GAIA-CLIM Report / Deliverable D3.5

Gap Analysis for Integrated Atmospheric ECV CLImate Monitoring:

Beta set of tools for quantification of co-location mismatch and smoothing uncertainties and associated documentation for integration in the development of the virtual observatory

A Horizon 2020 project; Grant agreement: 640276

Date: 30 April 2017

Beneficiary: TUT

Nature: R

Dissemination level: PU

Work Package	WP3 – Comparison error budget closure – Quantifying metrology related uncertainties of data comparisons
Deliverable	D3.5
Title	Beta set of tools for quantification of co-location mismatch and smoothing uncertainties and associated documentation for integration in the development of the virtual observatory
Nature	R
Dissemination	PU
Beneficiary	Tallinn University of Technology (TUT), Tallinn, Estonia
Date	April 2017
Status	final
Authors	Kalev Rannat (TUT), Kuldar Taveter (TUT), Hannes Keernik (TUT), Merik Meriste (TUT), Tijl Verhoelst (BIRA-IASB), Jean-Christopher Lambert (BIRA-IASB), Alessandro Fassò (UniBergamo), Arndt Meier (EUMETSAT), Fabio Madonna (CNR)
Contacts	kalev.rannat@ttu.ee, tijl.verhoelst@aeronomie.be, j-c.lambert@aeronomie.be
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.

Table of Contents

Executive summary

GAIA-CLIM, a H2020 project funded by the European Commission, aims to support Europe's Earth Observation programme Copernicus by assessing and improving the fitness-for-purpose of sub-orbital (ground- and balloon-based) reference measurements in the validation of observational data sets from satellites. In particular, the project aims at improved traceability and uncertainty characterization of the individual sub-orbital measurement (systems) and of the comparison with satellite data.

A key issue in the geophysical validation of satellite data sets with respect to sub-orbital reference measurements is the interpretation of their differences in terms of known, quantified uncertainties. This aspect includes not only the measurement uncertainties associated with the individual measurements, but also the additional uncertainties that appear when comparing different perceptions of the inhomogeneous and variable atmosphere, that is, when comparing data sets characterized with different sampling and smoothing properties, both in space and time. Those "comparison uncertainties" have been the main topic of investigation for GAIA-CLIM Work Package 3 (WP3) and the results on that have been published in D3.4 (Fassò et al., 2017).

Deliverable D3.5 describes the progress toward implementation of co-location mismatch and smoothing uncertainties in the "Virtual Observatory" (VO), which within GAIA-CLIM shall only serve as a proof-of-concept facility. The work done within WP3 T3.3 can be presented in 2 categories:

- Design of the system (i.e. the VO) with a cost/benefit evaluation of partners' software-tools for transferring WP3 results into the VO and giving a specific description of the chosen solution per ECV and measuring platform,
- Computation of LUTs (for total ozone).

Due to the dual-character of the work and the decisions made after the results of software analysis, D3.5 serves both as a report and a deliverable.

A key challenge addressed in D3.5 is the cost-effective implementation of the complex developments and CPU-intensive computing in WP3 (tasks T3.1 and T3.2, published in D3.4) and transferring this into a user-oriented platform such as the VO. The solutions presented in D3.5 favor pre-computed results where possible, while retaining some flexibility to address the specific user interactions envisaged for the VO. For example, integration of Observing System of Systems Simulator for Multi-mission Synergies Exploration (OSSSMOSE) simulations would have been completely impossible at this time within the VO due to the technical development status and computational demand requirements. However, implementation of look-up tables (LUT) reduces the complexity enabling the use of climatological results from OSSSMOSE within the VO. Technically, the VO in its current form relies on pre-computed numeric tables to calculate realistic smoothing and sampling errors. Remaining resources make it unlikely that OSSSMOSE could be integrated in the VO. In fact, this may not even be desirable given the complexity and horse power requirements of OSSSMOSE. Performing LUT calculations offline is working fine and only needs to be done once per reference instrument/product constellation. If pre-computed tables for all co-located instruments and Essential Climate Variables (ECVs) can be produced, then the VO could benefit from using the same technical solutions (beta set of tools) for comparison of all co-located observations. Following

GAIA-CLIM, further development of the VO may allow the use of more dynamically produced co-location uncertainty estimates.

1. Introduction

This deliverable gives an overview of how the results obtained in WP3 will be embedded in the VO. It utilises the results, tools, and methods used for calculating co-location mismatch and smoothing uncertainties for observations of ECVs. It describes the ways in which these results can be implemented either by upgrading the existing tools/features already implemented (as a current status of the VO prototype) or by developing and integrating new features.

The VO developed within WP5 of GAIA-CLIM enables end-users to explore, interrogate, extract and analyse co-locations between satellite data and high-quality reference and baseline network data.

The VO follows a Client-Server architecture. The Client part of the VO consists of the Graphical User Interface (GUI). The Server part of the VO consists of the Web Server supporting the GUI functionalities, a database containing co-located observational data, utilities to manage and access the database, and utilities for filtering and managing the data. The architecture of the VO is depicted in Figure 1. An overview of the VO and GUI as a whole can be found in Section 3 of GAIA-CLIM deliverable D5.3.

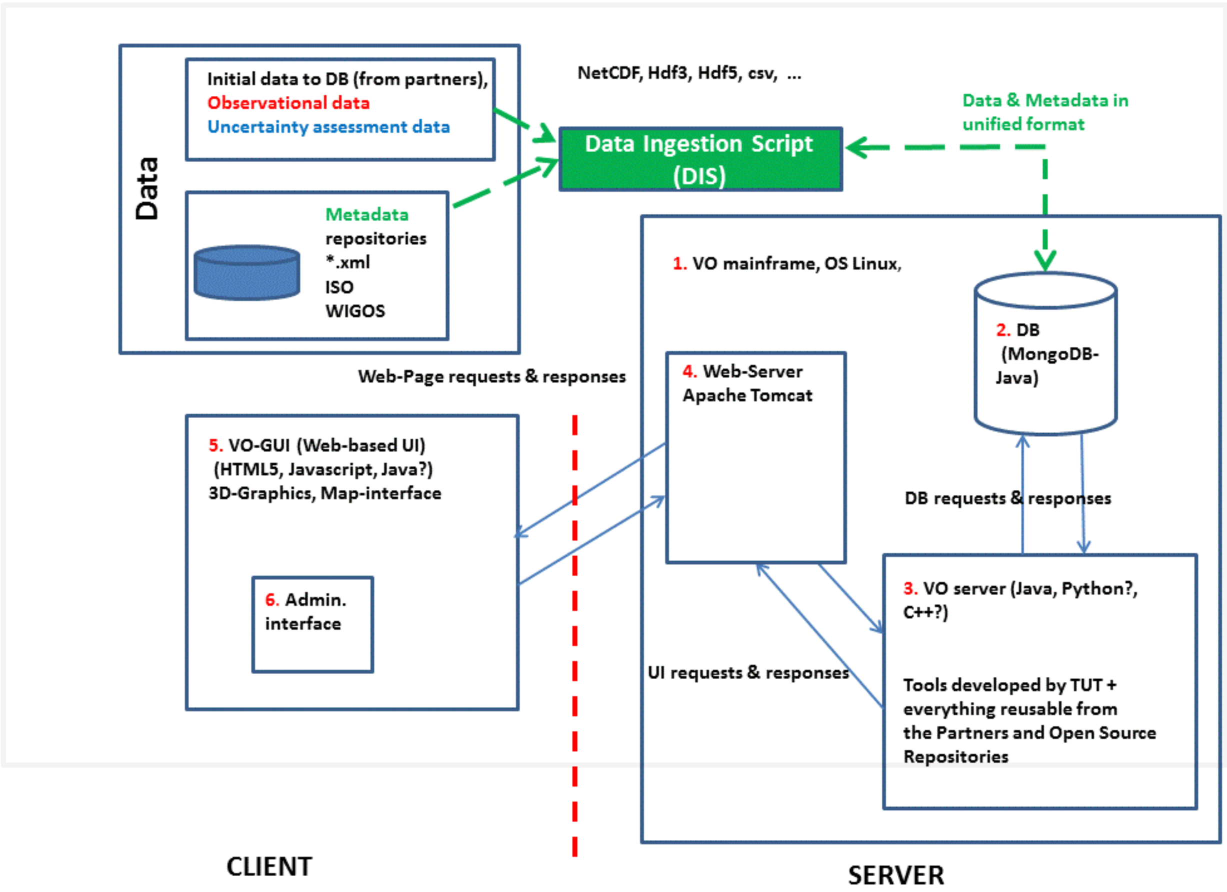

Figure 1: The implemented VO architecture

The operations by the VO for data management and processing include finding the co-locations, presenting the measurement uncertainties, and calculating additional errors due to spatio-temporal mismatch, and smoothing. The operations dealing with uncertainties arising from mismatch effects have been designed and will get implemented in the VO based on the results originating in WP3 of GAIA-CLIM. The scientific results have been published in D3.4 (Fassò et al., 2017). In most cases, the calculation of individual unique spatio-temporal mismatch effects and the related data processing are rather complicated and require a lot of computing power. Additionally, the software-tools used by different project partners are not only technically complex, but also not developed for integration in an environment like the VO. No software integration is possible without an Application Programming Interface (API) developed by a software provider. The time-line and resources available don't allow developing the necessary APIs for this project.

Therefore, it was decided in the project that only pre-calculated and pre-stored smoothing and sampling uncertainties would be used in the VO, either as part of pre-co-located data sets or as Look-Up Tables (LUTs) covering a wide range of viewing geometries, measurement principles and co-location criteria, from which the mismatch uncertainty for a specific comparison can be extracted. The rationale for this decision was that because of the computing power required, a fully operational VO is best (and cost-effectively) served by integrating into it precomputed results rather than the full software suites like the Observing System of Systems Simulator for Multi-mission Synergies

Exploration (OSSSMOSE) or other statistical tools. Indeed, several large operational systems (e.g. satellite retrieval systems) rely on LUTs for crucial processing steps that have to be executed quickly. This does not preclude the use of more dynamically estimated uncertainty estimates in the future, were the VO development to be taken forward and operationalized.

Although the LUTs do not by themselves constitute a traditional software tool, they can be considered as tools or core-components of any script or module used by the VO for visualization of uncertainties originating from co-location mismatch and smoothing errors.

For now, pre-calculated LUTs offer a fast and relatively easy way to implement in the VO the results demonstrated in D3.4 (Fassò et al., 2017). This kind of solution technically means combining measurement data and uncertainties with uncertainties originating in spatio-temporal mismatch and with co-location uncertainties presented in the form of pre-calculated numerical tables. This kind of approach avoids any traditional (often technically complicated and time-consuming) software integration, matches well with the paradigm of the VO as a demonstrator, and does not increase the computing load of the VO-server. By means of pre-calculated LUTs, the results obtained within WP3 can be fully embedded into the VO. The main open technical issue left for the VO and the WP3 tools will be the graphical visualization of the pre-computed data and the related uncertainties.

As reflected in Figure 1, the uncertainties pre-calculated within WP3 will be ingested into the VO database by means of a Data Ingestion Script (DIS). To allow end users of the VO to view the uncertainties pre-computed by means of the methods described in D3.4 (Fassò et al., 2017) and embedded in the VO in the manner described above, the VO needs to be complemented within WP5 with some additional tools such as an extended plot engine, simple statistics toolbox, and tools enabling the usage of the LUTs.

The methods and tools that will be addressed by the respective sections of this deliverable are as follows:

- OSSSMOSE LUTs for total ozone columns in Section 2
- Tools for IASI – RAOB comparisons in Section 3
- Tools for AATSR – AERONET comparisons in Section 4
- Tools for CALIPSO – EARLINET comparisons in Section 5
- Additional developments for Data Ingestion Script – in Section 6

All the tools or technical means (including LUTs) developed or being developed in frames of WP3 and WP5 as components of the VO and the VO GUI, the data coming from different sources and partners (delivered off-line but ingested into the MongoDB), the VO Server, its GUI and MongoDB will be installed at and hosted by EUMETSAT server(s).

The Annexes A.1-A.3 describe presently available LUTs, the construction of LUTs, and the related file formats. As such, the Annexes A.1-A.3 are an essential part of the documentation of the tools under development in the VO.

2. Observing System of Systems Simulator for Multi-mission Synergies Exploration look-up tables for total ozone column

An extensive analysis of smoothing and co-location mismatch uncertainties for total ozone column measurements and comparison between them was finalized by Verhoelst et al. (2015) in the early stages of the GAIA-CLIM project, building on work carried out in the EC FP6 project GEOmon and in various nationally funded projects. In particular, smoothing properties of total ozone column measuring instruments were analysed and parametrized by Lambert & Vandenbussche (2011) and Vandenbussche et al. (2011). In Verhoelst et al. (2015), the OSSSMOSE was used to construct reanalysis-based simulations of measurements, including a detailed treatment of the spatio-temporal smoothing and sampling properties of the measurements and comparison between them.

While these detailed simulations allowed for an in-depth characterization of the errors due to smoothing and sampling differences, embedding the OSSSMOSE system directly into the VO developed within the GAIA-CLIM project would be both outside the scope of the project and difficult to reconcile with the need for fast and straightforward user interactions. Consequently, it was decided to create an intermediate tool in the form of LUTs, which are based on extensive (time-consuming) OSSSMOSE simulations but require only minimal manipulations for use by the VO. The LUTs represent the “climatological” properties of the smoothing errors, as derived from a multi-year simulation for different instruments at different locations and under different measurement conditions. For new total ozone data ingested into the VO, the LUTs allow for a straightforward estimate of the smoothing uncertainties to be associated with the ozone data. This is done simply by finding the LUT entry for the appropriate location, time and other defining measurement characteristic (e.g. solar zenith angle). Similarly, when total ozone column data are co-located within the VO (e.g. a satellite sounder versus a ground-based reference instrument), the “co-location LUT” provides a direct estimate of the co-location mismatch errors based solely on the location of the ground site, the time-of-year and the adopted spatio-temporal co-location criteria.

In terms of performance, i.e. accuracy in estimating the actual smoothing and co-location uncertainties, early tests indicate that the LUTs are in general in very good agreement with the specific simulations performed in Verhoelst et al. (2015). Their main limitation is that, by definition, they can not capture specific atmospheric events that do not conform to the climatological behaviour. Another limitation of the LUT approach (as implemented now) is that it does not allow a rigorous handling of the potential correlations between smoothing errors for two measurements that are being compared under similar viewing geometries and measurement principles. Specific usage recommendations are detailed in Section A.3.

Full details about the construction of LUTs and the contents of the corresponding files are provided in Annex A. The following LUTs have already been embedded into the VO database:

1. Smoothing uncertainties for UV-Vis zenith scattered-light differential optical absorption spectroscopy (ZSL-DOAS) measurements (see Section A.1.1),
2. Smoothing uncertainties for direct-sun UV-Vis measurements (Brewers and Dobsons, see Section A.1.2),

3. Smoothing uncertainties for UV-Vis nadir backscatter measurements (e.g. GOME/ERS-2 and GOME-2 on MetOp-A/B, see Section A.1.3)
4. Co-location mismatch uncertainties for various spatial and temporal co-location constraints (see Section A.2)

Further developments are required to actually apply these LUTs within the VO, but these represent only a minor technical effort, consisting of implementing appropriate reading, interpolation, and visualization routines. Specific usage scenarios of the LUTs as appropriate for total ozone column validation work are described in Section A.3.

3. Tools for comparing Infrared Atmospheric Sounding Interferometer with radiosonde observations

Three main outputs will be offered to the end users through the VO regarding the considered study area described in D3.4 (Fasso et al., 2017) in Central EUrope (CEU):

- Sparseness and processing uncertainties of radiosonde observation (RAOB) products. These uncertainty profiles will be available only for the GRUAN site at Lindenberg, Germany, which performs four times daily ascents. They should be displayed as uncertainties of RAOB profiles, and not necessarily in a RAOB-IASI co-location comparison. They can be represented either as precomputed LUTs with the entries for pressure, season, and day or night, attached to each RAOB profile, or pre-attached to the corresponding RAOB profile.
- Vertical smoothing has the following two types:
  a) Harmonized RAOB profiles. These are associated with the corresponding RAOB profiles and should be displayed by the VO as an option chosen by the user. Harmonized RAOB profiles provide a graphical representation of the IASI smoothing effect of the IASI profile co-located with the considered RAOB profile. Harmonized RAOB profiles are precomputed and associated with each RAOB profile. In CEU one harmonized RAOB profile can be associated with each RAOB.
  b) Vertical smoothing uncertainty. This is the component of the total co-location mismatch uncertainty due to vertical smoothing. It should be displayed as a profile for each IASI-RAOB co-location. Vertical smoothing can be provided as a precomputed LUT with the entries for pressure, latitude, longitude, season, and day or night, and it can be attached to the observational data for each co-location. Note that vertical smoothing is computed by averaging over CEU and does not depend on co-location distance. Alternatively, vertical smoothing can be attached to each co-location in CEU.
- Harmonized co-location uncertainty adjusted for differences in vertical smoothing. Spatio-temporal mismatch uncertainty has been estimated for temperature and humidity as a function of air distance and delay at pressure levels available for IASI. This uncertainty can be given as a LUT with the entries for pressure, season, day or night, horizontal distance, and time distance (i.e. delay). LUT of this kind enables the user to easily find the *harmonized co-*

location uncertainty adjusted for differences in vertical smoothing. For the given values of the entries, the harmonized co-location uncertainty has to be considered as an average over CEU. This uncertainty may be provided to the VO both as a precomputed stand-alone table or attached to each co-location, if required. This information can be displayed by the VO either as a 3D-plot or as a contour plot. In principle, this information could also be used by the VO to select all the co-locations meeting a certain threshold of co-location uncertainty. All these outputs for IASI-RAOB comparisons developed and provided by WP3 may be used as inputs for the VO, which are uploaded offline onto the VO.

Regarding smoothing with splines (either Hermite interpolating splines described in D3.4 (Fassò et al., 2017) or smoothing B-splines), the possibility to implement a “spline engine” in the VO was considered, but was later replaced with the alternative of uploading pre-computed high resolution spline outputs to the VO.

Finally, the VO will visualize observed parameters (e.g. temperature, humidity) together with the measurement uncertainty and additional uncertainties due to vertical smoothing and harmonization of RAOB profiles, and sparseness and processing uncertainties. Some new features must be added to the GUI – for example, 3D-contour plots for the co-location uncertainty of temperature and humidity as a function of air distance and delay, at pressure levels available for IASI.

At the present state of the VO development, we have not integrated any tools into the VO in the context of IASI-RAOB comparisons. However, there is a clear roadmap for their integration and currently no issues in their eventual inclusion are foreseen.

4. Tools for comparing Advanced Along Track Scanning Radiometer with Aerosol Robotic Network Sun photometers

This section describes the development efforts for the VO undertaken for allowing comparison of the total atmospheric column aerosol optical depth (AOD) values obtained with Advanced Along Track Scanning Radiometer (AATSR) and Aerosol Robotic Network (AERONET) sun photometers. The work done and presented in D3.4 (Fassò et al., 2017) identifies the main sources of co-location mismatch uncertainties in comparison with AATSR and AERONET AOD retrievals. Quantitative indicators of these uncertainties have been developed and tested by EUMETSAT. In addition, the effect of comparing sampling parameters has been assessed.

The VO will be accompanied by the description of methods that can be used to assess the co-location mismatch uncertainty in comparison with AOD retrievals made by satellite instruments and AERONET sun photometers. The methods are based on the general knowledge of satellite product validation against AERONET, and on the case studies presented Section 6 of D3.4 (Fassò et al., 2017). The description includes identification of parameters characterizing the uncertainties and recommendations for the default sampling parameters used in the comparison.

The VO should provide the user with an option to choose the parameters and the co-location uncertainties by means of the GUI and visualize AATSR-AERONET measurements. The data ingested into the VO is to be pre-co-located based on AERONET to AATSR observations. The main enhancement required for the VO is working with co-located data and enhancing plotting capabilities. The data is provided off-line with pre-computed co-location mismatch uncertainties.

At the present state of development, we do not plan to integrate any additional tools into the VO in the context of AATSR-AERONET comparisons. This is caused by open questions on the metrological characterisation of AERONET instruments, which are currently under consideration by WP2. Therefore, we are not certain as it stands that it will be possible to use the AATSR-AERONET comparisons within the VO.

5. Tools for comparing Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations with European Aerosol Research Lidar Network data

WP3 has made an effort to identify the main contributions to co-location mismatch uncertainties in comparison with the Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALIPSO) and European Aerosol Research Lidar Network (EARLINET) aerosol backscatter profiles. In the LIght Detection and Ranging (LIDAR) framework described in D3.4 (Fassò et al., 2017), a tool for end users has been planned to be integrated with the VO. This tool will implement an independent method for co-locating LIDAR retrievals from EARLINET with the best horizontally smoothed CALIPSO retrievals, for sites located in Évora, Granada, Naples, Potenza and Leipzig. However, at the present stage of development, it has been agreed that this kind of tool will not be implemented within this project's timeframe and therefore the corresponding data will be provided offline and the uncertainties will be pre-computed and ingested into the VO database.

The datasets of CALIPSO-EARLINET comparisons may be provided either in the original format – the hdf-format for CALIPSO and the netCDF-format for EARLINET – or in the csv-format. In particular, the datasets may include original data, namely measurements and their uncertainties, the estimated co-location uncertainties, and the optimally horizontally smoothed CALIPSO retrievals, which are currently in the csv-format.

For the VO, it means adding capabilities to work with different data formats (netCDF, hdf and csv). The VO has been initially designed and implemented based on the paradigm of accessing observational data through its metadata and having all the data stored in MongoDB. When designing the VO, it has been considered that data from different networks can be obtained in different formats (e.g. netCDF or hdf). Therefore, a Data Ingestion Script has been developed for converting data and metadata into the unified format used by the VO database (Figure 1). In general, the VO does not need additional data stored in its original unconverted format. However, this option is not excluded and files containing such metadata and observation data can be kept on the VO server for the purposes of data export in its original format and linked to the VO DB through dedicated identifiers (IDs) included by the metadata files.

From the visualization point of view, the needs for visualization described in D3.4 (Fassò et al., 2017) can be met by general functionalities of the VO GUI. With the assumption that pre-co-located data with co-location mismatch uncertainties are stored in the VO DB, the visualization of the results for certain sites and smoothing categories can be achieved by modification of data filtering capabilities. The GUI will offer checkboxes for choosing the following smoothing categories for Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP) at the sites Évora, Granada, Napoli, Potenza, and Leipzig, and with the horizontal smoothing range 5-205 km:

- Co-locations for minimum smoothing (5 km);
- Co-locations for standard smoothing (45 km);
- Co-locations for optimal smoothing (site specific).

At the present state of development, we do not plan to integrate any additional tools into the VO in the context of CALIPSO–EARLINET comparisons. However, it might be possible to integrate such tools by the end of the project.

6. Passing co-located data to and from the Virtual Observatory database

The Data Ingestion Script (DIS) implemented in the Python language has been developed for ingesting the information from source files into the DB of the VO. The DIS can handle observational data files represented in the netCDF-, hdf- and csv-formats. In addition, the script can be used for ingesting XML files represented in the WIGOS and ISO-19115 formats. However, information from source files is not ingested directly as it is, but is rather converted into a unified format. The unification is applied for field names as well as for values (e.g. conversion to SI-units). This is achieved by the following three steps:

1. After detecting the data format and ensuring the data is not already ingested, the DIS calls two conversion scripts, one for metadata, and one for observational data;
2. Conversion scripts use format-specific input files containing pre-defined field names and/or units for unification. When necessary, LUTs are used in this phase and interpolation is carried out to the actual coordinates in the format (*lat, month, sza*) or (*dr, dt, lat, lon, month*), where *lat, lon, month, sza, dr and dt* denote the latitude, longitude, month, solar zenith angle, space- and time differences accordingly. Following that, computed uncertainty is added as a separate field such as "TOC_directsun_sza_smoothing". During the conversion, when needed, any original field from the source files is split into several different fields. Also, in order to make data querying more convenient at a later stage, several other fields may be added, such as the field in a metadata document containing the list of all the variables, the values of which have been estimated during the observations, links between co-located observations, uncertainties calculated based on LUTs, the coordinate field together with the 2D-sphere index for geospatial queries, and so on. Any restrictions based on quality information are also considered, such as the *qcflag* and *qcinfo* values in case of the GRUAN processor. Unified data is returned to DIS in the json-format;

3. DIS ingests metadata and observational data into separate Mongo DB collections. One document per observational variable is created. Output files containing information about the ingestion results are written. The user will be notified about any failures or new undefined data fields identified during the ingestion.

In the VO database, each piece of observational data is linked with its metadata using the Mongo ObjectId value. Two co-located observations are also linked in the same way.

Queries for retrieving co-located observations can be made by station, variable, start date and time, or by co-location criteria – time window in hours and spatial difference in kilometres. Co-located observations that meet user-defined conditions are always selected in pairs. Figure 2 shows an example document from the collection of observational data. The observational data in Figure 2 only corresponds to a single observation, meaning that the co-location criteria are missing from this collection.


```
> db.measurements.findOne( { 'station_platform_name': 'Barrow', 'measurand_variable': 'Air temperature' })
{
        "_id" : ObjectId("5834c5b9ddb5932e14306222"),
        "measurand_variable" : "Air temperature",
        "metadata_CCI_CF_ID" : ObjectId("5834c5b9ddb5932e1430621f"),
        "station_platform_name" : "Barrow",
        "start_date" : "2015-01-07",
        "start_time" : "19:17:30",
        "dimensions" : [
                "Time"
        ],
        "name" : "temp",
        "standard_name" : "air_temperature",
        "units" : "K",
        "long_name" : "Temperature",
        "format_type" : "FLT",
        "format_format" : "F6.2",
        "format_width" : "6",
        "format_nan" : "NaN",
        "processing_flag" : "raw, uncertainy calculated, corrected, internal QC passed, additional QC passed, outlier removed",
        "source_desc" : "FRAWPTU_RC",
        "column_type" : "original data",
        "resolution" : "10.0 s (time)",
        "comment" : "Air temperature, uncertainty estimated with GRUAN correction scheme, ",
        "related_columns" : "u_cor_temp u_std_temp u_temp ",
        "coordinates" : "lon lat alt",
        "values" : [
                247.0367889404297,
                246.1272735595703,
                246.03549194335938,
                246.2278594970703,
                246.57928466796875,
                247.29397583007812,
                247.99705505371094,
                248.6519775390625,
                249.27975463867188,
                249.8970947265625,
                250.5074005126953,
                251.11354064941406,
                251.6936492919922,
                252.258544921875,
                252.7930450439453,
                253.2933349609375,
                253.74514770507812,
                254.14988708496094,
                254.50672912597656,
                254.8199920654297,
                255.0972900390625,
                255.34898376464844,
                255.58627319335938,
                255.8186798095703,
                256.0525817871094,
                256.28826904296875,
```


Figure 2: Example document from the collection of observational data

7. Subsequent developments

The technical stage of the VO's development does not yet allow to work with co-located data. However, the main workflow – either with or without co-located data – will remain unchanged and is based on the following logic for searching, retrieving and visualizing the data:

1. Choose an ECV from a reference network.
2. Select the search area by providing either a point (latitude and longitude) and a search radius around the given point, or a search polygon.
3. Provide the time interval in the format "From YYYY-MM-DD To YYYY-MM-DD" and "From HH:MM:SS To HH:MM:SS".
4. Send a data request.
5. Wait for and receive the response from the VO.
6. Visualize and extract the data by choosing appropriate options for plotting.

Steps 2 and 3 serve as spatio-temporal criteria for the DB. Step 2 determines the area and Step 3 determines the total time interval for which the user is willing to obtain the data in the form of individual data points, co-locations, and differences over a certain time span (e.g., 10 years). Co-located data for a user session always form a subset of all the data stored in the VO database. The solution chosen for the first prototype of the VO – using pre-co-located data – can be seen as an initial version. For handling co-located measurements, the VO uses additional data fields in the metadata (please see Section 6 of this deliverable). These additional data fields contain (a) criteria for calculating the distance between points of co-located measurements and (b) co-location time – time difference between spatially co-located observations. Consequently, the next steps in the development of the VO will be concerned with:

1) Additional utilities for handling and unification of data/metadata (integrated into DIS);
2) Tools for the ingestion of co-located data to be integrated with the DIS.

The GUI of the VO should be complemented with the features for handling and visualization of co-located data, in addition to the following already implemented features for handling non-co-located data:

1) Plotting total column values for any user selected ECV at a chosen or all sites, such as, for example, Total Ozone Column (TOC) values with their respective uncertainties. Typically, the results are visualized as time series of observations with error bars. Plotting the time series can include an option to plot multiple stations on the same graph, using different colors for different sites.
2) Plotting ECV profiles with uncertainties in observational data at a site as an individual profile with error bars, or as a contour plot showing the time on the x-axis, altitude or pressure on the y-axis, and contours for a set of ECV concentrations.

Handling of additional uncertainties by the VO and the corresponding data flow are described in Figure 3. The data for the VO is delivered off-line by different data providers. Observational data is always accompanied by the relevant uncertainties. In Figure 3, data flows of the types A, B, C, and D denote different options for handling uncertainties in the VO, u_m, u_{m1}, u_{m2} – measurement uncertainties in observational data, u_c – co-location uncertainty, and u_s – uncertainty from smoothing.

Currently, the VO needs to distinguish between the "raw" data and already pre-co-located data. Up till now, the VO has worked with data flows of the type A, whereby the user has a possibility to obtain observational data and visualize any ECV together with its measurement uncertainty values.

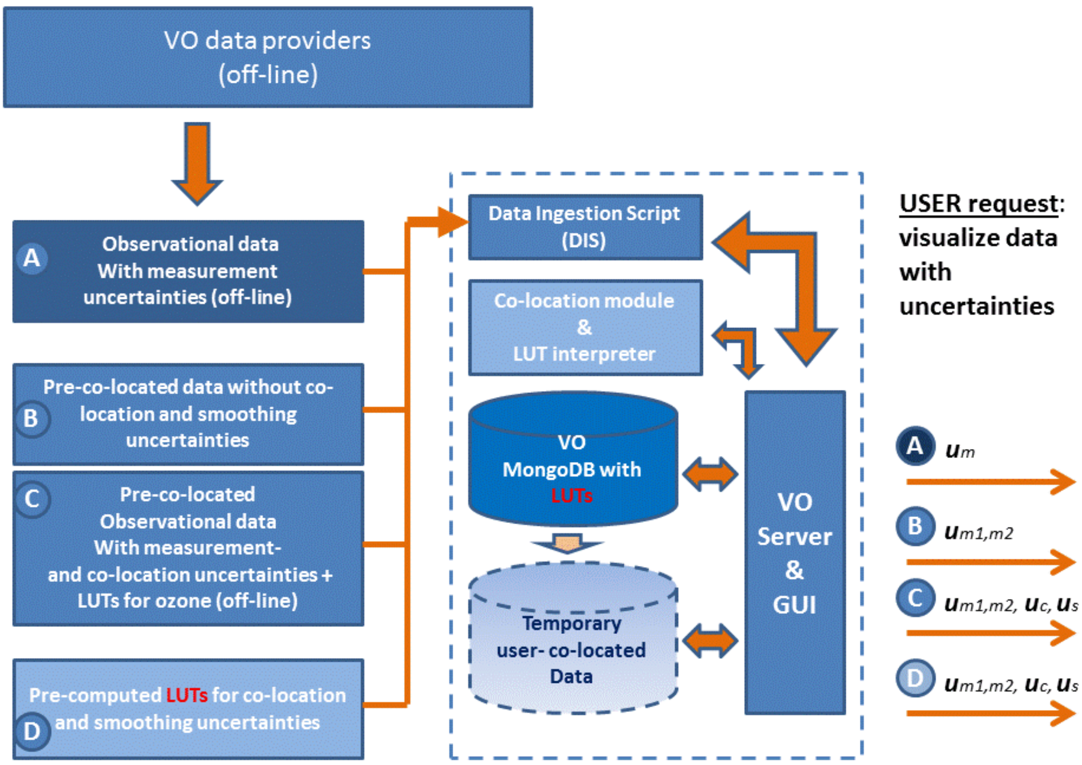

Figure 3: Uncertainties in VO, delivered by data providers.

Further development work with the VO will include implementing the algorithms developed within WP3 for quantification of co-location mismatch and smoothing uncertainties. As a result, the VO users will be able to visualize pre-co-located data with additional uncertainties originating in co-location mismatch and smoothing. The development of the features of the VO enabling to use pre-co-located data is at its starting phase, where initial tests have been conducted with synthetic data belonging to the data flow of type B.

The next step will be the implementation of the data flow of type C with real pre-co-located datasets and with smoothing and co-location uncertainties for ozone presented in the form of LUTs. Initially, the observational data is processed by DIS, as is shown in Figure 3, where the smoothing and co-location uncertainties are calculated by means of LUTs and are ingested into the VO database together with the original LUTs for ozone. We do not yet have any LUTs for other ECVs. However, for some ECVs and observation methods, the collocation mismatch errors are much smaller than for the example of ozone from UV-VIS instruments discussed. If one uses microwave sounders on an ECV with a large correlation length, then the mismatch errors may become negligible or close to. It needs to be further investigated where the usage of LUTs is feasible for other ECVs and co-located instruments.

The drawback of not using pre-computed LUTs within the VO is that for the comparisons without LUTs for smoothing and co-location uncertainties, we cannot allow for flexible co-location criteria for certain types of ECVs and instruments. Flexibility can be offered only either by having all the required tools integrated into the VO or by using pre-calculated LUTs adaptable to user-chosen co-location criteria. In fact, the use of LUTs as in the example of ozone, demonstrates a remarkable technical simplification of the VO, because it avoids the complex and laborious integration of OSSSMOSE into the VO. At the same time, neither the functionality of the VO nor the quality of data products gets notably reduced. With the data flow of type C, the VO user is still limited to pre-ingested co-located data.

The development step of the VO enabling the usage of data flow of type D will allow flexible co-location criteria for the VO user and will demonstrate full efficiency of the concept of LUTs (as opposed to using them only once for a pre-co-located dataset as in the type C data flow). At this step, the VO will be complemented with an additional functional module – Co-location module & LUT interpreter (Figure 3). This module will allow for creating new temporary sets of co-located data and calculating the smoothing and co-location uncertainties. Initially the module could work with only the ozone because the LUTs for the ozone are already available. For other ECVs additional development is foreseen either by creating similar LUTs by the partners or integrating with the VO software systems by the GAIA-CLIM partners for calculating in the VO smoothing and co-location uncertainties.

The Co-location module & LUT interpreter introduced in the previous paragraph can be seen as an independent tool allowing for calculating uncertainties by means of LUTs and creating temporary co-located data-sets based on user-defined co-location criteria. The tool itself may contain different utilities for operating with LUTs and finding new co-locations. Currently, a part of the mentioned functionalities have been developed for DIS, allowing for additional uncertainty calculations before the data ingestion phase.

Upgrading the VO to the level fulfilling the requirements for handling co-location mismatch and smoothing errors comprises the steps of:

- enhancing the plot engine of the VO;
- complementing the VO with some statistical tools;
- extending the capabilities of the DIS to take full advantage of LUTs.

The GUI of the VO will be enhanced with the following new options for visualization of co-located data:

1. If the visualization of TOC has been chosen by the VO user (i.e. one TOC value at a time for a site at (lat, lon) for the time point DDMMYYYY at HH:MM:SS with TOC from the co-located instrument(s)), a TOC time series is plotted with uncertainties presented in the format (measured value + measurement uncertainty + co-location mismatch uncertainties from LUTs). The following options exist for this kind of plotting:

    a. Plotting the time series for the 2 instruments, with error bars showing only the observational uncertainties and error bars represented in a different color showing smoothing uncertainties retrieved from the smoothing LUTs, and error bars showing co-location mismatches;
    b. Plotting the time series of the absolute (i.e. in Dobson units) differences, where the error bar is the combined uncertainty – quadratic sum of different uncertainties;
    c. Plotting the time series of relative differences presented in per cent, with the error bars and a shaded zone with the co-location uncertainty;
    d. Histogram of the differences.
2. If the visualization of the ozone profile has been chosen by the VO user for the site at (lat, lon) for DDMMYYYY at HH:MM:SS together with the co-located ground-based or satellite instruments, a set of profiles is plotted for co-located pairs in the same way as for any non-co-located data in the VO DB.

Graphical visualization for all the subtasks – including the IASI – RAOB, AATS – AERONET, AATS – AERONET, and CALIPSO – EARLINET comparisons, and ozone observations with LUT tables – can be supported by the same graphical libraries. However, the usage of these libraries should be modified according to the specific needs for different sets of observational data.

After completing all the activities described above – extending the DIS, enhancing the plot engine and creating necessary additional tools for using LUTs – the VO will be able to work with co-located data and will have the capability to compare observational data originating from different sources. Regarding the visualization of uncertainties, the VO will show not only uncertainties in observational data, but also uncertainties from co-location mismatch and smoothing errors.

8. Conclusions

Development of the VO and its GUI is one of the tasks of WP5 and is in continuous progress. The basic functionalities of the VO, based on the results of user surveys and user feedback have already been implemented, including data handling and basic visualization capabilities. The developments and outcomes from WP3 will not require the introduction of any substantial changes into the VO architecture, business logic, or functionalities. Instead of that, implementation of the results from WP3 means complementing the VO with new features by enhancing and upgrading its current technical capabilities.

The tools for quantification of co-location mismatch and smoothing uncertainties that are being developed within WP3 and have been presented in D3.4 (Fassò et al., 2017) are still at the experimental stage and therefore not ready for integration with the VO. In addition, some of the tools are too complex for direct integration with the VO or rely extensively on proprietary software (e.g. MATLAB). For these reasons, it has been decided to provide the VO input in the form of pre-computed LUTs of parameters that allow calculating individual error components for each particular collocation with high precision. Easy-to-use intermediate tools based on the Look-Up Table (LUT) approach will be created for the VO, depending on the ECVs and underlying methodologies explored in WP3. The first set of such LUTs, based on the OSSSMOSE simulations for total ozone column measurements, has been implemented and is presented in Section 2 and Annex A of this deliverable.

The most important decision made for the VO development is therefore to provide co-located observational data with pre-computed uncertainty parameters. This simplification allows for enhancement of the VO functionalities without integrating it with any additional software systems. The production of these pre-computed uncertainty estimates is currently under way and initial results on total ozone columns are ready to be integrated with the VO. This does not preclude integration of more advanced systems down the line where the VO functions to be operationalized.

All the developments together described in Sections 2-7 form the “Beta set of tools for quantification of co-location mismatch and smoothing uncertainties”.

References

1. BIPM, IEC, IFCC, ILAC, IUPAC, IUPAP, ISO, OIML: The international vocabulary of metrology—basic and general concepts and associated terms (VIM), 3rd edn. JCGM 200:2012, 2012.
2. Fassò, A., Ignaccolo, R., Madonna, F., Demoz, B. B., and Franco-Villoria, M.: Statistical modelling of collocation uncertainty in atmospheric thermodynamic profiles, Atmos. Meas. Tech., 7, 1803-1816, doi:10.5194/amt-7-1803-2014, 2014.
3. Fassò, A., Negri, I., Finazzi, F., Verhoelst, T., Lambert, J.-C., Virtanen, T. H., Kivi, R., Gardiner, T., Madonna, F., Mona, L., Papagiannopuolos, N., Thorne, P.: Measurement mismatch studies and their impact on data comparisons, GAIA-CLIM Report / Deliverable D3.4, 2017.
4. Iarlori, M., Madonna, F., Rizi, V., Trickl, T., and Amodeo, A.: Effective resolution concepts for lidar observations, Atmos. Meas. Tech., 8, 5157-5176, doi:10.5194/amt-8-5157-2015, 2015.
5. Immler, F. J., Dykema, J., Gardiner, T., Whiteman, D. N., Thorne, P. W., and Vömel, H.: Reference Quality Upper-Air Measurements: guidance for developing GRUAN data products, Atmos. Meas. Tech., 3, 1217-1231, doi:10.5194/amt-3-1217-2010, 2010.
6. Lambert, J.-C. and S. Vandenbussche: "EC FP6 GEOmon Technical Note D4.2.1 - Multi-dimensional characterisation of remotely sensed data - Chapter 1: Ground-based measurements," GEOmon TN-IASB-OBSOP / Chapter 1, BIRA-IASB, April 2011.
7. Seidel, D. J., B. Sun, M. Pettey, and A. Reale: Global radiosonde balloon drift statistics, J. Geophys. Res., 116, D07102, doi:10.1029/2010JD014891, 2011.
8. Vandenbussche, S.; Lambert, J.-C. & Spurr, R.: EC FP6 GEOmon Technical Note D4.2.1 - Multi-dimensional characterisation of remotely sensed data - Chapter 5: Satellite measurements of solar/stellar occultation BIRA-IASB, 2010
9. Verhoelst, T., Granville, J., Hendrick, F., Köhler, U., Lerot, C., Pommereau, J.-P., Redondas, A., Van Roozendael, M., and Lambert, J.-C.: Metrology of ground-based satellite validation: co-location mismatch and smoothing issues of total ozone comparisons, Atmos. Meas. Tech., 8, 5039-5062, doi:10.5194/amt-8-5039-2015, 2015.

Annex A: OSSSMOSE-based look-up tables for smoothing and co-location mismatch uncertainties of total ozone measurements

In Verhoelst et al. (2015), the OSSSMOSE system was used to construct a model-based simulation of each individual measurement in a total-ozone column comparison exercise. Consequently, for every measurement pair, it was possible to estimate the co-location error and the smoothing difference error. These could then be aggregated over a certain temporal or spatial window to estimate corresponding uncertainties and systematic errors (biases). The downside of simulating each measurement is the need to go through a lengthy chain of operations: to set up the observation operators, to read the model/reanalysis fields, to average/integrate the fields using the observation operators, to calculate the statistics on the resulting errors to derive uncertainties, etc. Consequently, for use in the GAIA-CLIM Virtual Observatory – or for other applications which require a fast estimate of spatio-temporal mismatch uncertainties without direct access to the observation-operator and reanalysis libraries – it was decided that Look-Up Tables (LUTs) based on a large set of full OSSSMOSE simulations would be valuable. These LUTs are stand-alone in the sense that they no longer require access to the reanalysis fields and observation operators. Instead, they provide a direct link between on the one hand measurement type, location, time and other defining characteristic (which is part of the measurement metadata), and on the other hand the corresponding smoothing and/or co-location uncertainty. In other words, we constructed from a large set of OSSSMOSE simulations a kind of climatology of smoothing and co-location mismatch uncertainties which is tabulated as a function of the main parameters determining the uncertainties.

A distinction is made between:

A. LUTs for *smoothing uncertainties*, related to measurements of a single instrument. So far this includes (1) ground-based zenith-sky UV-Vis DOAS instruments (ZSL-DOAS), (2) ground-based direct-sun UV-Vis instruments (Brewers and Dobsons) and (3) satellite UV-Vis nadir-looking instruments (GOME/ERS2, GOME-2 on Metop-A/B, …)

B. LUTs for *co-location mismatch*, related to the intercomparison of imperfectly co-located measurements. These do not take into account their different smoothing properties, only the different location and time of the nominal measurement coordinates.

The uncertainties and systematic errors estimated in the LUTs of type A need to be taken into account whenever measurements are used without taking into account explicitly their specific spatio-temporal smoothing properties. For instance, time series of measurements at a given station will contain random and systematic components, arising neither due to the atmospheric phenomena under study nor due to the measurement uncertainty, but due to random and systematic spatio-temporal smoothing errors. The random component can be considered to be an additional uncertainty (which can be added quadratically to the measurement uncertainty), the systematic component corresponds to a bias. The latter is not necessarily constant in time, which is why most of the LUTs contain monthly estimates of the systematic smoothing error.

The uncertainties and systematic differences estimated in the LUTs of type B need to be taken into account when comparing different measurements, which are not perfectly co-located. When the sampling pattern from the satellite instruments doesn't coincide with the location and measurement times of the ground network, additional random and systematic differences arise which must be taken into account when interpreting the comparison results. For instance, when calculating differences (per co-located pair), the total uncertainty ("error bar") on that difference should include the measurement uncertainties of both measurements plus this co-location mismatch uncertainty. Also, the smoothing uncertainties can/should be added, though this is rigorously speaking not entirely correct as the smoothing *difference* uncertainty is often not the quadratic sum of the smoothing uncertainties (due to correlations when both instruments have similar viewing geometries). It is therefore advised to only add the smoothing uncertainties when both instruments have fundamentally different viewing geometries (e.g. ZSL-DOAS versus nadir). More guidelines on the proper use of the LUTs are provided in following Sections.

A.1. Smoothing uncertainties

Smoothing uncertainties pertain to a single measurement or instrument: they are the consequence of the measurement not being point-like (in space and time) at the nominal measurement location and time. As such, the measured value will deviate from the true field at the measurement coordinates, and this deviation depends on the inhomogeneity and variability of the atmospheric field and the spatio-temporal smoothing properties of the measurement. A single smoothing error is defined as:

$$\varepsilon_{smoothing} = \frac{TOC_{smoothed} - TOC_{nominal}}{TOC_{nominal}}.$$

These smoothing errors have both a random component (e.g. due to the turbulent nature of the atmosphere), which leads a smoothing uncertainty that can be characterized by the spread of the smoothing errors, and a systematic component that can be characterized by the mean of the smoothing errors. Separate LUTs are provided for the spread and the mean of the simulated smoothing errors. Note that it is impossible to construct meaningful LUTs for individual smoothing errors, due to the stochastic components in the atmospheric variability.

Below, we describe briefly the measurement geometry and how it causes smoothing errors for 2 types of ground-based measurements and 1 type of satellite measurement of the total ozone column. We also describe the way the LUTs were generated and what the files look like. For the description of the measurement principle and the corresponding smoothing properties, the text below is based heavily on work performed in the EC FP6 project GEOMON, and further details can be found in the deliverable D4.2.1 of that project, in particular chapters 1 (Lambert & Vandenbussche, 2011) and 5 (Vandenbussche et al., 2011). For more details on the computation of the smoothing errors we refer to Verhoelst et al. (2015).

A.1.1. ZSL-DOAS

Measurement principle

Atmospheric trace constituents with a strong stratospheric component can easily be detected in zenith scattered light during twilight regimes, when the optical path in the stratosphere is very long. Using appropriate air mass factors, the apparent slant column (derived with an iterative procedure from the observed absorption features and the known absorption cross sections) can be converted into a vertical column. This technique is called zenith scattered light differential optical absorption spectroscopy (ZSL-DOAS), and it is for instance implemented in the SAOZ (Système d'Analyse par Observation Zénithale) series of instruments. The observing geometry of such a ZSL-DOAS measurement is illustrated in Figure 4, immediately revealing the large horizontal smoothing effect.

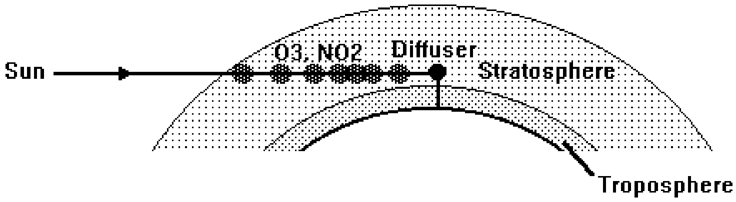

Figure 4: The optical path of a twilight measurement with a ground-based UV-Vis zenith-sky looking instrument. The large horizontal offset and smoothing are obvious.

Smoothing properties

Using a spherical ray tracing technique to quantify (1) the sensitivity of the AMF to perturbations in the O3 concentrations at different altitudes and (2) the corresponding horizontal location of that sensitivity, Lambert & Vandenbussche (2011) arrive at the results presented in the left-hand panel of Figure 5. Combining these results with the solar azimuth angle variation during a twilight measurement sequence (corresponding to a SZA of 91° to 86°) results in the polygons depicted in the right-hand panel of Figure 5.

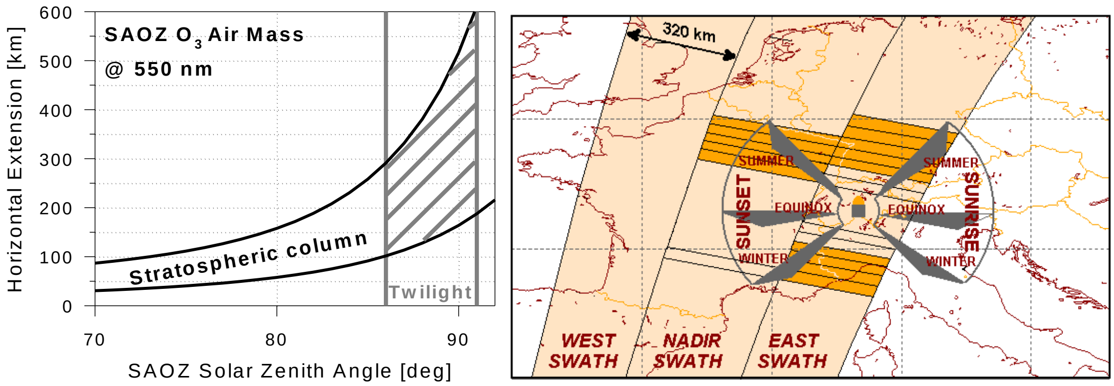

Figure 5*: Left-hand panel*: Horizontal extension of the air mass to which a zenith-sky UV-Vis measurement of total O_3 is sensitive. *Right-hand panel*: Illustration of the horizontal projection of the air masses to which a ZSL-DOAS measurement is sensitive, for both sunset and sunrise, and at summer solstice, equinox, and winter solstice (grey polygons). The satellite pixels (in orange) correspond to GOME measurements but are of no importance here. Both figures from Lambert & Vandenbussche (2011).

LUT construction

The LUTs are based on 10^6 simulated smoothing errors, at random latitude, longitude, and day-of-year, covering a total period of 4 years (2005–2008). Both sunrise and sunset measurements were simulated. The reanalysis fields used are those computed with IFS-MOZART in the context of MACC (now CAMS). Smoothing errors were then binned by latitude (5˚ resolution) and month (all 4 years folded onto one "climatological" year), and both the spread on the smoothing errors (as a proxy of smoothing uncertainty) and the mean smoothing error (as a proxy of smoothing bias) were computed and tabulated. The result is shown in Figure 6.

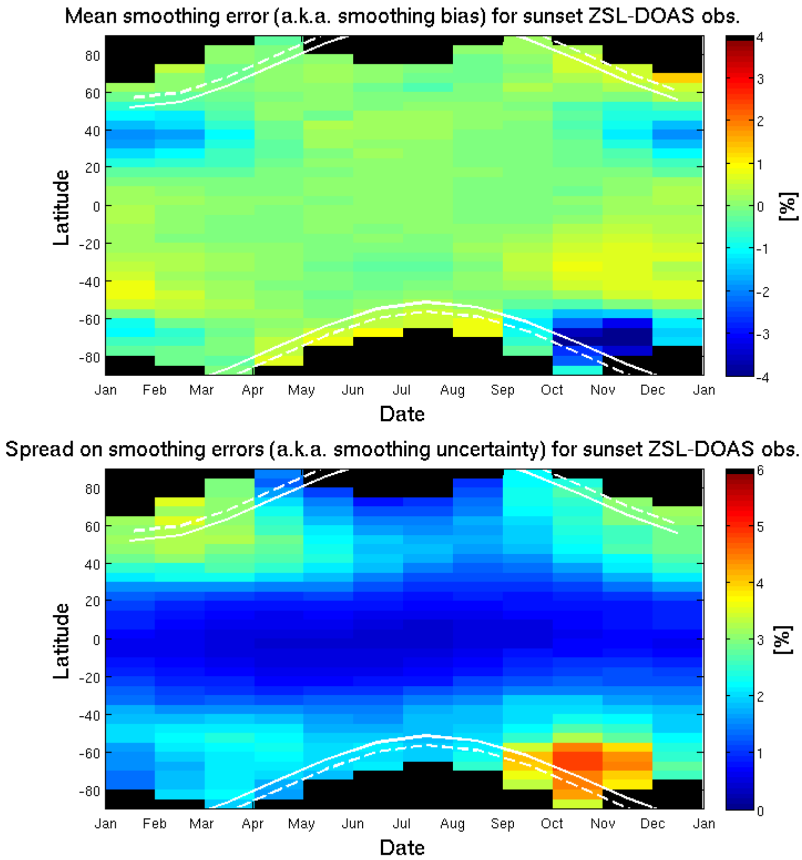

Figure 6: Visualization of the smoothing bias LUT (upper panel) and smoothing uncertainty LUT (lower panel) for ZSL-DOAS measurements of total ozone at sunset. Similar LUTs exist for sunrise measurements. Black regions correspond to conditions not allowing ZSL-DOAS measurements, i.e. there is no sunset (polar day or night). For reference, the solid and dashed white lines represent 75˚ and 80˚ SZA (maximum during the day) respectively.

File description

Two LUTs for ZSL-DOAS measurements were constructed: one for **sunrise** and the other for **sunset** measurements. These were stored in hdf5 files which are named *TOC_SAOZ_sunrise_smoothing.h5* and *TOC_SAOZ_sunset_smoothing.h5* respectively. Each file is only 10KB large. As described above, the influence quantities taken into account are latitude and month-of-year. Figure 7 illustrates the file contents. The use of attributes should make these files self-explanatory.

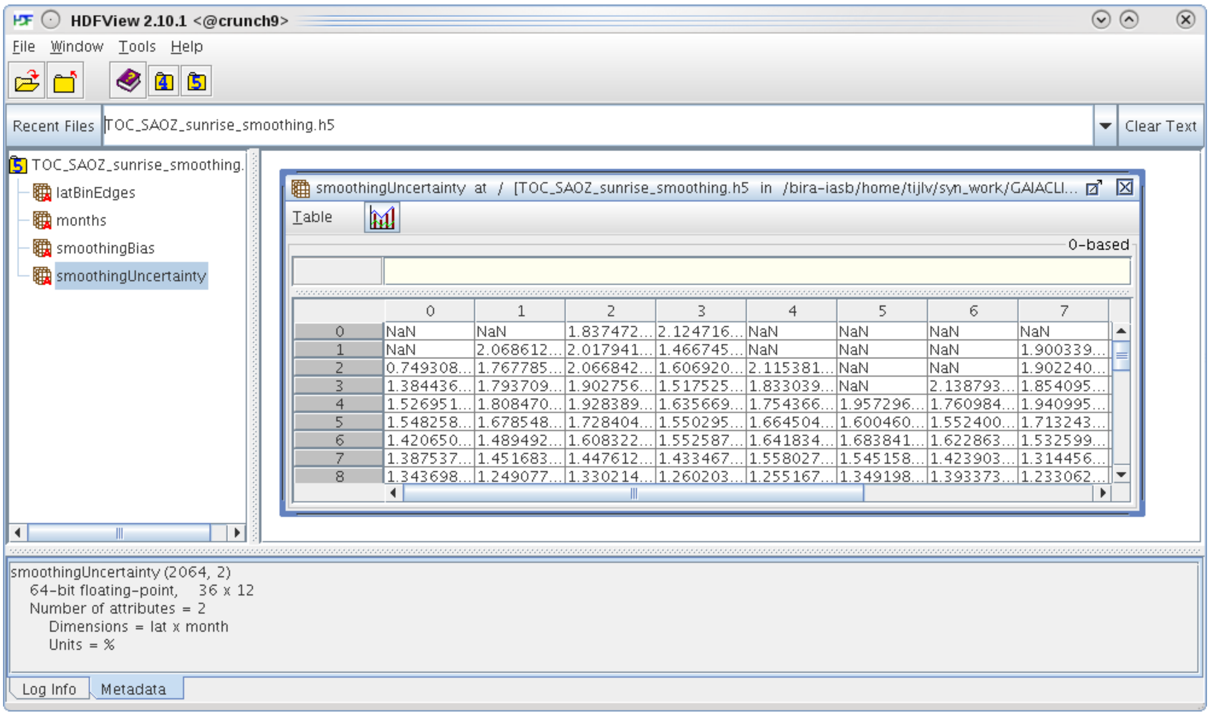

Figure 7: File contents for TOC_SAOZ_sunrise_smoothing.h5. The sunset file has exactly the same layout. Note the NaN values at locations and times when no sunrise or sunset occurs (i.e. polar night or day).

A.1.2. Direct sun: Dobsons and Brewers

Measurement principle

Direct-sun measurements such as those obtained with Dobson and Brewer instruments rely on the absorption by ozone of UV light in the Huggins band at 1 or 2 specific wavelength pairs. The ratio of the absorption strength in these pairs allows a direct estimate of the ozone column. The geometry of a direct-sun measurement is illustrated in Figure 8. It is clear from this graph that for non-zero solar zenith angles, the actual measurement sensitivity will not be located directly above the station. The resulting horizontal smoothing is determined by the absorption and scattering processes along the light-path.

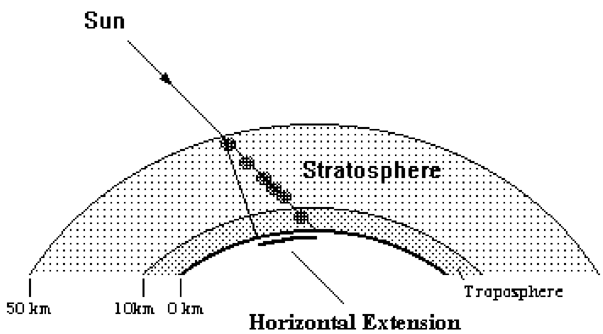

Figure 8: Schematic view of the optical path and corresponding horizontal extension of a direct-sun measurement of an ozone total column.

Smoothing properties

The horizontal smoothing properties of direct sun looking instruments were characterized and parametrized by Lambert and Vandenbussche (2011). In short, it is argued that the horizontal sensitivity of an individual UV/Vis direct-sun measurement can be estimated by first projecting the ozone profile on the slant optical path defined by the solar zenith and azimuth angles, and consequently projecting vertically onto the ground. As illustrated in Figure 9, withholding only that part of the projection which contains 90% of the total ozone column, the horizontal smearing is limited for low and moderate SZA, but increases steeply beyond 60-70˚ SZA. For daily means of direct-sun measurements, this 1-D estimate of the horizontal sensitivity can be extended to a 2-D polygon by taking into account the evolution of SZA and SAA throughout the day.

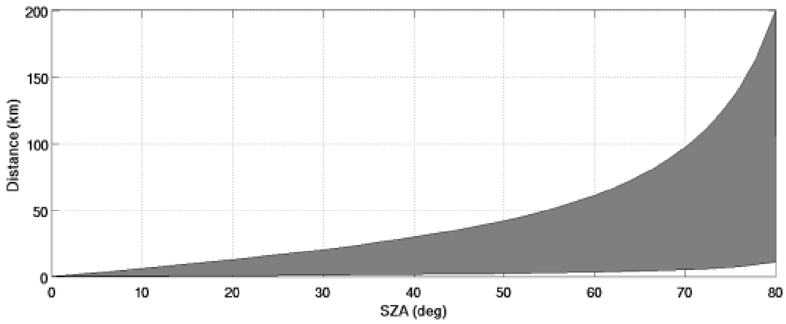

Figure 9: Horizontal extent of the air mass containing 90% of the total ozone column, expressed as distance from the station as a function of SZA.

LUT construction

Two LUTs were constructed: one for daily means of direct-sun measurements, where the mean is assumed to be based on a homogeneous sampling of that part of the day during which the sun is above 75˚ SZA, and one for individual direct-sun measurements, which requires as an additional dimension the SZA at the time of the measurement. Both LUTs are based on 10^6 simulated smoothing errors, at random latitude, longitude, and day-of-year, covering a total period of 4 years (2005–2008). The reanalysis fields used are those computed with IFS-MOZART in the context of MACC (now CAMS). Smoothing errors were then binned by latitude (5˚ resolution), month (all 4 years folded onto one “climatological” year), and for the individual direct-sun measurements also by SZA, where the bin sizes were chosen in such a way that the steep increase in corresponding air mass at higher SZA is sampled more densely than the small SZA regime. For these bins, both the spread on the smoothing errors (as a proxy of smoothing uncertainty) and the mean smoothing error (as a proxy of smoothing bias) were computed and tabulated. An excerpt from the LUT for individual direct-sun measurements is shown in Figure 10.

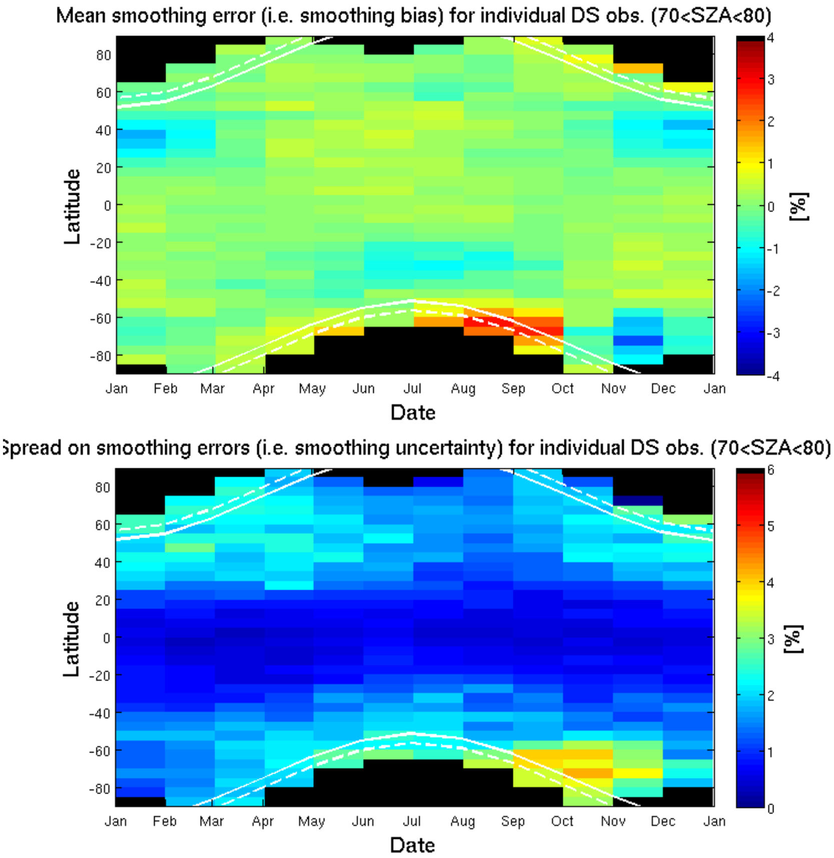

Figure 10: Visualization of the smoothing bias LUT (upper panel) and smoothing uncertainty LUT (lower panel) for direct sun (Dobson or Brewer) measurements of total ozone when the SZA is between 70˚ and 80˚. Black regions correspond to latitude/time-of-year combinations at which the sun doesn't reach angles between 70˚ and 80˚ SZA. The LUT files contain similar tables for a set of SZA ranges. For reference, the solid and dashed white lines represent 75˚ and 80˚ SZA (maximum during the day) respectively.

File description

The LUTs are saved as hdf5 files which are named *TOC_directsun_dm_smoothing.h5* and *TOC_directsun_sza_smoothing.h5.* The files are only 10kB and 51kB large. As described above, the influence quantities taken into account are latitude and month-of-year for the former file, and additionally also SZA for the latter. Figure 11 illustrates the file contents. The use of attributes should make these files self-explanatory.

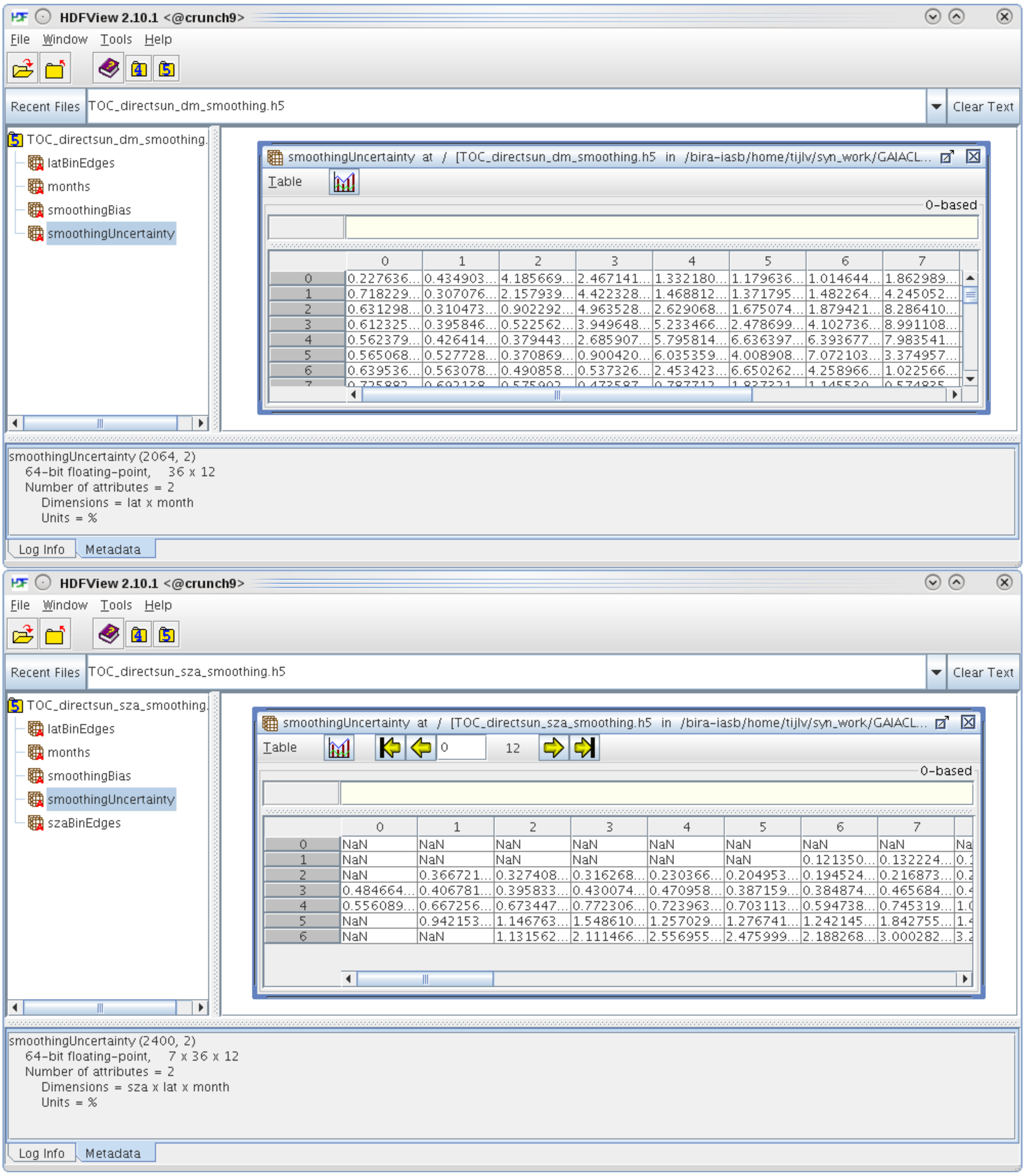

Figure 11: File contents for TOC_directsun_dm_smoothing.h5 (upper panel) and TOC_directsun_sza_smoothing.h5 (lower panel). Note the NaN values in the latter file at locations and times when the solar zenith angles of that SZA bin are not covered by the apparent motion of the sun, e.g. the smallest SZA are only reached at low latitudes.

A.1.3. GOME-2 on MetOp-A/B

Measurement principle

Ozone columns can be derived from the absorption features of O_3 in nadir measurements of backscattered UV-Vis light, for instance using the Differential Optical Absorption Spectroscopy (DOAS) technique. This is the method used for the operational products of key nadir UV-Vis sounders such as GOME on ERS-2 and GOME-2 on MetOp-A and MetOp-B.

Smoothing properties

While the horizontal resolution of such backscatter UV-Vis measurements is usually considered to be equal to the dimension of the ground pixel, Figure 12 illustrates that the information actually originates in two successive light paths: first between the sun and the scatterer and then between the scatterer and the instrument. Consequently, the effective horizontal resolution of such a measurement is not limited to the ground pixel, and it depends on influence quantities such as SZA, ground albedo, and vertical profile of the absorber. An extensive characterization and parametrization of this true horizontal resolution was performed by Vandenbussche et al. (2011) using a ray tracing method and the vertical averaging kernels projected on the ground along the light path.. In Verhoelst et al. (2015), it was found that for the swath width of GOME-2, a non-zero satellite viewing zenith angle (VZA) has only a minor impact on the horizontal dilution. For this reason, and to avoid an additional dimension in the LUTs, the VZA is hereafter assumed to be zero.

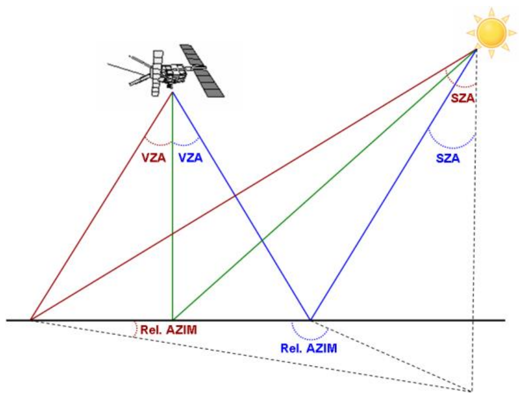

Figure 12: Nadir viewing geometry and the “dilution effect” of the measured information.

LUT construction

The LUT is based on 10^6 simulated smoothing errors, at random latitude, longitude, and day-of-year, covering a total period of 4 years (2005–2008). The actual overpass time of MetOp A/B (9:30am local solar time) was used to determine the SZA and SAA. These angles therefore do not require a dedicated dimension in the LUT. This does imply that these LUTs are not directly applicable to other nadir sounders with different overpass times, although it would be facile to produce new LUTs for such sounders. The ground pixel size was taken to measure the default 80x40km^2. The reanalysis fields used are those computed with IFS-MOZART in the context of MACC (now CAMS). Smoothing errors were then binned by latitude (5˚ resolution) and month (all 4 years folded onto one “climatological” year), and both the spread on the smoothing errors (as a proxy of smoothing uncertainty) and the mean smoothing error (as a proxy of smoothing bias) were computed and tabulated. The result is shown in Figure 13.

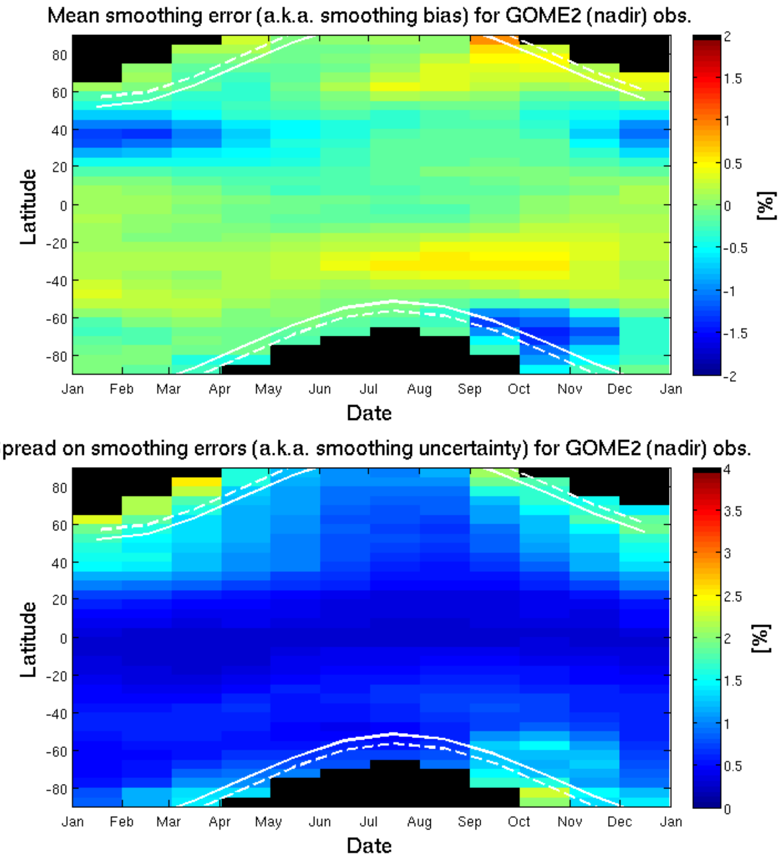

Figure 13: Visualization of the smoothing bias LUT (upper panel) and smoothing uncertainty LUT (lower panel) for GOME-2 measurements of total ozone, assuming an 80x40km^2 pixel size and a measurement in descending node at 9:30am local solar time. Black regions correspond to conditions not allowing nadir UV-Vis measurements, i.e. there is no sunlight. For reference, the solid and dashed white lines represent 75° and 80° SZA (maximum during the day) respectively.

File description

The LUT is saved as an hdf5 files which is named *TOC_GOME2_generic_smoothing.h5.* The file is only 10KB large. As described above, the influence quantities taken into account are latitude and month-of-year. Figure 14 illustrates the file contents. The use of attributes should make this file self-explanatory.

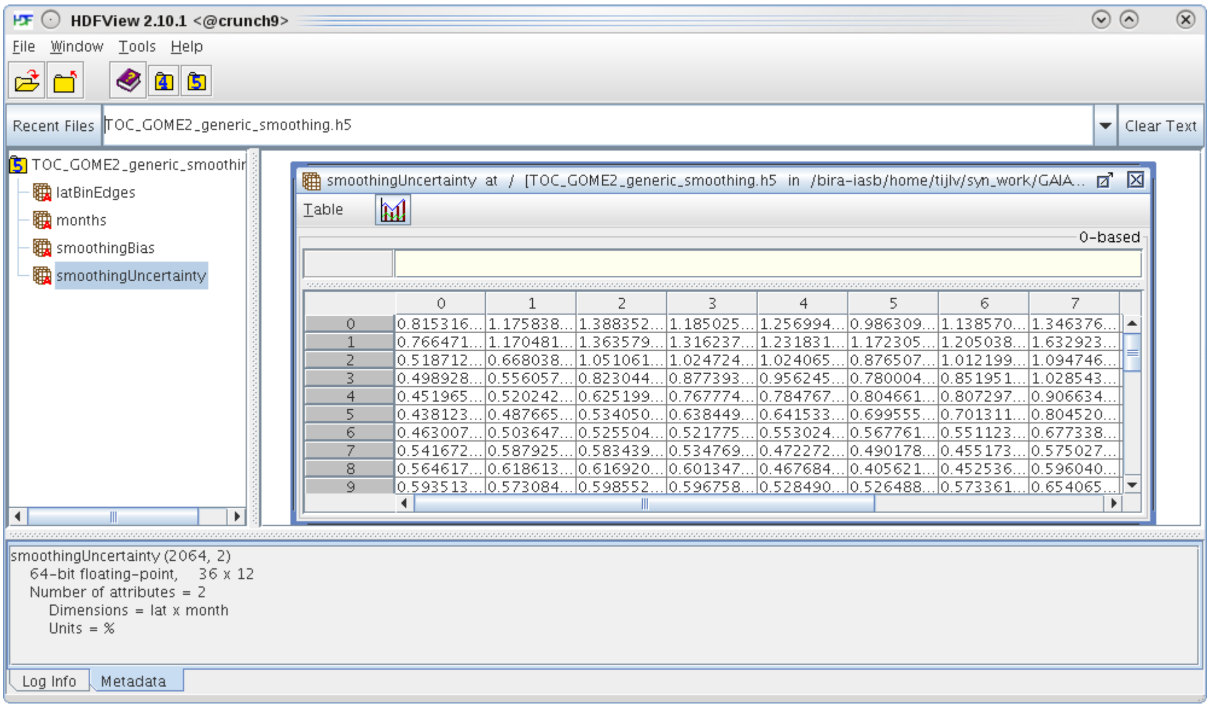

Figure 14: File contents for TOC_GOME2_generic_smoothing.h5.

A.2. Co-location uncertainties

In the previous sections, we have described the LUTs that quantify the smoothing uncertainty and bias related to a single measurement of the total ozone column. When comparing two measurements from different instruments, an additional source of errors is the mismatch in nominal measurement location and time. In practice, the amount of mismatch is governed by the co-location criteria that were used to construct the measurement pairs. To be able to add this uncertainty term to the total uncertainty budget of a comparison, LUTs were constructed that contain for a wide range of spatial and temporal co-location constraints the expected additional spread on the differences due to co-location mismatch. A single co-location error is defined as:

$$\varepsilon_{co-location} = \frac{\text{TOC }(lat1, lon1, t1) - \text{TOC}(lat2, lon2, t2)}{\text{TOC}(lat2, lon2, t2)}.$$

In the case of satellite validation with ground-based reference measurements, index “1” would refer to the satellite measurement and index “2” to the ground-based reference. Typical co-location criteria put constraints on *dr*, the great-circle distance between (*lat1,lon1*) and (*lat2,lon2*), and on *dt* = |*t2-t1*|.

LUT construction

The LUT was constructed by simulating roughly 10^9 co-location mismatch errors, distributed over the entire globe and covering 6 years, and this for each possible (listed) co-location criterion (specifically: choosing random separations in space and time, uniformly distributed up to the co-location criterion, where the spatial distribution is uniform in 2D). Verhoelst et al. (2015) found that for a typical TOC validation exercise comparing nadir UV-Vis satellite measurements with ground-based direct sun or ZSL-DOAS measurements, this source of uncertainty mostly dominates over the smoothing (difference) errors of the individual measurement systems, at least for co-location criteria less stringent than 100km maximum spatial separation and 3h maximum temporal separation. Moreover, it was found that the impact depends not only on latitude and time-of-year, but also on longitude. For these reasons, this LUT is sampled at high resolution (1˚x1˚) in latitude and longitude. Spatial co-location criteria are sampled as 0.5, 1, 3, 5, 10 degrees, and the temporal criteria as 0.1. 0.3, 1, 3, 10, 23 hours. Typical criteria for total ozone column validation work would be 3 degrees and 3 hours. The values read from this LUT represent the additional spread on an ensemble of pairs due to co-location mismatch within the given co-location criteria. As such, they need to be used when constructing the final comparison statistics. They cannot be used to generate a co-location uncertainty estimate for a single comparison pair. LUTs to be used in that sense are in progress and will be available in the final set of tools (D3.7).

File description

To keep the file size limited, a separate file was generated for each month of the year, named TOC_colocUncert_Month*xx*.h5. Figure 15 illustrates the contents of one such a file. The LUT has 4 dimensions: longitude, latitude, temporal co-location constraint, and spatial co-location constraint.

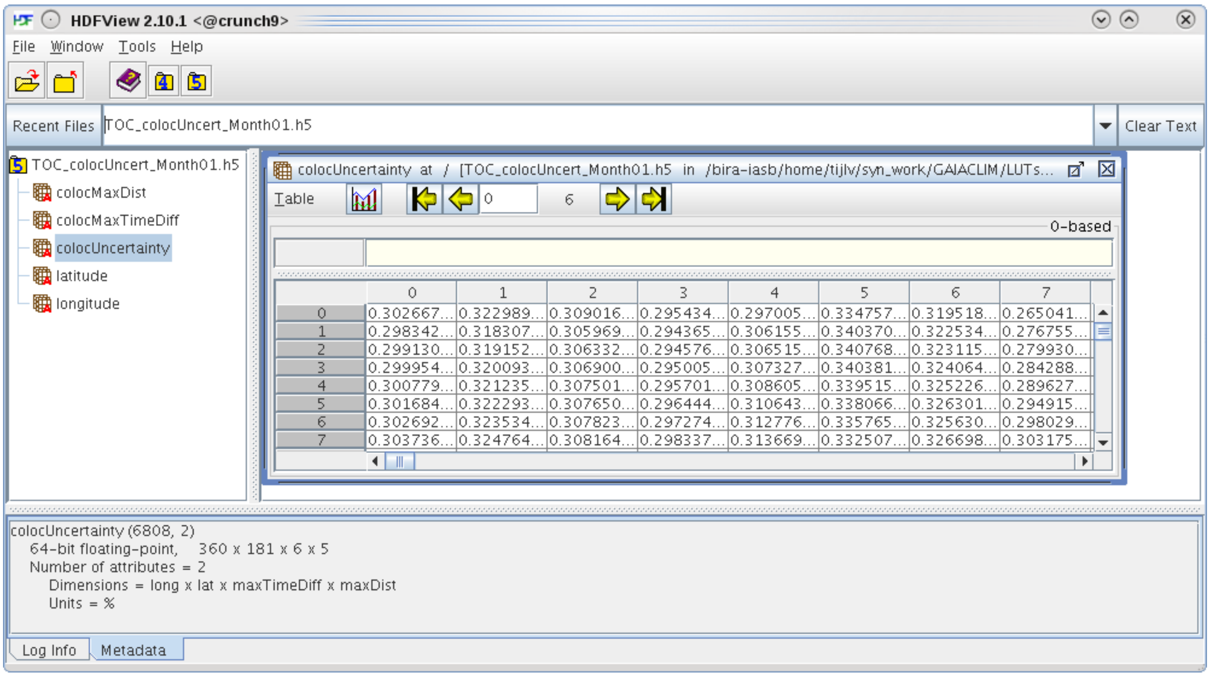

Figure 15: File contents for TOC_colocUncert_Month01.h5.

Prospects on co-location LUTs

As already mentioned, the LUTs described above represent the expected additional spread due to co-location mismatch on a large set of differences where the pairs represent different spatial and temporal offsets within the maxima allowed for by the co-location criteria. Work is ongoing to generate LUTs for a given spatial and temporal distance, which allow an estimate of co-location mismatch uncertainty for a single comparison pair.

A.3. Recommended use

The LUTs presented above can be used in two ways:

- Either they are used as true “Look-Up Tables”, meaning that one searches for the entry that corresponds to the correct (latitude, month, sza, ...) or (colocMaxDist, colocMaxTimeDiff, latitude, longitude, month) bin and reads out the corresponding uncertainty (spread on random errors) and – if available – bias (systematic error).
- Or, alternatively, and possibly providing more accurate results, one can interpolate the LUTs to the actual (lat, month, sza, ...) coordinate. The underlying variables are fairly continuous functions so interpolation methods can be both linear or higher-order. The grid steps in the co-location criteria in the co-location LUT don’t actually represent bins, and it is therefore advised to use the closest value. Interpolation can in principle still be done in latitude and longitude, but the spatial resolution is probably already at the limit of the information content of the simulations.

The practical use of these LUTs in the GAIA-CLIM VO will depend mostly on the functionalities developed within WP5. Three main use cases can be anticipated:

1. When TOC measurements are visualized, e.g. as time series, smoothing uncertainties can be added to the error bars (quadratic sum), or shown as additional error bars with a different colour/marker. This includes the possibility to differentiate between random and systematic components.
2. LUTs can be shown as stand-alone figures, to aid the user in selecting geographical regions or time periods for which smoothing and co-location uncertainties are within certain limits. This can be in the form of global or regional maps (with stations indicated), or as time series at a selected location (existing station, hypothetical new station).
3. The LUTs can be used for data filtering purposes, e.g. to extract only measurements for which the expected smoothing uncertainty is below a certain threshold, or to use only those co-located pairs for which the co-location mismatch is below a certain threshold.

For easy and fast visualization, it is suggested here to complement the measurement data with (interpolated) LUT data already at the point of data ingestion, e.g. as additional uncertainty fields. For the mapping functionalities, also the full LUTs need to be stored in the VO database.

Acronyms

AATSR	Advanced Along Track Scanning Radiometer
AERONET	Aerosol Robotic Network
AOD	Aerosol Optical Depth
CALIOP	Cloud-Aerosol LIdar with Orthogonal Polarization
CALIPSO	Cloud-Aerosol LIdar and Infrared Pathfinder Satellite Observations
CEU	Central Europe
DIS	Data Ingestion Script
EARLINET	European Aerosol Research LIdar NETwork
ECV	Essential Climate Variable
GRUAN	Global climate observing system Reference Upper-Air Network
GUI	Graphical User Interface
IASI	Infrared Atmospheric Sounding Interferometer
LIDAR	LIght Detection And Ranging
LUT	Look-Up Table
OSSSMOSE	Observing System of Systems Simulator for Multi-mission Synergies Exploration
RAOB	RAdiosonde OBservations
TOC	Total Ozone Column
VO	Virtual ObservatoryAOD Aerosol Optical Depth

CALIOP	Cloud-Aerosol LIdar with Orthogonal Polarization
CALIPSO	Cloud-Aerosol LIdar and Infrared Pathfinder Satellite Observations
CEU	Central Europe
DIS	Data Ingestion Script
EARLINET	European Aerosol Research LIdar NETwork
ECV	Essential Climate Variable
GRUAN	Global climate observing system Reference Upper-Air Network
GUI	Graphical User Interface
IASI	Infrared Atmospheric Sounding Interferometer
LIDAR	LIght Detection And Ranging
LUT	Look-Up Table
OSSSMOSE	Observing System of Systems Simulator for Multi-mission Synergies Exploration
RAOB	RAdiosonde OBservations
TOC	Total Ozone Column
VO	Virtual Observatory