Gap Analysis for Integrated Atmospheric ECV CLImate Monitoring:

Initial input from WP5 to the gap analysis and impacts document

A Horizon 2020 project; Grant agreement: 640276

Date: draft June 2015

Lead Beneficiary: EUMETSAT

Nature: R

Dissemination level: PU

Work-package	WP 5 (Creation of a 'virtual observatory' visualization and data access facility)
Deliverable	D5.1
Title	Initial input from WP5 to the gap analysis and impacts document
Nature	Report
Dissemination	Public
Lead Beneficiary	EUMETSAT, Germany
Date	30 June 2015
Status	Preliminary
Authors	M. Doutriaux-Boucher (EUMETSAT), J. Schulz (EUMETSAT), T. Verhoelst (BIRA), J.-C. Lambert (BIRA), Jacques Descloitres (USTL), Tony Reale (NOAA)
Editors	?
Reviewers	?
Contacts	marie.doutriauxboucher@eumetsat.int, joerg.schulz@eumetsat.int
URL	http://www.gaia-clim.eu/

This document has been produced in the context of the GAIA-CLIM project. The research leading to these results has received funding from the European Union's Horizon 2020 Programme under grant agreement n° 640276. All information in this document is provided "as is" and no guarantee or warranty is given that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability. For the avoidance of all doubts, the European Commission has no liability in respect of this document, which is merely representing the authors' view.

Identification

Date: 29 June 2015

Gap Identifier G<wp>.<no>	Gap Type [2])	Keywords [3]) [Up to 10 (max)]	ECV(s) [Specify if not generic]	Gap Description (<100 characters)	Trace (both underlying WP deliverable(s) as well as external papers, reports etc)	Gap Impacts (Bulleted summary)	Envisaged Remedy (including timescale and cost estimate if possible)	Remedy addressed in GAIA-CLIM (Yes/No)
G5.01	**Technical**	**Data access**	**All**	**Access to data in multiple locations with different data policies and accessibility (e.g. speed of retrieving and unpacking, password protected, etc).**	• **See existing platforms:** **http://www.gruan.org** **http://tccon.ornl.gov/** **http://www.ndsc.ncep.noaa.gov/data/**	• **Lack of access or low speed access will be a problem for an interactive web tool.**	• **Store sample data locally.** • **Strive for agreement on WMO data policy or develop shared data policy.** • **Develop user friendly access, e.g. using Earth System Grid as for CMIP5.**	**To some extent**
G5.02	**Technical**	**Data format and structure**	**All**	**Access to data in multiple data format and structure (e.g. granularity of data). Lack of standardized metadata.**	• **http://www.ucar.edu/tools/applications_desc.jsp**	• **Loss of metadata due to reformatting of data by user.**	• **Sample data at the highest possible level to minimize the time in data transfer.** • **Employ meta data standard such as WIGOS.**	**To some extent**
G5.03	**Technical**	**Efficient collocation algorithm**	**All**	**Efficient data management to collocate observations.**	• **CCI toolbox** • **Giovanni**	• **Enables collocations for long time series of satellite data.** • **Can impact the visualization tools.**	• **Develop further existing colocation tools (NPROVS, ICARE, STAMP)** • **Metadata should be well documented to help the collocation.**	**To some extent**
G5.04	**Scientific**	**Subset definition**	**All**	**Lack of pertinent reference subset database (spatial extent, time range, sampling, resolution, variables, etc.).**	• **CCI** • **GEOmon projects** • **CFMIP**	• **Direct comparisons are very complex without prior colocation.** • **Direct comparisons and analyses are complex if reference data sets are not consistent.**	• **Define reference subsets based on user requirements**	**To some extent**

G5.05	**Technical**	**Subset format**	**All**	**Usability of reference database.**	• **CCI** • **CFMIP**	• **Analyses may be impaired if tools cannot run consistently across databases.** • **Tool development may be impaired by format issues and lack of data consistency**	• **Specify subset format using appropriate standards**	**To some extent**
G5.06	**Technical**	**Analysis/Visualization tools**	**All**	**Need for analysis tools to exploit reference database (visualization, intercomparison, statistics, etc.).** **Data sets to analyze are very diverse: time-series / instantaneous, spatially localized / large extent, column integrated / profile.**	• **ICARE multibrowse and associated graphical modules** • **Felyx project** • **NOAA EDGE**	• **Reference data base is of little use of pertinent analysis tools are lacking.** • **Overly complex tools may hinder analysis.**	• **Develop further existing visualization and analysis tools to accommodate data set diversity.**	**To some extent**
G5.07	**Methodology**	**Validation; Quality assurance; Comparisons; Traceability;**	**All**	**Incomplete development and/or application and/or documentation of an unbroken traceability chain of Cal/Val data manipulations for atmospheric ECV validation systems.**	• **D5.1** • **CEOS QA4EO** • **EU FP7 project QA4ECV** • **Traceability chain in Keppens et al., AMT, 2015** •	• **General lack of documentation** • **Missing Quality Indicators in many validation studies.** • **Quality Indicators not always fit for purpose.** • **Incoherent and poorly traceable validation results.** • **Potential impact of ground-based validation not maximized.**	• **Development for several ECVs ongoing in EU FP7 project QA4ECV.** • **Further application in the Multi-TASTE Cal/Val system foreseen in GAIA-CLIM.**	**Key aspects will be implemented in the VO. Synergies with QA4ECV will be sought.**
G5.08	**Metrology**	**Validation; Comparisons; Uncertainties; Error budget;**	**All**	**Missing quantification of additional uncertainties introduced in the comparison results due to differences in (multi-dimensional) sampling and smoothing of atmospheric inhomogeneity.**	• **D5.1, D3.1** • **EU FP6 GEOmon Technical NotesD4.2.1 and D4.2.2 (2008-2011)** • **Lambert et al., ISSI book on Atmospheric Water Vapour, Chap. 9, 2012** • **Verhoelst et al., paper on metrology of ozone data validation, AMTD 2015**	• **Dominates random uncertainty in satellite-ground comparisons for most ECVs.** • **Significant contribution to systematic uncertainty in satellite-to-ground data comparisons.** • **Obstructs the interpretation of comparison results.**	• **Model-based (OSSSMOSE in Multi-TASTE) and statistical studies will address these issues for key ECVs in GAIA-CLIM WP3.** • **Awareness raised through the GAIA-CLIM VO.**	**Yes**

Suggested full text format for accompanying text entries in the underlying WP deliverables

G5.01 Access to data in multiple locations with different data policies and accessibility

Gap Type: *Technical*
Gap Keywords: *Data access*
ECV(s): *All*
Trace (external refs):

Gap Description

All available data for the virtual observatory will come from a range of locations with different data policies and accessibility. The access to some datasets may be restricted or may be difficult to automate. The access to other datasets may be restricted by retrieving and unpacking speed.

Gap Impacts

- Lack of access or low-speed access is incompatible with an interactive web-tool

Gap Remedy

At the first stage, it would be better to restrict platform to a few well identified available data. It could be useful to copy some database locally. The focus will be on the choice on pertinent sample dataset.

References

Data are for example accessible from those platforms:

http://www.gruan.org
http://tccon.ornl.gov/
http://www.ndsc.ncep.noaa.gov/data/

G5.02 Access to data in multiple data formats and structures

Gap Type: *Technical*
Gap Keywords: *Data format and structure*
ECV(s): *All*
Trace (external refs):

Gap Description

The virtual observatory platform will have to access to data in multiple data formats (HDF, NetCDF, BUFR, ASCII, other) and structures (granule vs. global datasets, level 1 vs level 2 data). In particular the granularity of available data may differ between data sources.
The metadata should be as complete as possible.

Gap Impacts

- Too large and useless transfer of datasets.
- Loss of metadata during data conversion using available tools.

Gap Remedy

The virtual observatory can rely on existing data conversion tool (e.g. cdo). There is the need to develop new tools to convert the dataset in the appropriate format. The VO data will have to be sampled at the highest possible level to minimize data transfer. Some effort will have to be made to standardized metadata.

References

http://www.ucar.edu/tools/applications_desc.jsp

G5.03 Lack of universal efficient data management system to collocate observations

Gap Type: *Technical*
Gap Keywords: *Efficient collocation algorithm*
ECV(s): *All*
Trace (external refs):

Gap Description

It is unlikely that a suitable data management system exist to achieve the goal of the WP. A good data management system underlies a good visualization tool.

Gap Impacts

- Impact on capability of visualization tool.

Gap Remedy

The project should focus project resources on this issue. If possible it could rely on existing tools as much as possible.

References

Some tools already exist. They should be analyzed and could serve as a baseline for the development of the virtual observatory. Here are examples of interesting projects/tools:
- Giovanni, interactive visualization and analysis, NASA GSFC, http://disc.sci.gsfc.nasa.gov/giovanni
- CCI toolbox, Community Intercomparison Suite: https://www2.physics.ox.ac.uk/research/climate-processes/projects/cis

G5.04 Lack of pertinent reference subset database

Gap Type: *Scientific*
Gap Keywords: *Subset definition*
ECV(s): *All*
Trace (external refs):

Gap Description

A collocation database must be developed to fulfill the project objectives. The first step is to identify all pertinent EO and ground-based datasets that would be of interest for direct comparison or to support analysis: sensors, variables, resolution if multiple resolutions are available, etc. When applicable, pertinent temporal sampling (i.e. frequency of observation) must be specified (e.g. geostationary measurements, ground-based high frequency observations).

For satellite observations, pertinent spatial extent around colocation sites must be specified. Similarly, for ground-based observations, pertinent time range around satellite overpass time must be specified.

For some variables, further work may be required to make them consistent across sensors or observation site: for example, for aerosol products, retrievals may be available at different wavelengths requiring further calculation to derive a reference variable: e.g. AOD (Aerosol optical depth) at 550 nm. Similarly, some retrieval may provide fine mode AOD and coarse mode AOD, others may provide total AOD and fine mode fraction.

Note: an underlying constraint if the database volume that needs to be of reasonable size to be manageable.

Gap Impacts

- Direct comparisons are very complex without prior colocation.
- Direct comparisons and analyses are complex if reference data sets are not consistent.

Gap Remedy

Define reference subsets based on user requirements

References

CCI and GEOmon projects

G5.05 Usability of reference database

Gap Type: *Technical*
Gap Keywords: *Subset format*
ECV(s): *All*
Trace (external refs):

Gap Description

Once reference measurements and subsets characteristics are defined, the reference subset database must be created. For referencing purposes and traceability, subsets will be created and transferred to one central location. One should take special care when creating a subset from the original data sets, with respect to data integrity and traceability. Although native formats will be preferred, some reformatting may be necessary if native formats are not suitable. In particular, conversion to self-descriptive standard formats (e.g. NetCDF, HDF, etc.) may be required for obvious data management reasons, usability and support by analysis tools. This gap adds to gap G5.02, which relates to the complexity of the source data sets.

Gap Impacts

- Analyses may be impaired if tools cannot run consistently across databases.
- Tool development may be impaired by format issues and lack of data consistency

Gap Remedy

Specify subset format using appropriate standards

References

CCI

G5.06 Analysis/Visualization tools

Gap Type: *Technical*
Gap Keywords: *Analysis tools*
ECV(s): *All*
Trace (external refs):

Gap Description

Analysis tools must be developed to provide graphical outputs, statistics and various indicators that meet user requirements. Special attention must be paid to the specification of graphical representation of individual parameters.
Flexibility will be favored with adequate level of parameterization.
Also need for solution to allow comparison of multiple collocated parameters at the same time to circumvent the complexity of comparing data sets of various nature and various geometries (e.g. time series and instantaneous, spatially localized and large spatial extent observations, column-integrated observations and vertical profiles, etc.)
Need for capability to display multiple reference sites at the same time for regional to global analyses.

Gap Impacts

- Reference data base is of little use of pertinent analysis tools are lacking.
- Overly complex tools may hinder analysis.

Gap Remedy

References

Some tools already exist and could be reused:

e.g. the Felyx project, the primary concept of Felyx is to work as an extraction tool, subsetting source data over predefined target areas (which can be static or moving) : these data subsets, and associated metrics, can then be accessed by users or client applications either as raw files, automatic alerts and reports generated periodically, or through a flexible web interface allowing for statistical analysis and visualization. http://hrdds.ifremer.fr/project/concept

ICARE multibrowse and associated graphical modules:

http://www.icare.univ-lille1.fr/browse/

G5.07 Incomplete development and/or application and/or documentation of an unbroken traceability chain of Cal/Val data manipulations for atmospheric ECV validation systems

Gap Type: *Methodology*

Gap Keywords: *Validation; Quality Assurance; Comparisons; Traceability*

ECV(s): *All*

Trace (external refs): *Keppens et al, AMT 2015; QA4EO and QA4ECV documentation.*

Gap Description

In the context of sustainable Earth Observation data services such as those in development for the Copernicus Climate Change Service (C3S) and Atmospheric Monitoring Service (CAMS), Quality Assurance (QA) and geophysical validation play a key role in enabling users to assess the fitness of available data sets for their purpose. User requirements, e.g., those formulated for the Global Climate Observing System (GCOS), have to be identified and translated into QA and validation requirements; in turn, QA and validation results must be formulated in the form of appropriate Quality Indicators (QI) to check and document the compliance of the data with the user requirements. Metrology practices recommend the development and implementation of traceable end-to-end QA chains, based on Système International d'Unités (SI) and community-agreed standards (as identified for instance in the GEO-CEOS QA4EO framework). Generic guidelines for such QA systems applicable virtually to all atmospheric and land ECVs are being developed within the EU FP7 QA4ECV project (2014-2018), while more specific guidelines developed in projects like ESA's CCI and dedicated to atmospheric ECVs are being published (e.g., Keppens et al., 2015a). Generic and specific QA systems and guidelines established in those recent projects are not known in the global community, where validation purposes, methodologies and results can differ significantly from one report to another. Harmonized practices should now be advertised and applied more universally across the community.

Gap Impacts

The impact of not adopting a traceable end-to-end validation approach is diverse. First, important quality indicators may be missing in the analysis, e.g. information on spatio-temporal coverage, resolution, dependences of the data quality on particular physical parameters (e.g. solar zenith angle, could cover, thermal contrast) etc. Second, results may be incoherent between several validation exercises on the same data set, and the origin of the discrepancies unclear due to insufficient traceability. Third, methodological uncertainties in, e.g., regridding, in the use of vertical averaging kernels, or in unit conversions using auxiliary data, may lead to unreliable results. Finally, all this may imply sub-optimal use of the true validation capabilities of the ground-based reference network.

Gap Remedy

Development of a generic end-to-end QA and validation chain is ongoing for atmospheric ECVs in the EU FP7 QA4ECV project, with application to 3 pilot ECV precursors in QA4ECV and to ozone in ESA's CCI Ozone project. This work needs to be extended to other ECVs, and the implementation of these QA4EO compliant practices must be illustrated (e.g. in the GAIA-CLIM VO) and operationalized (ongoing in the Multi-TASTE Cal/Val system operated at BIRA-IASB, e.g. Keppens et al. 2015b).

References

Quality Assurance framework for Earth Observation, http://www.qa4eo.org
Quality Assurance for Essential Climate Variables, http://www.qa4ecv.eu
Keppens et al., Round-robin evaluation of nadir ozone profile retrievals: Methodology and application to MetOp-A GOME-2, Atmos. Meas. Tech., 8, 2093-2120, 2015a.
Keppens et al., Harmonized Validation System for Tropospheric Ozone and Ozone Profile Retrievals from GOME to the Copernicus Sentinels, ESA's ATMOS2015 conference proceedings, 2015b.

G5.08 Missing quantification of additional uncertainties introduced in the comparison results due to differences in (multi-dimensional) sampling and smoothing of atmospheric inhomogeneity

Gap Type: *Metrology*
Gap Keywords: *Validation; Comparisons; Uncertainties; Error budget*
ECV(s): *All*
Trace (external refs): *D3.1*

Gap Description

When comparing two different measurements of an atmospheric variable, there almost inevitably exists a mismatch in measurement location, time, and smoothing properties (see e.g. Lambert et al. 2012 for the case of water vapour comparisons). As a result, spatio-temporal atmospheric variability and structures will impact these comparisons and introduce additional errors, not accounted for by the (instrumental and retrieval) uncertainties reported with the data. To be able to draw meaningful conclusions from the comparisons, these additional errors must either be minimized to well below the measurement uncertainties with the use of specific co-location criteria, or they must be reliably quantified. In practice, to obtain a sufficiently large number of co-located pairs to derive meaningful statistics, most adopted co-location criteria result in a significant contribution from natural variability to the comparison error budget. This contribution is only rarely quantified because of several hampering issues (cf. gaps G3.1 – G3.6).

Gap Impacts

For many ECVs, such as T, q and ozone, atmospheric variability is known to contribute significantly to the spread of the differences between satellite and ground-based measurements, even when using tight co-location criteria (e.g. Ridolfi et al. 2007, Cortesi et al. 2007, Fassò et al. 2014). Verhoelst et al. (2015) show that even the mean (or median) difference can be affected by atmospheric variability. Consequently, a reliable interpretation of the comparison results is impossible without a quantified understanding of the uncertainties due to differences in sampling and smoothing of the variable and inhomogeneous atmosphere.

Gap Remedy

Remedies to this gap are described in detail in gaps G3.1-G3.6 in the GAID. Significant work in this direction will be undertaken within GAIA-CLIM, in particular within WP3, where the comparison uncertainties due to differences in sampling and smoothing are quantified either using Observing System Simulation Experiments with the OSSSMOSE metrology simulator, or using advanced statistical modelling tools. The Virtual Observatory in WP5 will serve as a demonstrator of potential approaches in dealing with this issue in comparisons between selected satellite ground-based instruments, at the same time raising awareness of these issues with the user base.

References

Cortesi et al., Geophysical validation of MIPAS-ENVISAT operational ozone data, Atmos. Chem. Phys., 7, 4807-4867, 2007
Fasso et al., Statistical modelling of collocation uncertainty in atmospheric thermodynamic profiles, Atmos. Meas. Tech., 7, 1803-1816, 2014
Lambert et al., Ground-based remote sensing and in-situ methods for monitoring atmospheric water vapour – Chapter 9: Comparing and merging water vapour observations: A multi-dimensional perspective on smoothing and sampling issues, pp. 177-199, ISSI, 2012
Ridolfi et al., Geophysical validation of temperature retrieved by the ESA processor from MIPAS/ENVISAT atmospheric limb-emission measurements, Atmos. Chem. Phys., 7, 4459-4487, 2007

Verhoelst et al., “Metrology of ground-based satellite validation: Co-location mismatch and smoothing issues of total ozone comparisons”, accepted for publication in AMTD, 2015