Ref. Ares(2016)3285507 - 08/07/2016

GAIA-CLIM Report / Deliverable D5.2

Gap Analysis for Integrated Atmospheric ECV Climate Monitoring:

Update for GAIA-CLIM Gaps Assessment and Impacts Document from WP5

A Horizon 2020 project;
Grant agreement: 640276
Date: 8th July 2016
Lead Beneficiary: EUMETSAT
Nature: R
Dissemination level: PU

Work-package	**WP5 (Creation of a virtual observatory visualisation and data access facility)**
Deliverable	D5.2
Nature	R
Dissemination	PU
Lead Beneficiary	EUMETSAT
Date	31/07/2016
Status	Final
Authors	Arndt Meier (EUMETSAT), Jacques Descloitres (University of Lille), Marie Doutriaux Boucher (EUMETSAT), Kalev Rannat (Tallinn Technical University), Kuldar Taveter (Tallinn Technical University), Tijl Verhoelst (IASB-BIRA)
Editors	Jörg Schulz (EUMETSAT)
Reviewers	Peter Thorne (NUIM), Corinne Voces (NUIM), Richard Davy (NERSC)
Contacts	Marie.Doutriaux-Boucher@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

Document history

Version	Author(s) / Reviewers	Date	Changes
0	Peter Thorne, Michiel van Weele	13/05/16	Document template provision to all WP leads.
1.0	Arndt Meier, Jacques Descloitres, Marie Doutriaux-Boucher, Kalev Rannat, Kuldar Taveter, Jörg Schulz	06/07/2016	Updated and new entries for WP5.
1.1	Jörg Schulz	08/07/2016	Final edits following reviewers comments.

Table of Contents

Executive Summary

This deliverable constitutes further input from WP5 (Creation of a virtual observatory visualisation and data access facility) to the drafting of the living Gaps Assessment and Impacts Document (GAID) of Task 6.2 (WP6), led by KNMI. The purpose of the GAID is to collate and document gaps directly relevant to the aims of the GAIA-CLIM project. The GAIA-CLIM project is concerned with increasing the utility, use and value of non-satellite observations to characterise satellite observations. Further project details are available at www.gaia-clim.eu.

This deliverable refers to the second official release of the GAID (D6.4) and builds upon the gaps identified therein. In addition, it arises new gaps that have been identified in relation to the Work Package activities.

WP5 develops a so called Virtual Observatory (VO) that enhances the exploitation of ground-based reference data for satellite sensor and product validation through organising access to data and comparison results. The VO integrates ground-based reference and satellite data, and the methodologies to assess their uncertainty including mismatch uncertainties, with existing satellite-satellite data comparisons and observation feedback from NWP models and reanalysis. It allows users to interrogate multiple data sources in a seamless way and permits remote data analysis. By doing this the VO shall increase awareness among users of satellite and non-satellite data of the concept of traceable uncertainty estimates and provides a facility that may support Copernicus Services in the future for analysis of product quality in a sustainable routine mode.

This deliverable further expands upon the gaps identified in the initial work package input, relevant gaps sourced externally, and new gaps that have been identified by participants. The gaps discussed herein are exclusively those related to the WP aims and remit (see prior paragraph). A key focus of the current iteration is to make the gaps and their remedies more SMART (Specific, Measurable, Actionable, Relevant and Timebound) with realistic cost estimates and assessments of the risk / cost of leaving the gap unremedied. In year 3 the GAID shall inform the development of a list of prioritised recommendations and this shift in emphasis is expected to help inform such an exercise.

1. Document rationale and broader context

The purpose of this document is to provide input to the Gaps Assessment and Impacts Document (GAID) of the GAIA-CLIM project arising from WP5. This WP is concerned with providing a so called Virtual observatory (VO): a service available via the internet that allows exploitation of comparisons of ground-based reference data with corresponding satellite measurements. To assure that results of such comparisons are useful to characterise the quality of satellite measurements, all inputs are accompanied by traceable uncertainty estimates and the uncertainty estimates originating from the conditions under which the comparison is undertaken are also included to fully characterise the uncertainty budget of the comparison. As meaningful uncertainties are not always easy to derive for geophysical parameters retrieved from satellite measurements, the VO also uses observation operators that transfer non-satellite reference measurements into the satellite measurement space to allow a more meaningful comparison.

The VO provides access to all information resulting from such comparisons, including the input data with their meta-data, and output statistics and visual representations of the comparison results. The inputs to the VO in the form of data, meta-data and the methodology to estimate the uncertainty are provided by the other underlying work packages of the GAIA-CLIM project. In addition to the comparisons, the VO also offers the functionality to display the meta-data of non-satellite reference data networks.

The establishment of the VO is addressing an important required function for climate research and services by providing an extendable service for the characterisation of satellite data that does not presently exist. The integration of state-of-the-art knowledge on uncertainty characterisation with modern technical standards to provide access to the results has the potential to accelerate research and to broaden the ability of the scientific community to analyse data from various sources in a uniform framework.

The development of the VO is steadily progressing and major elements such as data-bases for meta-data and comparison data, as well as tools to transfer reference data into satellite measurement space, have been developed. The interface for the user is still under development and will be presented in its first version at the next GAIA-CLIM user workshop in November 2016.

The GAID has now gone through 2 iterations. The first iteration was based upon a combination of the user survey and individual inputs from this and the four remaining underlying Work Packages. The second iteration built upon this by incorporating feedback from the first user workshop and additional informal input delivered from this and other Work Packages. The third version shall build upon the second by considering input arising from this current set of deliverables. That version shall be discussed at the second user workshop to be held in Brussels in November 2016 and the input received shall lead to a further iteration, which shall form the initial basis for a set of prioritised recommendations arising from Task 6.3.

Feedback from the science advisory panel, the first General Assembly, and the review pointed collectively to the need to evolve the GAID to go beyond characterising the gap to considering in more detail implications, potential SMART remedies, costs, and the benefits of resolving them. This then shall help allow external and internal users to more fully explore and appreciate the gaps identified prior to work by Task 6.3 to collate a set of prioritised recommendations.

2. Summary of gaps from GAID v2 relevant to the current WP

The gaps identified in the GAID that shall be considered in further detail in Section 4 are summarised below. This is a direct subset of relevant entries from Table 2.2 of the version 2 release of the GAID. These gaps arose from either the initial Deliverable from this WP (D5.1) or from subsequent external input.

Table 1: List of gaps identified in the first two versions of the GAID identified as being in the sphere of WP5. Some of the descriptions have been updated as described and justified in the accompanying text.

Gap Identifier	Gap Type	ECV(s)	Gap Short Description	Trace
G5.01	Technical	H_2O, O_3, T, CO_2, CH_4, aerosols	Access to data in multiple locations with different user interfaces constitutes a barrier to usage and makes use in general difficult.	Examples of distributed access locations for reference data are: http://www.gruan.org, http://tccon.ornl.gov/,http://www.ndsc.ncep.noaa.gov/data/
G5.02	Technical	H_2O, O_3, T, CO_2, CH_4, aerosols	Access to and use of reference and satellite data provided in different data formats and structures (e.g. granularity of data) prevents easy exploitation.	Communities serving satellite and non satellite data are using a large variety of formats and structures.
G5.03	Technical	H_2O, O_3, T, CO_2, CH_4, aerosols	No common source for co-located data exists which prevents use of reference data to validate reference measurements to each other and to evaluate satellite data.	Many individual sources exist but only very few provide partly access to comparisons including uncertainties.
G5.06	Technical	H_2O, O_3, T, CO_2, CH_4, aerosols	Extraction, analysis and visualization tools to exploit the potential of reference measurements are currently only rudimentarily. This in particular includes tools to display uncertainty of the comparison results due to differences in sampling and so called smoothing error.	The required tools exist in a scattered way for individual measurement systems but are not integrated.
G5.07	Technical	H_2O, O_3, T, CO_2, CH_4, aerosols	Incomplete development and/or application and/or documentation of an unbroken traceability chain of Cal/Val data manipulations for atmospheric ECV validation systems prevents progress in the characterization of satellite products.	D5.1 of GAIA-CLIM Traceability chains as developed in the FP7 QA4ECV project (http://www.qa4ecv.eu/); General guidelines as provided by QA4EO (http://qa4eo.org/).

A number of gaps associated with the present Work Package identified in version 2 of the GAID have been retired and/or rationalised in this input to version 3 of the GAID. The retired gaps were previously numbered G5.04, G5.05 and G5.08, and retain these numbers, although they are no longer considered part of this contribution to the GAID. In addition, gaps G5.01-G5.03, G5.06 and G5.07 have been revisited to make their formulation more consistent and actionable. The rationale for retired and modified gaps is outlined below.

G5.01: This gap has been reduced to the technical aspect. The governance aspect addressing potential data policy issues for the reference data providing networks will now be addressed as a separate gap of type Governance under WP6;

G5.02: The formulation of this gap has been adapted to represent a gap and not an action. In addition gaps G5.04 (Usability of reference data needs to be improved: high functionality in subset selection) and G5.05 (Usability of reference data needs to be improved: format) have been integrated into G5.02 as they represent just sub categories of the refined gap.

G5.03 and G5.06: Both gaps have been revised to become actionable.

G5.04 and G5.05: Both gaps have been retired because of the inclusion into gap G5.02.

G5.07: The gap type characterisation Governance has been removed from the gap and the formulation has been changed to reflect a gap not an activity.

G5.08: This gap has been retired as it entirely overlaps with G3.06 that is kept by WP3.

3. New gaps identified by WP participants to date

Subsequent to the first official input to the GAID (D5.1), substantial work has been undertaken upon the Work Package in the following respects:

- Outcomes of the User Survey and 1st User Workshop have been analysed for the design of the VO with respect to all technical categories describing access to data and specific functionality of the VO in terms of data analysis and visualisation;
- The analysis results have been used to define use cases that allow a modular development of the VO;
- The readiness of uncertainty estimates for reference data and the ability for determination of smoothing errors has been analysed for all envisaged ECVs of GAIA-CLIM in collaboration with WP1, WP2, WP3 and WP4. The result has been used to derive an order for implementation of specific comparison configurations (reference vs. satellite measurements) for the ECVs, e.g., starting with ozone total column estimates, followed by water vapour and aerosol measurements;
- The architecture of the VO has been finalised after thorough analysis of existing similar solutions at ICARE and NOAA. The architecture is now under development and implementation;
- Core elements such as the data-bases for meta-data (in collaboration with WP1) and co-locations have been developed and implemented;
- An offline radiative transfer capability for GRUAN radiosonde data has been developed in collaboration with WP4. This allows comparison to a large fleet of satellite instruments (infrared and microwave spectral range) in their measurement space;
- A Graphical User Interface has been sketched and development started ;
- The technical configuration of the computer system has been developed and deployed.

In addition to the project specific activities, further consultation with the Copernicus Climate Change Service has taken place that is seen as a potential operational implementation target for the VO.

These activities, in addition to advancing the aims of the GAIA-CLIM project, have given cause to reflect further on potential gaps in our collective knowledge and capabilities. This has led to additional gaps being identified. These gaps will be further expanded in Section 4 and are summarised in Table 2.

Table 2: List of new gaps identified following the user survey, 1st user workshop and conversations with Copernicus Services.

Gap Identifier	Gap Type	ECV(s)	Gap Short Description	Trace
G5.09	Technical	H_2O and T	A readily accessible online tool to perform radiative transfer calculations to transfer reference measurements of ECVs, including their uncertainty estimates, into the space measurement space is missing to enable characterisation of satellite measurements	GAIA-CLIM User Survey and 1st GAIA-CLIM User Workshop
G5.10	Technical	All	Characterisation of different types of uncertainty has not been systematically addressed per ECV, e.g., for some ECVs a full measurement uncertainty chain has been established, but smoothing uncertainty has not been considered or vice-versa. This prevents and potentially delays inclusion of various instrument/ECV combinations into the Virtual Observatory.	GAIA-CLIM Reference Observation Readiness Table
GS5.11	Technical	All but GRUAN and EUMETSAT	Non operational provision of reference measurement data and some (L2) satellite products may prevent use in Copernicus operational product monitoring.	Evident from currently offered reference measurement, and some satellite, data products provisions and the needs of an operational service.

4. Detailed update on traces for the gaps arising from this Work Package for inclusion in the GAID

Within this section gaps that were detailed in Sections 2 and 3 are expanded to give a full trace of our current understanding of the gap, its impacts and its potential remedies. For those gaps identified in Section 2 we take as the starting point the corresponding text arising from the GAID (v2, Section 3) text and / or the initial deliverable text as we deem most appropriate. This is then expanded upon here in an attempt to better delineate the gap, its impacts, its potential remedies (including indicative costs and timelines) and the scientific impact of (non-)resolution. Gaps are ordered numerically and each given a specific subsection.

4.1 G5.01: Access to data in multiple locations with different user interfaces constitutes a barrier to usage and makes use in general difficult

Gap detailed description

The task of characterizing satellite measurements by means of comparison to reference measurements needs access to and documentation of various reference measurements needed for the analysis of the quality of satellite measurements and/or derived geophysical data products. This task can be massively complicated and time consuming arising from the need to collect data from multiple locations also often offering the data on various types of user interfaces, with which a user needs to become familiar. In many cases, data downloads do not follow specific data exchange standards, which makes it difficult to automate access to them. In addition, the available bandwidth at the provider side might be too small to serve many customers, which can result in extended waiting times for the data. This applies even more when co-located ground based and satellite data are to be offered to the user.

Activities within GAIA-CLIM related to this gap

WP5 activities address this gap partly but not comprehensively with the VO.

Gap remedy(s)

A general remedy of the gap would be to make data available through operational exploitation platforms which could be developed for different user communities. GAIA-CLIM provides this as part of the VO for a set of atmospheric ECVs and the specific application of characterizing satellite measurements. Many other ECV reference measurements – satellite data combinations are outside the scope of the GAIA-CLIM project and will not be addressed in this project.

Remedy #1

<u>Specific remedy proposed</u>

The considered ECVs and combinations with satellite data that will be implemented into the Virtual Observatory will become accessible, with the VO acting as a central hub for data access.

Measurable outcome of success

Success could be measured by data download statistics and how they evolve with time. The latter cannot be measured within the time frame of GAIA-CLIM.

Achievable outcomes

Technological viability: High

Indicative cost estimate: High (>5 million) for delivering exploitation platforms and low (<1 million) for achieving a demonstrator level as planned for GAIA-CLIM.

Relevance

The remedy would set an example of how data can be served in the context of the GAIA-CLIM project.

Timebound

The demonstrator takes 2 years, more complete exploitation platforms longer than 5 years.

Gap risks to non-resolution

Identified future risk / impact	**Probability of occurrence if gap not remedied**	**Downstream impacts on ability to deliver high quality services to science / industry / society**
The use of multiple locations with different set ups for data access continues to complicate work on data comparison and increases cost to delivery and analysis / exploitation of data.	High	Impact depends on diversifications of locations from which data are served.

4.2 G5.02: Access to and use of reference and satellite data provided in different data formats and structures (e.g. granularity of data) prevents easy exploitation

Gap detailed description

The comparison of satellite data and reference measurements is further complicated through the fact that data are provided in multiple data formats, e.g., HDF, NetCDF, BUFR, ASCII, etc, and in different structures (granules vs. global datasets, level 1 vs. level 2 data). In particular, the granularity of available data may differ between data sources. The use of such data is complex as the inclusion into a common data base that allows geographical and temporal sub-setting and the reliable use of data analysis tools requires format conversion modules for each format used on the input side. Format conversions always bring with them the danger of destroying information, in particular in the accompanying meta-data. For instance data flag definitions are often coded in meta-data and during format conversion they are not correctly transferred due to bugs in the conversion software, which can render flags in the data useless as they cannot be interpreted anymore.

Different granularity of the data creates work to collect and resample data until they represent the same area and time. Then, to perform a comparison, data need to be co-located using specific criteria. For this aspect, not having access to the highest temporal resolution for the reference data can really hamper the comparison, e.g., if they cannot be brought close to the satellite measurement in a meaningful way. For instance, if we have a vertical profile at one place that should be compared to a snapshot from a satellite with a certain geographical coverage and spatial resolution, one needs the reference data to be available at the highest possible frequency in order to average over timescales representative for the spatial variability, as seen by the satellite.

Work to achieve correct co-locations under the described conditions are repeated by users many times, which is a gross redundancy in effort and prone to processing errors.

Activities within GAIA-CLIM related to this gap

WP5 activities support the remedy of this gap by providing data format conversion tools for various input data and a data extraction function that makes the outputs available in user friendly formats.

Gap remedy

Develop a Virtual Observatory that converts all input data formats into a data base and also adapts the granularity of input data to ensure efficient use of the data in reference measurement and satellite data comparisons.

Remedy #1

Specific remedy proposed

The inputs to the VO must be reformatted which can partly rely on existing data conversion tools (e.g. cdo), but for some there is the need to develop new tools to convert the datasets into the appropriate format. Some effort will have to be made to standardise metadata. The granularity of the input data needs to be adjusted to support comparisons from different data-bases. Outputs need to be made available in 'most wanted' formats as indicated in the recent GAIA-CLIM user survey that clearly indicated the usefulness of NetCDF to target users.

Measurable outcome of success

Success can be measured by assessing if for all available data formats readily available conversion tools exist that allow integration into the GAIA-CLIM VO.

Achievable outcomes

Technological viability: High

Indicative cost estimate: low (<1 million)

Relevance

The proposed remedy will help to avoid the repetition of work for format conversions and conversions of data provided with different granularity into data sets that can be compared.

Timebound

The envisaged demonstrator developments are within the timebounds of the GAIA-CLIM project.

Gap risks to non-resolution

Identified future risk / impact	Probability of occurrence if gap not remedied	Downstream impacts on ability to deliver high quality services to science / industry / society
Continued need for data format conversion tools that are established by many different groups.	High	General higher cost for data handling before achieving results.

4.3 G5.03: No common source for co-located data exists which prevents use of reference data to validate reference measurements to each other and to evaluate satellite data.

Gap detailed description

Several sources for co-located data sets exist but most of them are specialized to compare mapped fields, e.g., obs4mips ESG data provisions, reference and other non-satellite data against models or satellite data, e.g., the NORS project. But most of these are not fully utilizing the potentially available information on uncertainty or including uncertainty arising from spatiotemporal mismatch of the compared data streams. Some of the existing datasets are publically available via the internet, while others are run internally to organizations like EUMETSAT to monitor data quality in real time. The effect of this gap is that many validation activities are performed, but do not use the available uncertainty information in an optimal way which has general effects on the quality of the research and the robustness of any conclusions drawn from such validation exercises.

A common source that integrates several reference data networks with satellite data considering traceable uncertainty does not exist but is needed according to the GAIA-CLIM user survey.

Activities within GAIA-CLIM related to this gap

WP5 develops a Virtual Observatory that addresses this gap.

Gap remedy(s)

The Virtual Observatory shall be developed to demonstrate the use of non-satellite reference data and NWP model data for the characterisation of satellite data. The Virtual Observatory must integrate the different measurements, their metadata, quantified uncertainty for the measurements, and the uncertainty arising from the comparison process. The major part of the VO is a co-location data base that enables various scenarios for comparison of the satellite and non-satellite data.

Remedy #1

Specific remedy proposed

As a major part of the VO a co-location data base must be developed to establish a foundation for the remedy of this gap. The first step is to identify all pertinent satellite and non-satellite reference datasets that are of interest for a comparison to a given satellite sensor data. This could either be via a forward modeling approach to derive an estimate of the satellite sensor data or a comparison to geophysical variables derived from the satellite data. The provided data need to be complemented by as complete as possible meta-data and traceable uncertainty information, including comparison mismatch uncertainties that need to be derived from the comparison setting and the variability of the geophysical variable to be compared.

For satellite measurements, suitable spatial extent around reference measurement sites must be specified. Similarly, for non-satellite measurements, pertinent time range around satellite overpass time must be specified to allow for a meaningful comparison.

Many satellite measurements contain information from different geophysical variables, e.g., the separation of temperature and humidity signals in a retrieval scheme working in the infrared spectral range is very complex. Instead, the co-located reference measurements, for this example containing temperature and humidity profiles, can be used to simulate the expected satellite measurement which allows a more meaningful comparison. Such a 'processor' is addressed in the remedy of gap G4.01 in WP4.

For some variables, further work may be required to make them consistent across satellite sensors or observation sites. For instance, 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.

A technical constraint for a co-location data base is its volume that needs to of reasonable size to remain manageable.

Measurable outcome of success

Established co-location database that is accessible for extraction, analysis, visualisation tools and allows for interrogation by users via a user interface.

Achievable outcomes

Technological viability: High

Indicative cost estimate: High (>5 million) for a database covering ECVs beyond the GAIA-CLIM ECV set; low (<1 million) for demonstrating the general capability as planned in GAIA-CLIM.

Relevance

The remedy proposed here is in full agreement with the results of the GAIA-CLIM User Survey and the results of the 1st GAIA-CLIM User Workshop.

Timebound

The remedy proposed here is a key focus, and deliverable of GAIA-CLIM WP5 due for delivery (D5.3, D5.4) in months 24 and 30, respectively.

Gap risks to non-resolution

Identified future risk / impact	Probability of occurrence if gap not remedied	Downstream impacts on ability to deliver high quality services to science / industry / society
Non-satellite reference measurements will have limited value for the characterisation of satellite measurements which can have negative effects for the funding of the networks.	High	Potential for full characterisation of the quality of satellite measurements is not realised with potential negative impacts on instrument developments in the future.

4.4 G5.06: Extraction, analysis and visualization tools to exploit the potential of reference measurements are currently only rudimentary. This in particular includes tools to display uncertainty of the comparison results due to differences in sampling and so called smoothing error.

Gap detailed description

Services that provide data extraction, analysis and visualization tools are currently only rudimentary. In particular, analysis capabilities that for instance allow analysis at different time or spatial scales are missing. The analysis capability provides a challenge to such services as no widely used tools for the comparison of satellite and non-satellite reference measurements exist presently. The GAIA-CLIM User Survey indicated a clear need for such a capability to be developed, but challenges remain because whatever analysis / visualization tool can be provided it will not necessarily match all individual needs. The GAIA-CLIM User Survey also indicated that the analysis of the co-locations provided by the Virtual Observatory may not solely be used to evaluate satellite measurements but also vice-versa the satellite measurements may be used to evaluate the quality of the reference measurements, e.g., their temporal consistency. Such a flexible tool does not exist to date.

While measurement uncertainties are at least displayed by some existing services, e.g., the FP7 NORS project, the visualization of uncertainty arising from differences in spatiotemporal sampling is generally not included, but is needed to fully understand the uncertainty budget of a specific comparison.

The impact of this gap is that a co-location data base without a comprehensive capability for data extraction, visualization and data analysis is of limited value to the scientific community and also for a potential later use in Copernicus Services.

Activities within GAIA-CLIM related to this gap

WP5 develops a Virtual Observatory that addresses this gap.

Gap remedy(s)

The Virtual Observatory shall demonstrate state-of-the-art approaches to the use of non-satellite reference data and NWP model data for the characterisation of satellite data by providing a comprehensive data extraction and visualisation capability together with an analysis capability for a number of static case studies.

Remedy #1

Specific remedy proposed

A data extraction capability can be developed that allows the export of data from the Virtual Observatory in user friendly formats. The GAIA-CLIM User Survey indicated a clear preference for the self-descriptive NetCDF format which also allows comparison data being amended by meta-data of the comparison, e.g., the used co-location criteria, etc. Such format also supports analysis of the data that may not be enabled, at least initially, in the Virtual Observatory. Data extraction tools also shall be capable of sub-setting each data source contained in the co-location data base by ECV, time and location, observing system and other boundary conditions such as surface type.

To exploit the co-location data base proposed as a remedy for gap G5.03, analysis tools must be developed to provide statistics and various indicators for a comparison that meet user needs as indicated by the GAIA-CLIM User Survey outcomes. These analysis tools must have some flexibility, such as interchanging the reference in a comparison and the ability to perform analysis at different time and eventually space scales.

Visualization tools need to be capable of displaying multiple collocated parameters to circumvent the complexity of comparing datasets of varying type and geometries, e.g. time series and instantaneous, spatially localized and large spatial extent observations, column-integrated observations and vertical profiles, etc. Special attention must be paid to the specification of graphical representation of individual parameters and various uncertainty measures, including the smoothing uncertainty.

Tool development shall benefit from existing elements and capabilities whenever possible. All developed tools need to be accessible via a graphical user interface that also needs to be developed.

Measurable outcome of success

Developed tools being accessible via a graphical user interface.

Achievable outcomes

Technological viability: High

Indicative cost estimate: High (>5 million) for a database covering ECVs beyond the GAIA-CLIM ECV set; low (<1 million) for demonstrating the general capability as planned in GAIA-CLIM.

Relevance

The remedy proposed here is in full agreement with the results of the GAIA-CLIM User Survey and the results of the 1st GAIA-CLIM User Workshop.

Timebound

The remedy proposed here is a key focus and deliverable of GAIA-CLIM WP5 due for delivery (D5.3, D5.4) in months 24 and 30, respectively.

Gap risks to non-resolution

Identified future risk / impact	**Probability of occurrence if gap not remedied**	**Downstream impacts on ability to deliver high quality services to science / industry / society**
Lack of the described tools prevents optimal use of reference measurements leading to potential issues with the justification of the measurements in the future.	High	Derived global products from satellite may suffer in quality from inadequate evaluation of the measurements and retrieval schemes used to generate them. This can hamper applications supporting decision and policy making.

4.5 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 prevents progress in the characterization of satellite products.

Gap detailed 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 sufficiently well recognized or understood 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.

The impact of not adopting a traceable end-to-end validation approach is diverse. Firstly, 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. Secondly, results may be incoherent between several validation exercises on the same data set, and the origin of the discrepancies unclear due to insufficient traceability. Thirdly, methodological uncertainties in, e.g., geographical mapping, in the use of vertically 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.

Activities within GAIA-CLIM related to this gap

The GAIA-CLIM project adds to other EU projects with respect to more ECVs and disseminates results via the Virtual Observatory.

Gap remedy(s)

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).

Remedy #1

Specific remedy proposed

The science for the end-to-end QA and validation chain is developed in several WPs in GAIA-CLIM. The Virtual Observatory can serve as a tool making the practises visible to users of the comparison data and results. In addition to remedies proposed for the gaps G5.03 and G5.06 described above the Virtual Observatory can present the traceability chains derived and also provide tutorials describing the methods used.

Measurable outcome of success

Full documentation available in the Virtual Observatory; and demonstrated use by VO users.

Achievable outcomes

Technological and organizational viability: High

Indicative cost estimate: Low (<1 million)

Relevance

The proposed remedy is fully in line with procedures established by QA4EO and provides a step forward to the currently available documentation.

Timebound

The remedy proposed here is a key focus and deliverable of GAIA-CLIM WP5 due for delivery (D5.3, D5.4 and D5.5) in months 24, 30 and 33, respectively.

Gap risks to non-resolution

Identified future risk / impact	**Probability of occurrence if gap not remedied**	**Downstream impacts on ability to deliver high quality services to science / industry / society**
Non availability of the documentation of the full traceability chains and methods may result in limited understanding of quantified uncertainties in data set comparisons in the larger science and service community.	Medium	Limited understanding of the QA and validation chain may prevent the optimal choice of products in science and services.

4.6 G5.09: A readily accessible online tool to perform radiative transfer calculations to transfer reference measurements of ECVs including their uncertainty estimates into the space measurement space is missing to enable characterisation of satellite measurements.

Gap detailed description

The GAIA-CLIM User Survey highlighted the need to have an online radiative transfer capability available in the planned Virtual Observatory to allow the transfer of reference measurements into the measurement space of satellite instruments. The impact of not comparing in measurement space is the need for uncertainty estimates for one or several retrieved geophysical parameters, which is usually more complex compared to assessing the uncertainty of a measured satellite count or radiance. This is because the forward calculation from the geophysical profile is unique, whereas the inverse calculation is non-unique in that several distinct geophysical profiles can be satisfied by a single radiative measurement.

Activities within GAIA-CLIM related to this gap

Gap remedy(s)

Implement a forward radiative transfer capability into the Virtual Observatory.

Remedy #1

Specific remedy proposed

Remedy for Gap4.01 (Lack of traceable uncertainty estimates for NWP and reanalysis fields & equivalent TOA radiances – relating to temperature) proposes the development of the GRUAN processor that is able to simulate measurements for many satellite instruments operating in the infrared and microwave spectral ranges consistent with GRUAN profile measures and their uncertainties. Here, it is proposed to integrate the GRUAN processor into the Virtual Observatory and make it online accessible to create simulated measurements for any satellite instrument for which co-locations with the GRUAN reference measurements exist in the Virtual Observatory database.

Measurable outcome of success

Online radiative transfer capability available as part of the Virtual Observatory.

Achievable outcomes

Technological viability: High

Indicative cost estimate: Low (<1 million)

Relevance

Implementing the proposed remedy would help to satisfy a clear user need expressed in the GAIA-CLIM User Survey.

Timebound

The remedy proposed here is part of deliverables of GAIA-CLIM WP5 due for delivery (D5.4) in month 30.

Gap risks to non-resolution

Identified future risk / impact	Probability of occurrence if gap not remedied	Downstream impacts on ability to deliver high quality services to science / industry / society
Limited uptake of Virtual Observatory as comparisons not possible at level 1b radiance space	Medium	Value of reference quality measurements for satellite characterisation not realised.

4.7 G5.10: Characterisation of different types of uncertainty has not been systematically addressed per ECV, e.g., for some ECVs a full measurement uncertainty chain has been established, but smoothing uncertainty has not been considered or vice versa. This prevents and potentially delays inclusion of various instrument/ECV combinations into the Virtual Observatory.

Gap detailed description

The development work of the Virtual Observatory was addressing the selection of reference data, available uncertainty estimates (measurements and smoothing) and the satellite data that shall be characterized. This exercise revealed that certain different types of uncertainty are not addressed systematically resulting in some cases in reference measurements that have quantified uncertainty but for which no means exist to address smoothing uncertainties and vice-versa. This leads to delays in integrating the full QA and validation chain into the Virtual Observatory. It can be expected that for other ECVs in atmospheric but also oceanic and terrestrial domains similar issues exist.

Activities within GAIA-CLIM related to this gap

For the ECVs dealt within GAIA-CLIM this issue is mitigated by focussing on a specific set of non-satellite reference measurements and satellite products and working out the full chain for them.

Gap remedy(s)

Gap remedy at full scale of the problem would be to assess for each GCOS ECV what elements of the uncertainty budget are quantified for pairs of reference measurements and satellite measurements/products and how the missing elements can be achieved. For the ECVs considered in GAIA-CLIM this has been completed.

Remedy #1

Specific remedy proposed

For the full scale problem international bodies such as the CEOS Cal/Val and WG Climate and the GCOS Secretariat should be involved to assess the status throughout the domains and to report to WMO and agency principals about the resulting issues. For the CEOS WG Climate cycles of updating ECV climate data record Inventories could be used to assess the quality of the validation activities in more depth than has been the case to date.

Measurable outcome of success

Knowledge of the status of uncertainty budget estimates for all GCOS ECVs.

Achievable outcomes

Technological viability: High

Organizational viability: medium to low

Indicative cost estimate: Low (<1 million)

Relevance

Implementing the proposed remedy would help to create awareness of the issue and if successful to make the use of more reference measurements and satellite products possible in a Virtual Observatory setting.

Timebound

Status assessment would take approximately 5 years including finding international acceptance for the action and executing it technically.

Remedy #2

For the establishment of the Virtual Observatory a systematic assessment of the considered ECVs is needed (and has partially been done) with respect to the available uncertainty budget elements. In order to find practical solutions it needs to be considered what is the practical importance of smoothing uncertainty for different ECVs (having variability at different space time scales) in addition of different parts of the electromagnetic spectrum. For instance, smoothing uncertainty is of large importance in the visible spectrum but may be of negligible size in the microwave spectrum.

Measurable outcome of success

Successful implementation of the complete QA and validation chains in the Virtual Observatory.

Achievable outcomes

Technological viability: High

Indicative cost estimate: Low (<1 million)

Relevance

The remedy proposed here is in full agreement with the project plan of GAIA-CLIM and is key to success.

Timebound

The remedy proposed here is part of deliverables of GAIA-CLIM WP5 due (D5.3, D5.4 and D5.5) in months 24, 30 and 33, respectively.

Gap risks to non-resolution

Identified future risk / impact	Probability of occurrence if gap not remedied	Downstream impacts on ability to deliver high quality services to science / industry / society
If no assessment of the status of uncertainty budgets for ECVs would be undertaken uncertainties would continue not to be systematically addressed.	Medium	Continued uncertainty about the quality of satellite products for many ECVs used in service relevant applications.
If the implementation in the VO is failing the concept would need to be revisited.	Low	Satellite products used in science and services will have only limited characterisation of uncertainty in

4.7 G5.11: Non operational provision of reference measurement data and some (L2) satellite products may prevent use in Copernicus operational product monitoring

Gap detailed description

Currently, some reference measurements are provided with specific delays due to requirements for certain quality control measures to be applied. The usage scenario for a Virtual Observatory within a Copernicus Service would likely need a close to real time availability of such data to enable the assessment of very recent satellite data products and the close to real time performed reanalysis. If the quality analysis and data provision cannot be operationalized leading to faster delivery, quality assessment at short time scales shall remain of limited nature reducing the value of the data for applications.

Activities within GAIA-CLIM related to this gap

None

Gap remedy(s)

Operationalize processing and delivery for non-satellite reference measurements.

Remedy #1

Specific remedy proposed

A first step would be to assess the current procedures for quality control and delivery mechanism, and to work out a proposal to further automate them. Depending on the needs, specific projects could be established to operationalize the processes and associated software. The dissemination of such data could be included into operational dissemination mechanisms used for operational data provisions such as over the WMO Information System.

Measurable outcome of success

Close to real time availability of non-satellite reference measurements to Copernicus Services.

Achievable outcomes

Technological viability: High

Organizational viability: medium

Indicative cost estimate: High (>5 million)

Relevance

The remedy would significantly increase the use of non-satellite reference data in Copernicus Services.

Timebound

Approximately 5-7 years.

Gap risks to non-resolution

Identified future risk / impact	Probability of occurrence if gap not remedied	Downstream impacts on ability to deliver high quality services to science / industry / society
If the remedy is not started, the use of non satellite reference measurements remains limited	Medium	Quality analysis for time critical services of Copernicus could be significantly increased by providing reference measurements closer to real time.

5. Summary

The input to the Gap Analysis and Impacts Document to date from Work Package 5 comprises gaps identified at the outset of the project. Some gaps have been revisited since the previous iteration of the GAID to make their formulation more consistent and actionable. Those high level gaps defined the work plan for GAIA-CLIM WP5 during Year 1 and have guided the development of the Virtual Observatory.

Several new gaps have been identified after the GAIA-CLIM User Survey and as a consequence of the work carried out during Year 1 of GAIA-CLIM. The User Survey provided important input to the required functionality of the Virtual Observatory. The new gaps identified in the first year of GAIA-CLIM, and recorded for the first time here in or input to v3 of the GAID relate mostly to technical aspects (G5.9, G5.10 and G5.11). In particular, Gap 5.9 derived from the User Survey findings identifies the need to have an online radiative transfer capability available in the planned Virtual Observatory to allow the transfer of reference measurements into the measurement space of satellite instruments to enable a more meaningful comparison to satellite data. The integration of the GRUAN processor, developed by WP4, into the Virtual Observatory and its online accessibility to create simulated measurements for any satellite instrument for which co-locations with the GRUAN reference measurements exist in the Virtual Observatory database will be delivered by WP5 towards the end of the project.

Looking forward towards a potential operational implementation of the Virtual Observatory in the context of the Copernicus Services, two gaps (G5.10 and G5.11) have also organizational aspects that have been included as well. Those gaps address the establishment of information on uncertainty quantification for more ECVs than considered in GAIA-CLIM and the provision of reference measurements with higher timeliness as this is potentially required to further assure the quality of satellite and modeling products for the Copernicus Services. The remedy of such organizational gaps is challenging as it needs the involvement of several international bodies and goes beyond the lifetime of the GAIA-CLIM project.

References

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.

Glossary

ASCII	American Standard Code for Information Interchange
BUFR	Binary Universal Form for the Representation of Meteorological Data
CEOS	Committee on Earth Observation Satellites
ECV	Essential Climate Variable
ESA CCI	European Space Agency, Climate Change Initiative
EO	Earth Observation
ESA-CCI	European Space Agency Climate Change Initiative
ESGF	Earth System Grid Federation
GCOS	Global Climate Observing System
GEOSS	Global Earth Observation System of Systems
GRUAN	GCOS Reference Upper-Air Network, http://www.gruan.org
HDF	Hierarchical Data Format
ICARE	ICARE Data and Services Center, Cloud-Aerosol-Water-Radiation Interactions http://www.icare.univ-lille1.fr/
NetCDF	Network Common Data Format
NOAA	US National Oceanic and Atmospheric Administration, http://www.noaa.gov/http://www.noaa.gov/
NORS	Network of Remote Sensing Ground-Based Observations in support of the Copernicus Atmospheric Service. http://nors.aeronomie.be/index.php/project
NWP	Numerical Weather Prediction
Obs4MIPs	Observations for Model Intercomparison Projects including data provision using the ESGF

QA4ECV	Quality Assurance for Essential Climate Variables, http://www.qa4ecv.eu
QA4EO	Quality Assurance framework for Earth Observation, http://www.qa4eo.org
WMO	World Meteorological Organisation, http://www.wmo.int