaporation, and in-situ Runoff data are consistent with each other, as they are compared with model results (ERA-Interim), which can be regarded as physically consistent system as it is constrained by the coupled model. On the other hand, the analysis schemes for the different components are separate and use different methodologies. For example, for the atmosphere-land domains, the screen-level parameter analysis is the first to be completed and is used as input for the soil moisture and snow analysis. The analyzed surface variables generate feedback for the upper-air analysis for the next assimilation window, through their influence on the first-guess forecast that propagates information from one cycle to the next. In this sense, this can be identified as weakly coupled system. The similar weakly coupled system exists for the atmosphere-ocean domains though.

The weakly coupled data assimilation scheme is widely used in reanalysis centers. One advantage of this weakly coupled approach is that, for example, the ocean initialization, obtained by running an ocean model with surface boundary conditions from an atmospheric (re)analysis, can benefit from the wealth of atmospheric observations and sophisticated atmospheric data assimilation methods. And, such uncoupled approach also permits modularity and easy implementation. The disadvantage is that the surface properties of both atmosphere (e.g. wind conditions and near surface temperature) and oceans (e.g. sea surface temperature) cannot be consistently assimilated within the separate assimilation systems.

The efforts in developing coupled data assimilation for atmosphere and ocean are currently on going at ECMWF (e.g. with ERA-CLIM2 project). It aims to develop a first coupled ocean-atmosphere reanalysis of the twentieth century, together with consistent estimates of carbon fluxes and stocks [*Dee et al.* 2013]. The new data assimilation scheme will be designed in a way to allow dynamic two-way exchange of information between ocean and atmosphere within a single analysis cycle, which can accommodate observations that are sensitive to both oceanic and atmospheric variables [*Dee et al.* 2013].

Nevertheless, it is recognized that both the observational record and the model have inherent uncertainties that are not always quantifiable. Therefore, it is important to expose all available information pertaining to these uncertainties, and make them accessible to the scientific community. It is also important to enable users to assess the observational information content of specific reanalyzed parameters as a function of space and time, depending on whether those parameters have been directly observed or indirectly constrained by observations of other parameters. All these information will allow end-users to draw meaningful inferences about the uncertainties in their own estimates, meeting the requirements for their specific applications [*Dee et al.* 2011].

Reference:

Bates, J. J. and J. L. Privette (2012). "A maturity model for assessing the completeness of climate data records." Eos, Transactions American Geophysical Union **93**(44): 441-441.

Berrisford, P., P. Kållberg, S. Kobayashi, D. Dee, S. Uppala, A. J. Simmons, P. Poli and H. Sato (2011). "Atmospheric conservation properties in ERA-Interim." Quarterly Journal of the Royal Meteorological Society **137**(659): 1381-1399.

Camacho, F., J. Cernicharo, R. Lacaze, F. Baret and M. Weiss (2013). "GEOV1: LAI, FAPAR essential climate variables and FCOVER global time series capitalizing over existing products. Part 2: Validation and intercomparison with reference products." Remote Sensing of Environment **137**(0): 310-329.

CEOS (2012). "The Response of the Committee on Earth Observation Satellites (CEOS) to the Global Climate Observing System Implementation Plan 2010 (GCOS IP - 10)." http://www.ceos.org/images/WGClimate/ceos_response_to_gcos_ip-10_24_september_2012_formatted.pdf, pp.2111.

CMUG (2013). "Delierable 4.1 Scientific Exploitation Plan." CMUG Deliverables: pp. 23.

CMUG (2014). "Deliverable 3.1 technical note on CMUG ECV Quality Assessment Report." CMUG Deliverables: pp.93.

Crosnier, L. and C. Le Provost (2007). "Inter-comparing five forecast operational systems in the North Atlantic and Mediterranean basins: The MERSEA-strand1 methodology." Journal of Marine Systems **65**(1–4): 354-375.

Dee, D. P., M. Balmaseda, G. Balsamo, R. Engelen, A. J. Simmons and J. N. Thépaut (2013). "Toward a Consistent Reanalysis of the Climate System." Bulletin of the American Meteorological Society **95**(8): 1235-1248.

Dee, D. P., E. Kallen, A. Simmons and L. Haimberger (2011). "Comments on "Reanalyses Suitable for Characterizing Long-Term Trends"." BAMS **DOIL: 10.1175/2010bams3070.1**.

Donlon, C. J., M. Martin, J. Stark, J. Roberts-Jones, E. Fiedler and W. Wimmer (2012). "The Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA) system." Remote Sensing of Environment **116**(0): 140-158.

EUMETSAT (2014). "CORE-CLIMAX System Maturity Matrix Instruction Manual." Deliverable D222, CC/EUM/MAN/13/002, http://www.coreclimax.eu/?q=Reports: pp.41.

Fernandes, R., S. Plummer, J. Nightingale and e. al. (2014). "Global Leaf Area Index Product Validation Good Practices." CEOS WGCV Land Product Validation Sub-group: pp. 76.

Ferry, N., L. Parent, G. Garric, C. Bricaud, C. E. Testut, O. Le Galloudec, J. M. Lellouche, M. Drevillon, E. Greiner, B. Barnier, J. M. Molines, N. C. Jourdai, S. Guinehut, C. Cabanes and L. Zawadzki (2012). "GLORYS2V1 global ocean reanalysis of the altimetric era (1992-2009) at meso scale." Mercator Newsleter **44**(29-39).

Fujiwara, M., S. Polavarapu and D. Jackson (2013). "A proposal of the SPARC Reanalysis/Analysis Intercomparison Project." SPARC Newsletter **38**: 14-17.

GCOS (2003). "GCOS Climate Monitoring Principles." http://www.wmo.int/pages/prog/gcos/index.php?name=ClimateMonitoringPrinciples.

GCOS (2010a). "Guideline for the Generation of Datasets and Products Meeting GCOS Requirements." GCOS-143: Geneva, Switzerland, pp. 12.

GCOS (2010b). "Implementation plan for the global observing system for climate in support of the UNFCCC (2010 UPDATE)." GCOS-138: Geneva, Switzerland, pp180.

GCOS (2011). "Systematic observation requirements for satellite-based data produts for climate, 2011 Update, Suplemental details to the satellite-based component of the "Implementation

Plan for the Global Observing System for Climate in Support of the UNFCCC (2010 Update)"." GCOS-154: Geneva, Switzerland, pp.139.
Gregow, H. W. l., A. Kaiser-Weiss, A. Obregon, D. Tan and P. Poli (2014a). "Deliverable D5.53 Procedure for comparing reanalysis, and comparing reanalysis to assimilated observations and CDRs." CORE-CLIMAX Deliverables: pp.45.
Gregow, H. W. l., H. Makela, K. Jylha, A. Kaiser-Weiss, A. Obregon, K. Kaspar, P. Poli, D. Tan, s. kEKKI, A.-N. Arslan and K. Eichler (2014b). "Deliverable D5.52 Reanalysis and user needs with respect to climate change services." CORE-CLIMAX Deliverables: pp.78.
Haylock, M. R., N. Hofstra, A. M. G. Klein Tank, E. J. Klok, P. D. Jones and M. New (2008). "A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006." Journal of Geophysical Research: Atmospheres **113**(D20): D20119.
Hofstra, N., M. Haylock, M. New and P. D. Jones (2009). "Testing E-OBS European high-resolution gridded data set of daily precipitation and surface temperature." Journal of Geophysical Research: Atmospheres **114**(D21): D21101.
Hofstra, N., M. New and C. McSweeney (2010). "The influence of interpolation and station network density on the distributions and trends of climate variables in gridded daily data." Climate Dynamics **35**(5): 841-858.
Houghton, J., J. Townshend, K. Dawson, P. Mason, J. Zillman and A. Simmons (2012). "The GCOS at 20 years: the origini, achievement and future development of the Global Climate Observing System." Weather **67**(9): 227-235.
IGACO (2004). "The chaning atmosphere. An integrated global atmospheric chemistry observation theme for the IGOS partnership. Report of the Integrated Global Atmospheric Chemistry Observation (IGACO) theme team." ESA SP-1282, GAW No. 159, WMO-TD No. 1235.
Kaspar, K., G. Müller-Westermeier, E. Penda, H. Mächel, K. Zimmermann, A. Kaiser-Weiss and T. Deutschländer (2013). "Monitoring of climate change in Germany – data, products and services of Germany's National Climate Data Centre." Adv. Sci. Res. **10**: 99-106.
Klok, E. J. and A. M. G. Klein Tank (2009). "Updated and extended European dataset of daily climate observations." Int. J. Climatol. **29**: 1182, doi:1110.1002/joc.1779.
Kyselý, J. and E. Plavcová (2010). "A critical remark on the applicability of E-OBS European gridded temperature data set for validating control climate simulations." Journal of Geophysical Research: Atmospheres **115**(D23): D23118.
Lambert, J.-C. (2013). "Service Validation Protocol (V2)." macc-II Deliverable D153.1 **Approved by A. De Rudder (BIRA-IASB) and V. -H. Peuch (ECMWF)**(Endorsed by Sub-Project Leaders: VAL/H. J. Heskes (KNMI), INT/T. Holzer-Popp (DLR), POL/Ø. Hov (MET.NO), GHG/F. Chevallier (IPSL-LSCE), GRG/M. Schultz (FZJ), AER/O. Boucher (CNRS-LMD), GDA/R. Engelen (ECMWF), H. EDA/Elbern (FRIUUK), ENS/V. Marécal (CNRM-MF), EVA/L. Rouil (INERIS), RAD/ M. Schroedter-Homscheidt (DLR) and L. Wald (Armines), OBS/L. Tarrason (NILU), EMI/C. Granier (UPMC-LATMOS), FIR/J. Kaiser (ECMWF)): PP.46.
Lawrimore, J. (2014). "Review of specific obsrving system activities related to observing systeme design (GCOS)." CBS/OPAG-IOS/OSDW1/Doc 5.4: pp.19.
Lellouche, J. M., O. Le Galloudec, M. Drévillon, C. Régnier, E. Greiner, G. Garric, N. Ferry, C. Desportes, C. E. Testut, C. Bricaud, R. Bourdallé-Badie, B. Tranchant, M. Benkiran, Y. Drillet, A. Daudin and C. De Nicola (2013). "Evaluation of global monitoring and forecasting systems at Mercator Océan." Ocean Sci. **9**(1): 57-81.
Lorenz, C. and H. Kunstmann (2012). "The Hydrological Cycle in Three State-of-the-Art Reanalyses: Intercomparison and Performance Analysis." Journal of Hydrometeorology **13**(5): 1397-1420.
Manninen, T. and P. Stenberg (2009). "Simulation of the effect of snow covered forest floor on the total forest albedo." Agricultural and Forest Meteorology **149**(2): 303-319.
Mapes, B. E. and J. T. Bacmeister (2012). "Diagnosis of Tropical Biases and the MJO from Patterns in the MERRA Analysis Tendency Fields." Journal of Climate **25**(18): 6202-6214.

Maraun, D., T. Osborn and H. Rust (2012). "The influence of synoptic airflow on UK daily precipitation extremes. Part II: regional climate model and E-OBS data validation." Climate Dynamics **39**(1-2): 287-301.

Metsämäki, S., O.-P. Mattila, J. Pulliainen, K. Niemi, K. Luojus and K. Böttcher (2012). "An optical reflectance model-based method for fractional snow cover mapping applicable to continental scale." Remote Sensing of Environment **123**(0): 508-521.

Parker, J. and Z. Stott (2013). "Coordinating Climate Science Observations (final report)." http://www.ukeof.org.uk/documents/CCOP2014FINALreport2011.pdf, pp.2104.

Peterson, T. C., D. R. Easterling, T. R. Karl, P. Groisman, N. Nicholls, N. Plummer, S. Torok, I. Auer, R. Boehm, D. Gullett, L. Vincent, R. Heino, H. Tuomenvirta, O. Mestre, T. Szentimrey, J. Salinger, E. J. Førland, I. Hanssen-Bauer, H. Alexandersson, P. Jones and D. Parker (1998). "Homogeneity adjustments of in situ atmospheric climate data: a review." International Journal of Climatology **18**(13): 1493-1517.

QA4EO "A Quality Assurance framework for Earth Observation, established by the CEOS. It consists of ten distinct key guidelines linked through an overarching document (the QA4EO Guidelines Framework) and more community-specific QA4EO procedures, all available on http://qa4eo.org/documentation.html A short QA4EO "user" guide has been produced to provide background into QA4EO and how one would start implementing it (http://qa4eo.org/docs/QA4EO_guide.pdf)."

Riihelä, A., T. Manninen, V. Laine, K. Andersson and F. Kaspar (2013). "CLARA-SAL: a global 28 yr timeseries of Earth's black-sky surface albedo." Atmos. Chem. Phys. **13**(7): 3743-3762.

Rudder, D. (2014). "Service Validation Report (SVR)." MACC-II Deliverable D153.3 **IASB-BIRA (#6)**(Reviewers: V. -H. Peuch (ECMWF), J. -C. Lambert (IASB -BIRA)): pp.22.

Saunders, R. C. t. (2014). "Deliverable 2.3 Technical note on validation and user assessment." CMUG Deliverables: pp.15.

Seidel, D. J., F. H. Berger, F. Immler, M. Sommer, H. Vömel, H. J. Diamond, J. Dykema, D. Goodrich, W. Murray, T. Peterson, D. Sisterson, P. Thorne and J. Wang (2009). "Reference Upper-Air Observations for Climate: Rationale, Progress, and Plans." Bulletin of the American Meteorological Society **90**(3): 361-369.

Simmons, A. J. and A. Hollingsworth (2002). "Some aspects of the improvement in skill of numerical weather prediction." Quarterly Journal of the Royal Meteorological Society **128**(580): 647-677.

Trenberth, K. E., J. T. Fasullo and J. Mackaro (2011). "Atmospheric Moisture Transports from Ocean to Land and Global Energy Flows in Reanalyses." Journal of Climate **24**(18): 4907-4924.

van den Besselaar, E. J. M., M. R. Haylock, G. van der Schrier and A. M. G. Klein Tank (2011). "A European daily high-resolution observational gridded data set of sea level pressure." Journal of Geophysical Research: Atmospheres **116**(D11): D11110.

Wang, H., L. Jia, H. Steffen, P. Wu, L. Jiang, H. Hsu, L. Xiang, Z. Wang and B. Hu (2013). "Increased water storage in North America and Scandinavia from GRACE gravity data." Nature Geosci **6**(1): 38-42.

Wang, X., C. de Linage, J. Famiglietti and C. S. Zender (2011). "Gravity Recovery and Climate Experiment (GRACE) detection of water storage changes in the Three Gorges Reservoir of China and comparison with in situ measurements." Water Resources Research **47**(12): W12502.

Wilson, J., M. Dowell and A. Belward (2010). "European capacity for monitoring and assimilating space based climate change observations - status and prospects." JRC Scientific and Technical Reports (EUR 24273 EN): European Commission, Joint Research Center, Institute for Environment and Sustainability, PP.47.

WMO "WMO Rolling Requirements: Based on WMO-TD No. 1052, SAT-26, available on http://www.wmo.int/pages/prog/sat/Requirements/Observational-requirements_web.xls."

WMO (2013). "Guide to the Global Observing System." WMO-488: Geneva, Switzerlands, pp. 174.

Zeng, Y., Z. Su, J. Schulz, J.-C. Calvet, P. Poli, D. Tang, T. Manninen, A. Riihelä, C.-M. Tanis, A.-N. Arslan, E. Swinnen, C. Tote, A. Obregon, K.-W. Andre, R. Roebeling, V. John, K. Kaspar, W. J. Timmermans and J. Timmermans (2014). "Generic Validation Strategy for CDRs/ECVs." CORE-CLIMAX Deliverables: pp. 1-31.

Appendix A: Overview of Space Observable ECVs

Table A-1 Overview of Space Observable ECVs from different domains [*GCOS* 2011]

Atmosphere ECV	Global Products requiring Satellite Observations	Fundamental Climate Data Records required for Product Generation (from past, current and future missions)
Surface Wind Speed and Direction	Surface wind retrievals	Passive microwave radiances and radar backscatter
Precipitation	Estimates of liquid and solid precipitation, derived from specific instruments and provided by composite products	Passive microwave radiances Geostationary VIS/NIR/IR radiances
Upper-air Temperature	Upper-air temperature retrievals Temperature of deep atmospheric layers	Passive microwave and IR radiances GNSS radio occultation bending angles
Upper-air Wind Speed and Direction	Upper-air wind retrievals	VIS/IR imager radiances Doppler wind lidar
Water Vapour	Total column water vapour Tropospheric and lower-stratospheric profiles of water vapour Upper tropospheric humidity	Passive microwave radiances UV/VIS imager radiances IR and microwave radiances Limb soundings
Cloud Properties	Cloud amount, top pressure and temperature, optical depth, water path and effective particle radius	VIS/IR imager radiances IR and microwave radiances Lidar
Earth Radiation Budget	Earth radiation budget (top-of-atmosphere and surface) Total and spectrally-resolved solar irradiance	Broadband radiances Spectrally-resolved solar irradiances Geostationary multispectral imager radiances
Carbon Dioxide, Methane and other GHGs	Retrievals of greenhouse gases, such as CO2 and CH4, of sufficient quality to estimate regional sources and sinks	NIR/IR radiances
Ozone	Total column ozone Tropospheric ozone Ozone profiles from upper troposphere to mesosphere	UV/VIS and IR/microwave radiances, from nadir and limb sounding
Aerosol Properties	Aerosol optical depth Aerosol single scattering albedo Aerosol layer height Aerosol extinction profiles from the troposphere to at least 35km	UV/VIS/NIR/SWIR and TIR radiances UV/VIS/IR limb sounding (scatter, emission, occultation) Lidar profiling
Precursors supporting the Ozone and Aerosol ECVs	Retrievals of precursors for aerosols and ozone such as NO2, SO2, HCHO and CO	UV/VIS/NIR/SWIR and TIR radiances UV/VIS/IR limb sounding (scatter, emission, occultation)
Ocean ECV	**Global Products requiring Satellite Observations**	**Fundamental Climate Data Records required for Product Generation (from past, current and future missions)**
Sea-surface Temperature	Integrated sea-surface temperature analyses based on satellite and *in-situ* data records	Single and multi-view IR and microwave imager radiances
Sea-surface Salinity	Datasets for research on identification of changes in sea-surface salinity	Microwave radiances
Sea Level	Sea-level global mean and regional	Altimetry

	variability	
Sea State	Wave height, supported by other measures of sea state (wave direction, wavelength, time period)	Altimetry
Sea Ice	Sea-ice concentration/extent/edge, supported by sea-ice thickness and sea-ice drift	Passive and active microwave and visible imager radiances, supported by Synthetic Aperture Radar (SAR) altimetry
Ocean Colour	Ocean colour radiometry – water leaving radiance Oceanic chlorophyll-a concentration, derived from ocean colour radiometry	Multispectral VIS imager radiances
Terrestrial ECV (and supporting Variable)	**Global Products requiring Satellite Observations**	**Fundamental Climate Data Records required for Product Generation (from past, current and future missions)**
Lakes	Lake levels and areas of lakes in the Global Terrestrial Network for Lakes (GTN-L)	VIS/NIR imager radiances, and radar imager radiances Altimetry
Snow Cover	Snow areal extent, supplemented by snow water equivalent	Moderate-resolution VIS/NIR/IR and passive microwave imager radiances
Glaciers and Ice Caps	2D vector outlines of glaciers and ice caps (delineating glacier area), supplemented by digital elevation models for drainage divides and topographic parameters	High-resolution VIS/NIR/SWIR optical imager radiances, supplemented by microwave InSAR and along-track optical stereo imaging
Ice Sheets	Ice-sheet elevation changes, supplemented by fields of ice velocity and ice-mass change	Radar and laser altimetry, supplemented by:SAR, gravity
Albedo	Reflectance anisotropy (BRDF), black-sky and white-sky albedo	Multispectral and multiangular imager radiances
Land Cover	Moderate-resolution maps of land-cover type High-resolution maps of land-cover type, for the detection of land-cover change	Moderate-resolution multispectral VIS/NIR imager radiances High-resolution multispectral VIS/NIR imager radiances, supplemented by radar
FAPAR	Maps of the Fraction of Absorbed Photosynthetically Active Radiation	VIS/NIR multispectral imager radiances
LAI	Maps of Leaf Area Index	VIS/NIR multispectral imager radiances
Biomass	Regional and global above-ground forest biomass	Long-wavelength radar and lidar
Fire Disturbance	Maps of burnt area, supplemented by active-fire maps and fire-radiative power	VIS/NIR/SWIR/TIR moderate-resolution multispectral imager radiances
Soil Moisture	Research towards global near-surface soil-moisture map (up to 10cm soil depth)	Active and passive microwave
Land-surface Temperature	Land-surface temperature records to support generation of land ECVs	High-resolution IR radiances from geostationary and polar-orbiting satellites; Microwave radiances from polar-orbiting satellites

Appendix B: Overview of “Satellite – In Situ” observing pairs

Table B-1 Overview of “Satellite-In Situ” observing pairs over Atmospheric domain-Surface [*GCOS* 2010b]

ATMOSPHERIC DOMAIN – SURFACE				
ECV	**Contributing Network(s)**	**Status**	**Contributing Satellite Data**	**Status**
Wind Speed/ Direction	GCOS Surface Network (subset of full WWW/GOS surface network). WWW/GOS surface synoptic network. Additional national networks. Buoys and ships (see Ocean Surface section).	Wind is still not included in GSN.	Scatterometer. Passive microwave for wind speed. Polarimetric microwave radiometry for wind vectors	Uncertain operational continuity of two-scatterometer constellation
Precipitation	GCOS Surface Network (subset of full WWW/GOS surface synoptic network). Full WWW/GOS surface synoptic network. Additional national meteorological and hydrological gauge networks; island networks. Surface-based radar networks. Buoys	At least 95 % of stations are active, but only about 80% transmit CLIMAT reports. Quality of data and quantity of reports are variable. Most countries operate national high-resolution precipitation networks, but data are often not available internationally, or available only with time delay. Radar data not globally exchanged; spatial and temporal sampling limitations.	Passive microwave, VIS/IR on GEO. Precipitation radar.	High priority for climate applications Uncertain continuity of precipitation radar, Temporal and spatial sampling limitations.
Water Vapour	GCOS Surface Network (subset of full WWW/GOS surface synoptic network); Full WWW/GOS surface synoptic network. Ships and moored buoys	Water vapour is only partly included in CLIMAT reports, and not monitored. VOSClim stable; VOS fleet declining; no measurement from drifting buoys and only from a subset of moored buoys		
Surface Radiation Budget	BSRN. WWW/GOS surface synoptic network.	High-quality data, but coverage should be extended and continuity secured. Quality and coverage of routine radiation data is inadequate for climate purposes.	GEWEX Surface Radiation Budget project	Solar from satellites For longwave, satellite data are used to estimate cloud parameters and near-surface thermodynamics fields are typically taken from NWP models

ATMOSPHERIC DOMAIN – SURFACE				
ECV	**Contributing Network(s)**	**Status**	**Contributing Satellite Data**	**Status**
Temperature	GCOS Surface Network (subset of full WWW/GOS surface synoptic network). Full WWW/GOS surface synoptic network. Buoys and ships. Additional national networks (see also Oceanic section, Sea-surface Temperature ECV).	At least 95 % of stations are active, but only about 80% transmit CLIMAT reports. Need data from entire network to be available for climate purposes; data receipt from many countries is inadequate. VOSClim stable; VOS fleet declining; no measurements from drifting buoys.	Sea-surface temperature (IR, microwave) has strong influence on analysis of air temperature over the ocean.	Operationally supported .
Pressure	GCOS Surface Network (subset of full WWW/GOS surface synoptic network). Full WWW/GOS surface synoptic network. Additional national networks. Buoys and ships (see Ocean Surface section).	At least 95 % of stations are active, but only about 80% transmit CLIMAT reports. Some inconsistencies in pressure reduction methods to mean sea level. Some national networks inadequate for climate studies.		

Table B-2 Overview of "Satellite-In Situ" observing pairs over Atmospheric domain – upper air [*GCOS* 2010b]

ATMOSP ERIC DOMAIN – UPPER AIR				
ECV	**Contributing Network(s)**	**Status**	**Contributing Satellite Data**	**Status**
Temperature	Reference network of high- quality and high-altitude radiosondes (GRUAN). GCOS Upper-Air Network (subset of full WWW/GOS radiosondes network) Full WWW/GOS radiosonde network. Commercial aircraft.	International cooperation continues to work towards establishing the reference network. About 90% of stations are reporting regularly. Many stations do not provide two observations each day. Aircraft observations are valuable but limited to specific routes and levels except near airports.	Microwave sounders GNSS radio occultation. Infrared sounders	Need to ensure continuity of MSU-like radiance bands. Continuity for GNSS RO constellation needs to be secured

Wind Speed and Direction	GCOS Upper-Air Network (subset of full WWW/GOS radiosondes network). Full WWW/GOS radiosonde network. PILOT balloons	About 90% of stations are reporting regularly; only two completely silent	Visible and infrared (atmospheric motion vectors) from geostationary and polar orbit satellites. Lidar	Continuity of some polar winds at risk. Awaiting ADM/Aeolus demonstration; no continuity planned.

ATMOSPHERIC DOMAIN – UPPER-AIR				
ECV	**Contributing Network(s)**	**Status**	**Contributing Satellite Data**	**Status**
Wind Speed and Direction (cont'd)	Radar (profilers). Commercial aircraft.	Radar data are not globally distributed. Aircraft observations are valuable but limited to specific routes except near airports. .		
Water Vapour	Reference network of high-quality and high- altitude radiosondes (GRUAN). GCOS Upper-Air Network (subset of full WWW/GOS radiosondes network). Full WWW/GOS radiosonde network. Ground-based GNSS receiver network. Commercial aircraft.	Accurate reference sondes measuring upper-tropospheric and lower-stratospheric humidity are needed. Accuracy of water vapour measurements is improving, but is still inadequate for climate purposes in the upper troposphere and lower stratosphere. Wider international exchange of data is needed Aircraft data are potentially useful.	Microwave imagers and sounders; Infrared sounders GNSS radio occultation; Infrared and micro-wave limb sounders Solar occultation NIR images over land	Continuity assured for operational microwave and IR sounders; Continuity uncertain for microwave imagery Continuity uncertain for research satellites and GNSS constellation.
Cloud Properties	Surface observations (GSN, WWW/GOS, VOS). Cloud radar and lidar.	Surface observations of cloud cover provide an historical but uncertain record, and continuity is a concern; Reprocessing of cloud data is needed. Research-based networks	Visible, infrared and microwave radiances from geostationary and polar orbiting satellites; Cloud radar and lidar (research).	Cloud top temperature, microphysical properties and coverage are all operational.
Earth Radiation Budget			Broadband short- and longwave and total	Continuity and good calibration of

Table B-3 Overview of "Satellite-In Situ" observing pairs over Atmospheric domain – Composition[*GCOS* 2010b]

ATMOSPHERIC DOMAIN – COMPOSITION				
ECV	**Contributing Network(s)**	**Status**	**Contributing Satellite Data**[69]	**Status**
Carbon Dioxide	WMO GAW Global Atmospheric CO2 Monitoring Network (major contribution to the GCOS comprehensive network for CO2) consisting of:		SWIR and high-resolution IR	Continuity in IR operational instruments but products are immature and limited; A dedicated research satellite mission to provide better global products has been launched in 2009 (GOSAT), but continuity of such SWIR measurements need to be assured.
	WMO GAW[70] continuous surface monitoring network.	Operational; Partial network; Operational data management.		
	WMO GAW surface flask sampling network.	Operational; Partial network; Operational data management.		
	Airborne sampling (JAL, CARIBIC).	Limited operational aircraft vertical profiling initiated.		
	WMO GAW TCCON network (ground-based FTIR)	Operational, partial network		
Methane and other long-lived greenhouse gases[71]	WMO GAW Global Atmospheric CH_4 Monitoring Network ((major contribution to the GCOS comprehensive network for CH_4), consisting of:		IR nadir sounders SWIR nadir sounders	Satellite measurements on CH_4 are maturing and are part of operational satellites.
	GAW continuous surface monitoring network.	Operational; Partial network; Operational data management.	IR and microwave limb sounders	MLS, HIRDLS performs N_2O measurements in the stratosphere as well as of the other GHGs. Future research satellites might continue this, but there is uncertain continuity of profiling limb sounders.
	GAW surface flask sampling network.	Operational; Partial network; Operational data management.		
	AGAGE, SOGE and University of California at Irvine, USA.	Operational; Partial network; Operational data management.		

ATMOSPHERIC DOMAIN – COMPOSITION				
ECV	**Contributing Network(s)**	**Status**	**Contributing Satellite Data[69]**	**Status**
Methane and other long-lived greenhouse gases (cont'd)	Airborne sampling (JAL, CARIBIC, MOZAIC). NDACC	Limited operational aircraft vertical profiling initiated. Operational; Partial network; Operational data management		
Ozone	WMO GAW GCOS Global Baseline Profile Ozone Network (GAW ozonesonde network, including NASA SHADOZ and NDACC). WMO GAW GCOS Global Baseline Total Ozone Network (GAW column ozone network (filter, Dobson and Brewer stations)). NDACC	Mature operational balloon sonde network. Mature operational ground-based total column network. Operational; Partial network; Operational data management	UV nadir and limb sounders IR nadir sounders ; IR and MW limb sounders	Operational continuity for column ozone; No future operational or research high vertical resolution profiling currently planned after 2015.
Precursors (supporting the Aerosol and Ozone ECVs)	WMO GAW observing network for CO (continuous and flasks measurements) WMO GAW network for reactive nitrogen EMEP (GAW contributing network) Research programmes using MAXDOAS, SAOZ, FTIR and other techniques (for NO_2) *In situ* network from environmental agencies Aircraft (IAGOS, CO) NDACC	Operational; Partial network; Operational data management Currently in the stage of establishment, several stations world-wide Operational European network for monitoring of primary pollutants, Sparse, research-oriented Operational at national level Limited operational aircraft vertical profiling initiated Operational, Partial network; Operational data management	UV/VIS/NIR/SWIR sounders Nadir IR sounders	Precursors are measured by research satellites and operational satellites in future. Information on high spatial and temporal resolution is limited.
Aerosol Properties	BSRN; WMO GAW and contributing networks (AERONET, GALION); backscatter lidar networks.	Operational; Operational; Global coordination in progress.	Solar occultation VIS/ IR imagers Lidar profiling UV nadir Polarimetry Multi-angular viewing	Planned operational continuity for column products; No operational missions planned for aerosol type and aerosol size Research missions for profiling tropospheric aerosols; No plans for continuity of stratospheric profiling.

Table B-4 Overview of "Satellite-In Situ" observing pairs over Oceanic domain [*GCOS* 2010b]

Ocean-Surface ECV	**Satellite**	**Coordinating Body**	**In situ**	**Coordinating Body**
SST	Satellite IR (polar orbit and geostationary) AMSR-class microwave SST satellite	CEOS, CGMS	– Global surface drifting buoy array on 5x5 degree resolution (1250) – Global tropical moored buoy network (~120) – VOSClim and VOS fleet – Global reference mooring network (30-40) – Carbon VOS – Argo	– JCOMM DBCP – JCOMM Tropical Moored Buoy Implementation Panel (TIP/DBCP) – JCOMM SOT – OceanSITES (JCOMM) – IOCCP, OOPC pilot activity – GODAE, CLIVAR
SSS	The Microwave Imaging Radiometer with Aperture Synthesis (MIRAS) Aquarius (radiometer/scatterometer)	CEOS, CGMS	– VOSClim and VOS fleet – Global reference mooring network (30-40) – Carbon VOS – Argo	– JCOMM SOT – OceanSITES (JCOMM) – IOCCP, OOPC pilot activity – Argo Steering Team
Sea Level; **Sea-level** anomaly from steady state, **Sea state**	High-precision satellite altimetry Complementary-orbit (sun-synchronous) satellite altimetry	CEOS, CGMS	– VOSClim and VOS fleet – Global reference mooring network (30-40) – GLOSS Core Sea-level Network, plus regional/national networks	– JCOMM SOT – OceanSITES (JCOMM) – JCOMMGLOSS
Sea Ice	Satellite IR (polar orbit and geostationary) AMSR-class microwave satellite Surface vector wind satellite (two wide-swath scattero- meters are highly desired) Satellite SAR Satellite Visible Satellite altimetry	CEOS, CGMS	Sea-ice buoys Ice Profiling Sonar Sea-ice in-situ drilling Observations of snow characteristics on sea ice Observations by coastal stations	International Arctic/Antarctic Buoy Programme (JCOMM DBCP)
Wind Speed Surface Vector Wind	AMSR-class microwave satellite Surface vector wind satellite (two wide-swath scattero- meters are highly desired) Scatterometer Polarimetric microwave radiometry for wind vectors	CEOS, CGMS	Global tropical moored buoy network (~120) VOSClim and VOS fleet Global reference mooring network (30-40)	– JCOMM Tropical Moored Buoy Implementation Panel (TIP/DBCP) – JCOMM SOT – OceanSITES (JCOMM)
Ocean Colour; Chlorophyll concentration (biomass of Phytoplankton)	Ocean colour satellite (SeaWiFS-class)	CEOS, CGMS	VOSClim and VOS fleet Global reference mooring network (30-40)	– JCOMM SOT – OceanSITES (JCOMM)
Sea State	High-precision satellite altimetry Satellite SAR	CEOS, CGMS	VOSClim and VOS fleet Global reference mooring network (30-40)	– JCOMM SOT – OceanSITES (JCOMM)
Ocean-Subsurface ECV	**Satellite**	**Coordinating Body**	**In situ**	**Coordinating Body**
Temperature	N/A	N/A	About 40 repeat XBT line network About 120 tropical moorings 30-40 reference moorings network Sustained and repeated ship-based hydrography network Argo network Critical current & transport monitoring	JCOMM SOOP JCOMM TIP OceanSITES (JCOMM) IOCCP, CLIVAR, other national efforts Argo Steering Team CLIVAR, IOCCP
Salinity	N/A	N/A	About 120 tropical moorings 30-40 reference moorings network	JCOMM TIP OceanSITES (JCOMM)

			Sustained and repeated ship-based hydrography network Argo network	IOCCP, CLIVAR, other national efforts Argo steering Team
Current	N/A	N/A	About 120 tropical moorings 30-40 reference moorings network Sustained and repeated ship-based hydrography network Argo network	JCOMM TIP OceanSITES (JCOMM) IOCCP, CLIVAR, other national efforts Argo steering Team
Autonomously observable ECVs feasible	N/A	N/A	30-40 reference moorings network Sustained and repeated ship-based hydrography network	OceanSITES (JCOMM) IOCCP, CLIVAR, other national efforts
All feasible ECVs, including those that depend on obtaining water samples	N/A	N/A	30-40 reference moorings network Sustained and repeated ship-based hydrography network	OceanSITES (JCOMM) IOCCP, CLIVAR, other national efforts
Heat, Freshwater, Carbon transports, mass	N/A	N/A	Critical current & transport monitoring	CLIVAR, IOCCP

Table B-5 Overview of

Observing networks and systems contributing to the Terrestrial Domain [*GCOS* 2010b]

ECV	Contributing Network(s)	Status	Contributing Satellite Data	Status
River Discharge	GCOS/GTOS Baseline GTN-R based on TOPC priority list	Stations selected and partly agreed by host countries, non-contributing stations approached	Research concerning laser/radar altimetry for river levels and flow rates.	Operational laser altimeters not scheduled; EO-based network only research.
Lakes	GCOS/GTOS Baseline Lake Network based on TOPC priority list. To include freeze-up/break-up.	Stations selected, approached by HYDROLARE; GTN-L needs to be established;	Altimetry, high-resolution optical and radar imagery and reprocessing of archived data.	Operational laser altimeters not scheduled. Question mark over high-resolution systems continuity. EO-based network only research.
Above ground Biomass	FAO's FRA; FLUXNET; No global data centre for non-forest biomass.	No designated baseline network exists; FRA data not currently applicable for high-resolution spatial analysis.	Low-frequency radar, optical and laser altimetry.	Laser/radar missions currently planned; need to be implemented
Soil Carbon	National soil carbon surveys	No designated global network or data centre exists; major geographical gaps; FAO-IIASA world soil map	Not directly applicable	
Fire Disturbance	GOFC Regional Networks, GFMC	Some geographical gaps exist.	Optical and thermal.	Geostationary and moderate to high-resolution optical systems continuity required.
Soil Moisture	FLUXNET; WWW/GOS surface synoptic network International Soil Moisture Network	No designated baseline network exists.	Active and passive microwave missions	Continuity after the research missions required

ECV	Contributing Network(s)	Status	Contributing Satellite Data	Status
Ground Water (Levels, Use)	None, but framework for GGMN exists; many national archives of ground-water level exist.	Collection of aggregated data for GGMN has started; GTN-GW needs to be established	Gravity missions	Gravity measurements operat'l, continuity needs to be secured
Water Use (Area of Irrigated Land)	No network, but a single geo-referenced database exists.		Any high-/medium-resolution optical/ radar systems.	Lack of high-resolution optical continuity.
Snow Cover	WWW/GOS surface synoptic network (depth). National Networks (depth and snow water equivalent).	Synoptic and national networks have significant gaps and are ALL contracting. Northern and Southern Hemisphere monitored operationally for extent and duration.	Moderate to high resolution optical for extent/duration. Passive microwave for snow water equivalent. Geostationary satellites	Moderate to high resolution optical and microwave sensor system follow-on is programmed.
Glaciers and Ice Caps	GTN-G coordinates national monitoring networks.	Major geographic gaps still need to be closed; especially concerning glacier mass balance measurements inadequate.	Visible and infrared high-resolution; Stereo optical imagery; Synthetic Aperture Radar Satellite altimetry.	Lack of high-resolution optical satellite continuity. Satellite altimetry research missions will help; Lack of laser altimetry mission continuity.
Ice Sheets	Program for Arctic Regional Climate Assessment; International Trans-Antarctic Scientific Expedition.	Large uncertainty in mass balances and dynamics Ocean ice interaction major weakness	Gravity mission, Synthetic Aperture Radar and laser altimetry	Satellite altimetry research missions will help; Lack of laser altimetry mission continuity
Permafrost	GTN-P coordinates National Monitoring Networks.	Major geographical gaps. National data centres need to be established	Derived near-surface temperature and moisture (e.g., from ERS/Radarsat, MODIS, AMSR-E).	No direct operational sensors to detect permafrost; no products.
Albedo	CEOS WGCV; MODLAND; Atmospheric Radiation Measurement sites.	No designated reference network.	Multi-angular sensors. Geostationary Polar orbiters. GCMPs applied to measurements.	Use of operational meteorological satellites (SCOPE-CM Pilot Project) and moderate-resolution optical polar orbiters; Continuation of multi-angular missions required
Land Cover	FAO Global Land Cover Network; GOFC-GOLD.	First generation products available.	Any high-/medium-resolution optical/ radar systems.	Moderate resolution good;High-resolution optical system continuity required.
FAPAR	CEOS WGCV; FLUXNET; GTOS Net Primary Productivity.	Still no designated baseline network exists.	Optical, multi-spectral and multi-angular.	Moderate spatial resolution multi-spectral good; Continuation of multi-angular measurements required.
LAI	CEOS WGCV; FLUXNET; GTOS.	Still no designated baseline network exists.	Optical, multi-spectral and multi-angular.	Moderate spatial required resolution multi-spectral good; Continuation of multi-angular measurements