Proceedings for the 2018 EUMETSAT Meteorological Satellite Conference, 17-21 September 2018, Tallinn, Estonia

QUALITY ASSESSMENT OF QA4ECV CLIMATE DATA RECORDS OF ATMOSPHERIC COMPOSITION: TERMINOLOGY, METHODOLOGY AND APPLICATION TO TROPOSPHERIC NO_2, HCHO AND CO FROM THE GOME-2, IASI AND OMI SATELLITES

Steven Compernolle[1], Jean-Christopher Lambert[1], Tijl Verhoelst[1], José Granville[1], Daan Hubert[1], Arno Keppens[1], Sander Niemeijer[2], Bruno Rino[2], Gaia Pinardi[1], Steffen Beirle[3], Folkert Boersma[4], Cathy Clerbaux[5], Pierre Coheur[6], Isabelle De Smedt[1], Henk Eskes[4], Maya George[5], François Hendrick[1], Alba Lorente[7], Joanne Nightingale[8], Enno Peters[9], Andreas Richter[9], Jos van Geffen[4], Michel Van Roozendael[1], Thomas Wagner[3], and Huan Yu[1]

[1]Royal Belgian Institute for Space Aeronomy (BIRA-IASB), Ringlaan 3, 1180 Brussels, Belgium,
[2]s[&]t Corporation, Delft, The Netherlands,
[3]Max Planck Institute for Chemistry (MPIC), Mainz, Germany,
[4]Royal Netherlands Meteorological Institute (KNMI), De Bilt, The Netherlands,
[5]LATMOS/IPSL, Sorbonne Université, UVSQ, CNRS, Paris, France
[6]Spectroscopie de l'Atmosphère, Chimie Quantique et Photophysique, Université Libre de Bruxelles (U.L.B.), Brussels, Belgium
[7]Wageningen University, Meteorology and Air Quality Group, Wageningen, the Netherlands,
[8]National Physical Laboratory (NPL), Teddington TW11 0LW, UK
[9]Institute of Environmental Physics, University of Bremen (IUP-B), Bremen, Germany

Abstract

Two major objectives of the EC FP7 project Quality Assurance for Essential Climate Variables (QA4ECV) were the development of a generic quality assurance system dedicated to C3S, and the demonstration of this generic quality assurance system on 6 pilot ECVs. In this work, we describe (i) the terms and definitions applicable to the quality assurance that have been agreed within QA4ECV, (ii) the generic validation protocol, (iii) the QA4ECV atmosphere ECV validation server, and (iv) quality assessment results for the 3 atmospheric QA4ECV climate data records (NO_2, HCHO and CO).

1. INTRODUCTION

Atmospheric nitrogen dioxide (NO_2), formaldehyde (HCHO) and carbon monoxide (CO) play a key role both in air quality and as precursors to Essential Climate Variables (ECV) like ozone (O_3), carbon dioxide (CO_2) and aerosols.

For these three trace gases the EC FP7 project Quality Assurance for Essential Climate Variables (QA4ECV, 2014-2018) has produced quality assured Climate Data Records (CDRs) from several satellite-based instruments: ERS-2 GOME, Envisat SCIAMACHY, Metop-A GOME-2, and EOS-Aura OMI for NO_2 [1] and HCHO [6], and Metop-A IASI for CO [14]. Here we report on a comprehensive quality assessment of the following CDRs: QA4ECV NO_2 from OMI and GOME-2A (tropospheric and stratospheric column), QA4ECV HCHO from OMI and GOME-2A (tropospheric column), and IASI-A FORLI CO (total column).

A major objective of the QA4ECV project was to prototype a generic quality assurance system dedicated to the Copernicus Climate Change Service (C3S), applicable virtually to all ECVs. This QA system is described in Nightingale et al. [19]. Part of this, a harmonized list of standard terms and definitions, relevant to the quality assurance of ECV Climate Data Records, has been compiled from international standards published by normalization institutes and complemented by community practices from space agencies. The quality assessment of the satellite CDRs for atmospheric ECVs follows a generic validation protocol virtually applicable to all atmospheric composition data records. It results in a wide range of quality indicators enabling potential users to verify the fitness of the data records for their own purpose. This validation protocol has been implemented in the prototype QA4ECV Atmosphere Validation Server (QA4ECV-AVS), building upon the heritage of the Multi-TASTE versatile satellite validation system and its operationalization in the context of EUMETSAT AC-SAF and ESA Multi-TASTE/CCI. An automated version of this validation server, accessible online, constitutes the backbone for the Validation Data Analysis Facility (VDAF) of the Sentinel-5p Mission Performance Centre (MPC), in charge of the routine validation service for the TROPOMI operational atmospheric data products.

Another achievement of the QA4ECV project has been to demonstrate the generic quality assurance system on six pilot CDRs, three from the land domain (surface albedo, leaf area index (LAI) and fraction of photosynthetically active radiation (FAPAR)) and three from the atmospheric domain (tropospheric columns of NO_2, HCHO and CO). The validation protocol and the QA4ECV-AVS have been applied on the QA4ECV atmospheric climate data records using, as reference measurements, the ground-based DOAS UV-Visible and FTIR data acquired at several sites of the Network for the Detection of Atmospheric Composition Change (NDACC).

In the following sections, we describe (i) the terms and definitions applicable to the quality assurance that have been agreed upon within QA4ECV, (ii) the generic validation protocol, (iii) the QA4ECV atmosphere ECV validation server, and (iv) validation results for the 3 atmospheric ECVs.

2. TERMS AND DEFINITIONS APPLICABLE TO THE QUALITY ASSURANCE OF CLIMATE DATA RECORDS

At the early stage of the QA4ECV project, it became clear that there was a need for a consistent set of terms and definitions pertaining to the quality assurance of climate data records. It was decided to use and further develop the CEOS WGCV list that resulted from the GEO-CEOS 2008 workshop for a GEOSS Data Quality Strategy, and which got several updates since then. It encompasses terms from the metrology, remote sensing, monitoring and modelling domains, and was implemented (at various stages of development) in QA/validation protocols for the EC MACC-I/II and PASODOBLE pioneering projects for the Copernicus Atmosphere Monitoring Service (CAMS), the ESA Multi-TASTE and CCI validation of multi-sensor and multi-species long-term data records, and of course QA4ECV. In complement, the EU Horizon2020 GAIA-CLIM project pursued consistency and compliance with standards, specifically regarding terminology in metrology [23].

The set of terms and definitions established for QA4ECV can be found at [5]. This document contains a selection of standard terms and definitions relevant to the quality assurance of ECVs data records. It reproduces appropriate terms and definitions published by normalization bodies, mainly BIPM/JCGM/ISO in their International Vocabulary of Metrology (VIM) [24] and Guide to the Expression of Uncertainties (GUM) [12]. It also reproduces selected terms and definitions related to the quality assurance and validation of Earth Observation (EO) data, available publicly on the ISO website and on the Cal/Val portal of the Committee on Earth Observation Satellites (CEOS) [2]. Several of those terms have been recommended by CEOS for the GEO-CEOS Quality Assurance framework for Earth Observation (QA4EO) [20] and, as such, are applicable to virtually all Copernicus data sets of EO origin. Another important source is ISO:9000 [15], containing generic terms on quality.

Below, we illustrate with a few case studies that the choice of nomenclature is not trivial and that improper use can lead to inconsistent communication of QA requirements and results.

2.1. Case studies

Use of terms ‘error’ and ‘uncertainty’:

This is arguably one of the best-known examples where every-day usage often deviates from the ISO norm set by VIM [24] and GUM [12]. A core problem is that in every-day usage *error* is sometimes used – often in the same text – to designate (i) the difference between measured value and true value, and (ii) a spread measure of likely values of the true value. GUM and VIM make proper distinction of both concepts, reserving the term *error* for case (i), and *uncertainty* for case (ii).

GCOS user requirements and the concept of accuracy:
GCOS establishes user requirements for ECV data products. Unfortunately, the terminology used in different GCOS documents is sometimes inconsistent, and not conform to ISO standards. The term *accuracy* is a case in point. The 'Systematic Observation Requirements for Satellite-based Products for Climate' [10] states

The user requirement for accuracy is a requirement for closeness of agreement between product values and true values. […]

On the other hand, the 'Guideline for the Generation of Datasets and Products Meeting GCOS Requirement' [8] defines *accuracy* as

Measured by the bias or systematic error of the data, i.e. the difference between the short-term average measured value of a variable and the true value. […]

The VIM [24] defines accuracy as *closeness of agreement between a measured quantity value and a true quantity value of a measurand*. The first GCOS accuracy definition (from [10]) is quite in line with the VIM definition, but the second GCOS definition (from [8]) is in clear contradiction with the VIM definition. The second GCOS definition is rather related to the VIM concepts *trueness* and *bias*. It goes without saying that the use of different definitions for the same term leads to confusion about the correct interpretation of the GCOS user requirements.

More recently, GCOS has replaced the term accuracy in favor of *uncertainty* [11] in the ECV product requirement tables (Annex A of the GCOS 2016 Implementation Plan [11]), with reference to the WMO 'Guide to Meteorological Instruments and Methods of Observation' [25], which itself draws from VIM [24] and GUM [12].

The concept 'structural uncertainty':
The concept *structural uncertainty* is described by Thorne et al. [21] as follows:

[…] a number of seemingly physically acceptable methodologies for constructing a dataset from the same raw data will yield a range of solutions rather than converge to a single point solution […] structural uncertainty arises through the choice of approach.

The term 'structural uncertainty' has been used in several works (including QA4ECV papers, see e.g., [18]). However, this concept can actually be directly related to the GUM's concept of *uncertainty of measurement method* ([12], section F.2.5):

Uncertainty associated with the method of measurement, as there can be other methods, some of them as yet unknown or in some way impractical, that would give systematically different results of apparently equal validity.

It is recommended that the ISO GUM reference is used to describe this concept.

Sensitivity versus representativeness:
Sometimes, the term *area (or volume) of representativeness* is used for the area (or volume) where the measuring system is sensitive to changes (see e.g., [13]). *Sensitivity* is defined in VIM [24] as *the quotient of the change in an indication of a measuring system and the corresponding change in a value of a quantity being measured*. On the other hand, *area (or volume) of representativeness* is defined as *the area (or volume) in which the concentration does not differ from the concentration at the station by more than a specific range* [17]. A system measuring a concentration can have a small area of sensitivity, but, if the concentration field is sufficiently flat, the corresponding measurement can have a high area of representativeness.

3. GENERIC VALIDATION PROTOCOL

The generic validation protocol builds further on the validation chain described in Keppens et al. [16]. Its different steps are outline in ***Table** 1*. For more information about the individual steps, see [3].

Designing (the validation process and expected Quality Indicators) STEP 1: Translation of user requirements into validation requirements - Collection of user requirements: random and systematic component of total uncertainty, long-term stability [e.g., %/decade], horizontal resolution [km], vertical resolution [km], geographical domain [e.g.,

latitude/longitude], vertical range [km], sampling frequency [overpass/day] or repeatability period [day], solar local time [hh:mm:ss]…
- Collection of known/potential features of the measurement and retrieval to be considered, as they might impact the quality of the data: sensitivity to solar and viewing angles, sensitivity to vertical profile of target constituent, sensitivity to surface albedo, sensitivity to cloud cover and parameters, sensitivity to temperature contrast…
- Translation of user requirements and sensitivities into traceable Quality Indicators: bias (mean or median) and variance (standard deviation or interpercentile) with respect to reference measurements, their variation over the user-defined domain and range, need to assess Backus-Gilbert spread of vertical averaging kernels as an estimate of vertical resolution…

Sub-setting (the satellite dataset)

STEP 2: Satellite data selection, filtering and post-processing
- Selection of satellite-based ECV data based on user-defined criteria for the spatial (geographical area/station) domain, vertical range and temporal domain.
- Application of data filtering as recommended by the data providers (e.g., no use of data under a threshold of thermal contrast or beyond a threshold of solar zenith angle, above a given value of DOAS fit RMS, over South Atlantic Anomaly…)

Characterizing (the resulting satellite dataset)

STEP 3: Data content study (DCS) of satellite-based ECV dataset
- Characterization of spatial, vertical and temporal domain over which data are available after selection, filtering and post-processing.
- Production of information on data availability and its variations: Identification of areas/periods/altitudes where data are not or less available, identification of patterns, cycles and trends in data representativeness etc.
- Means: visualization of datasets and associated statistical indicators.

STEP 4: Information content study (ICS) of satellite-based ECV dataset
- Characterization of atmospheric information contributed by the measurement, as opposed to non-measured a priori extracted from a climatology or imposed from retrieval constraints
- The output of this study must enable identification of the altitude range where the comparisons with reference measurements make sense, and must be reported on comparison graphs.
- Means: Linear analysis of vertical averaging kernels and covariance; derivation and study of diagnostics like Degree of Freedom of the System (DFS), vertical sensitivity, Measurement quality quantifier (MQQ), Backus-Gilbert spread…
- Main output Quality Indicators: sensitivity estimate, vertical resolution estimate and uncertainty on altitude registration, with their variations.

Co-locating (satellite and reference data)

STEP 5: Selection and characterization of correlative data
- Selection of ground-based reference ECV data based on user-defined criteria for the spatial (geographical area/station) domain, vertical range and temporal domain.
- Consideration of the vertical sensitivity of satellite data derived from STEP 4.
- DCS and ICS of the reference data. E.g. characterization of influence quantities on the reference data, consideration of the vertical sensitivity of reference data.

STEP 6: Identification and characterization of co-located data pairs
- Identification of co-located pairs based on different selection criteria, ranging from the classical radius around the ground-based station, to more sophisticated methods based on the use of observation operators describing the air mass really contributing to the measurement.
- Loop on all ground-based data, to generate a bank of co-located data pairs.
- Characterization of spatial, vertical and temporal domain over which co-located data pairs are available.
- Purpose: Production of information on the domain of validity of the data comparison results: Identification of areas/periods/altitudes where co-locations are available.
- Means: visualization of datasets and associated statistical indicators.
- Repetition of the DCS and ICS, but now solely on the co-located satellite data points. Comparing the plots with those produced during DCS and ICS learns how representative the co-located satellite data set is with respect to the original set.

Harmonizing (satellite and reference data)

STEP 7: Data homogenization
- Harmonization of representation systems and units.
- Resampling and smoothing vertically to converge towards a common vertical grid.
- Means: algorithms and routines detailed in the DPM [3].

Comparing (satellite and reference data)

STEP 8: Data comparison process
- Pair-by-pair bias: Calculation of the absolute difference between the satellite-based ECV data and the co-located reference ECV measurement, co-located data pair by co-located data pair,
- Loop on all co-located data pairs to generate a bank of comparison data.
- Parallel handling of uncertainties and metadata associated with the satellite-based and ground-based data

STEP 9: Derivation of statistical comparison Quality Indicators

- First examination of the comparison results to establish potential for statistical analysis. E.g., looking at histograms of the absolute differences, do the biases follow a normal distribution, for which mean and standard deviation estimates make sense, or do they follow a more complex distribution which needs further classification of the results? Looking at time series, can we detect a drift which will alter the derivation of a mean bias and increase the spread of the absolute differences over time?
- In addition to the mean and standard deviation, calculation also of third and fourth momentum of the distribution to quantify skewing and flattening/sharpening of the distribution with respect to a Gaussian.
- To check if outliers alter significantly the validity of mean and standard deviation estimates, calculation of median and interpercentile estimates.
- If needed, further classification of the comparison results to enable their statistical analysis and the detection of dependences on influence quantities (like solar zenith angle, thermal contrast, clouds etc.)
- Derivation of statistical quantities like mean bias, spread, stability, dependences and their variation with time, latitude, pollution level…
- All operations make sense only over the identified domain and range of sensitivity.
- Note that there are other Quality Indicators than the statistical quantities on data comparisons derived here, e.g., results of data content studies from STEP 3, estimates of vertical and horizontal resolution from STEP 4 etc.

Monitoring and reporting

STEP 10: Production of user-oriented report

- Collection and transfer of fit-for-purpose Quality Indicators from the different STEPS into user-oriented reports and monitoring facilities.
- Web-based report accompanying the traceability chain of the selected validation process.

Verifying (fitness-for-purpose of the data)

STEP 11: External verification of compliance with user requirements

- Examination of the comparison results and other Quality Indicators: Verification that the estimated dispersion of satellite data with respect to reference data does not exceed user requirements in terms of uncertainty, stability, etc.
- Verification that estimates of sensitivity range, resolution and other characteristics comply with user requirements.
- Verification that features, patterns, cycles, drifts detected in the satellite data set do not exceed limits and thresholds fixed by user requirements.
- *A posteriori* assessment of the value and appropriateness of reference data for the validation process.

***Table** 1*. Steps of the generic validation protocol, as applied to the ECV products NO_2, HCHO and CO. Taken from [3].

4. ATMOSPHERE ECV VALIDATION SERVER

Starting from the generic validation protocol, detailed validation schemes were developed for NO_2 (tropospheric and stratospheric column), HCHO (tropospheric column) and CO (tropospheric column), which are described in the Detailed Processing Model [3]. Based on the latter document, the QA4ECV Atmosphere ECV Validation Server (QA4ECV-AVS) was demonstrated on the following QA4ECV products: QA4ECV NO_2 (OMI and GOME-2A), and QA4ECV HCHO (OMI and GOME-2A). Both are compared to QA4ECV NDACC MAXDOAS products. The comparison results can be inspected at https://qa4ecv-dev.stcorp.nl/ (see also **Figure 1**).

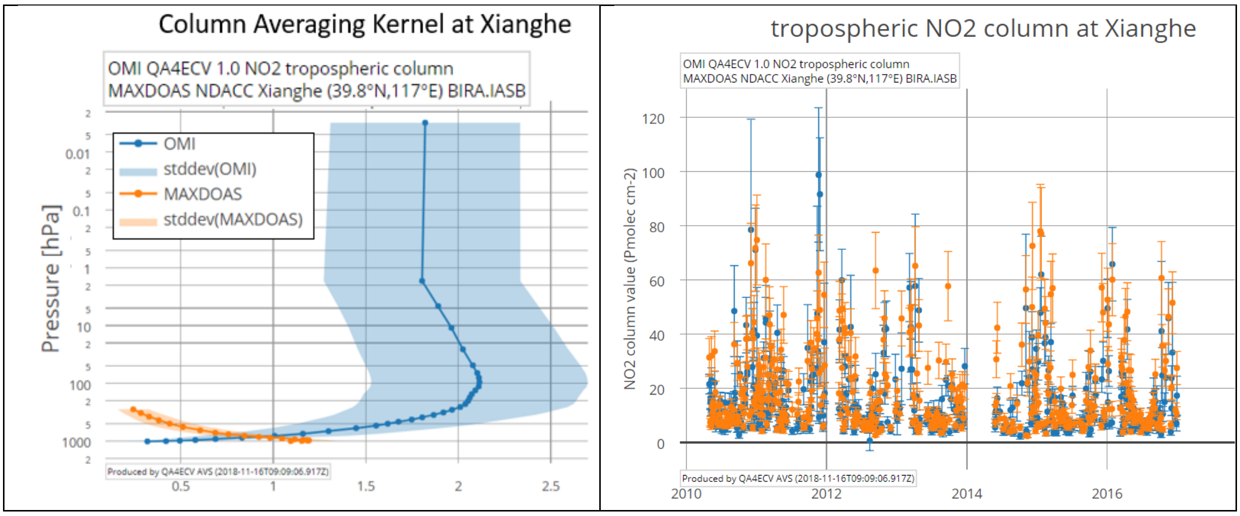

Figure 1. **Figures generated by the QA4ECV Atmosphere ECV Validation Server, for the comparison QA4ECV NO2 OMI vs NDACC MAXDOAS at Xianghe. Left: column averaging kernel (indicating sensitivity of retrieved column to true profile). This is a typical plot for an information content study (see STEP 4 in Table 1). Right: Time series of satellite and ground-based co-located data points, a typical plot for in the data comparison process (see STEP 8 in Table 1).**

5. QUALITY ASSESSMENT RESULTS

It was verified whether the QA4ECV data products comply with user requirements formulated in the GCOS 2016 implementation plan [11]. Here we focus on the required measurement uncertainty; more information can be found in the QA4ECV Product Validation and Intercomparison Report [4]. Reference data (NDACC QA4ECV MAXDOAS for NO2 and HCHO, NDACC QA4ECV FTIR for CO) are described in [13] and [7].

ECV	Required measurement uncertainty
NO_2 tropospheric column	MAX(20%, 0.8 10^{15} molec/cm^2)
HCHO tropospheric column	MAX(30%, 1.1 10^{15} molec/cm^2)
CO tropospheric column	MAX(20%, 5.4 10^{17} molec/cm^2)

***Table 2*. Target requirements on measurement uncertainty, taken from the GCOS 2016 implementation plan [11].**

For the products QA4ECV NO_2 1.1 and QA4ECV HCHO 1.1, the bias and/or the comparison spread mostly exceed the requirement on measurement uncertainty. However, it should be noted that other errors contribute to these quality indicators: (i) errors in the reference measurement, and especially (ii) errors associated with the data comparison method, i.e., errors not related to the actual measurement uncertainties but due to e.g., irreducible co-location mismatches in space and time. The latter are addressed in the EU Horizon2020 project GAIA-CLIM (see [23] and [22]). For IASI FORLI CO, the requirement on measurement uncertainty is met.

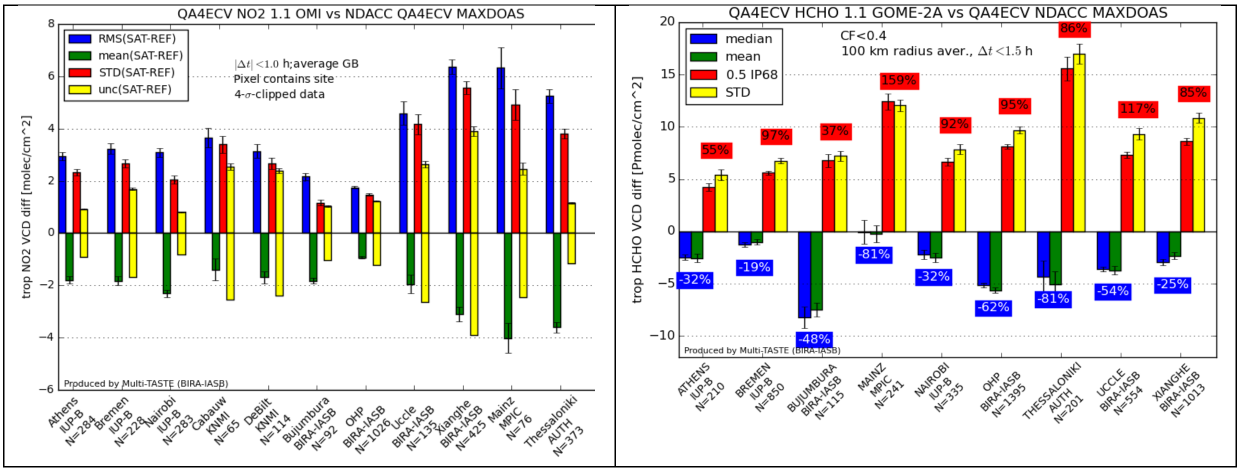

Figure 2. **Left: per-station overview of comparison results for QA4ECV NO2 1.1 OMI vs NDACC QA4ECV MAXDOAS. The following statistical quality indicators (QI) based on SAT-REF difference (after 4-sigma clipping the data) are provided: root-mean-square, mean and standard deviation (STD). The combined ex-ante uncertainty is also provided. Right: per-station overview of comparison results for QA4ECV HCHO 1.1 GOME-2A vs NDACC QA4ECV MAXDOAS. The following statistical QI, derived from the difference, are provided: median, mean, STD and half of the 68% interpercentile (½IP68). Regarding QI derived from relative difference, median and ½IP68 are indicated in percentages.**

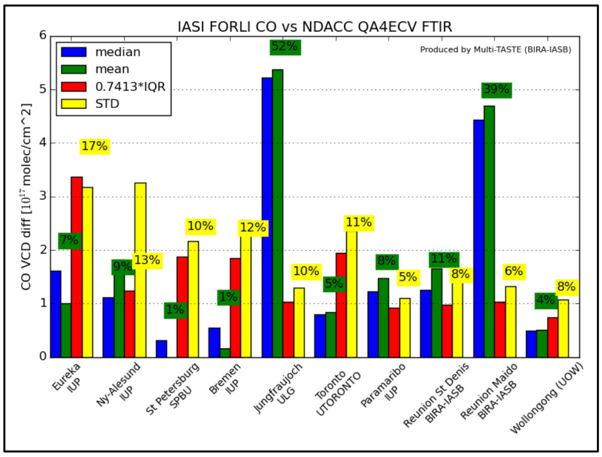

Figure 3. **Per-station overview of comparison results for IASI FORLI CO vs NDACC QA4ECV FTIR. The following statistical QI are derived from the SAT-REF difference: median, mean, normalized inter quartile range (=0.7412IQR) and STD. Regarding QI derived from relative difference (SAT-REF)/REF, mean and STD are indicated in percentages.**

6. CONCLUSION

The FP7 QA4ECV project was a successful pioneering implementation of the GEO-CEOS QA4EO framework for the Copernicus Climate Change Service (C3S). A set of standard terms and definitions, applicable to the quality assurance of ECV data records, was first established [5]. Advances in the generic validation protocol and specific validation techniques were integrated in the QA4ECV prototype Atmosphere ECV Validation Server and the Multi-TASTE versatile satellite validation system, and applied to the following CDRs produced by the project: QA4ECV HCHO, NO_2 and CO from GOME-2A, OMI and IASI-A. The technology of the QA4ECV Atmosphere ECV Validation Server has been transferred successfully to the operational validation facility for the Sentinel-5 Precursor satellite, with extension to many other data products relevant to climate, air quality, stratospheric ozone, ultraviolet radiation and volcanic hazards: methane (CH_4), O_3 (tropospheric column, profile, total column), sulfur dioxide (SO_2), aerosols, clouds... QA4ECV achievements have also been instrumental in improving data quality assessment activities for several C3S data procurement services. Despite the noticeable achievements accomplished within the QA4ECV project and transferred to several Copernicus components, the interpretation of data comparisons, and especially between satellite and ground-based measurements, remains a difficult task due to errors associated with irreducible co-location mismatches. Further work is needed with respect to the harmonization of measurement uncertainty – both its calculation and its expression – and to the closure of the data comparison error budget.

ACKNOWLEDGEMENTS

This work was funded by the EU FP7 Project Quality Assurance for Essential Climate Variables (QA4ECV), grant no. 607405, and the H2020 project GAIA-CLIM (Ares(2014)3708963/Project 640276).

7. REFERENCES

[1] Boersma, K. F.; Eskes, H. J.; Richter, A.; De Smedt, I.; Lorente, A.; Beirle, S.; van Geffen, J. H. G. M.; Zara, M.; Peters, E.; Van Roozendael, M.; Wagner, T.; Maasakkers, J. D.; van der R. J., A; Nightingale, J.; De Rudder, A.; Irie, H.; Pinardi, G.; Lambert, J.-C. & Compernolle, S., (2018) Improving algorithms and uncertainty estimates for satellite NO_2 retrievals: Results from the Quality Assurance for Essential Climate Variables (QA4ECV) project. Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2018-200.

[2] CEOS - Committee on Earth Observation Satellites (CEOS): Terms and Definitions and other documents and resources publicly available on http://calvalportal.ceos.org/

[3] Compernolle, S.; Lambert, J.-C.; and Niemeijer, S., (2016) Prototype QA/Validation Service for Atmospheric ECV Precursors : Detailed Processing Model Version 2. QA4ECV EC FP7 project deliverable D2.5. Available at http://www.qa4ecv.eu/sites/default/files/QA4ECV_BIRA-IASB_D-2-5_AVS-DPMv2_20160623.pdf

[4] Compernolle, S.; Lambert, J.-C., et al., (2017) Report with quality assessment of Atmosphere ECV data products. Product Validation and Intercomparison Report for: QA4ECV NO2 OMI 1.0, QA4ECV HCHO OMI 1.0, IASI FORLI CO v20100815+ v20140922, QA4ECV EC FP7 project deliverable D5.6. Available at http://www.qa4ecv.eu/sites/default/files/Attachment_0.pdf.

[5] Compernolle, S., & Lambert, J.-C. (2017). Standard terms and definitions applicable to the quality assurance of Essential Climate Variable data records. Report for the QA4ECV project (EC FP7). Royal Belgian Institute for Space Aeronomy, Belgium, https://doi.org/10.18758/71021041

[6] De Smedt, I., Theys, N., Yu, H., Danckaert, T., Lerot, C., Compernolle, S., Van Roozendael, M., Richter, A., Hilboll, A., Peters, E., Pedergnana, M., Loyola, D., Beirle, S., Wagner, T., Eskes, H., van Geffen, J., Boersma, K. F., and Veefkind, P.: Algorithm theoretical baseline for formaldehyde retrievals from S5P TROPOMI and from the QA4ECV project, (2018) Atmos. Meas. Tech., **11**, pp 2395-2426, https://doi.org/10.5194/amt-11-2395-2018.

[7] Dils B.; Gielen C.; et al., (2016) Quality indicators on uncertainties and representativity of atmospheric reference data, QA4ECV EC FP7 project deliverable D3.9. Available at http://www.qa4ecv.eu/sites/default/files/D3.9.pdf.

[8] GCOS (2010a) Guideline for the Generation of Datasets and Products Meeting GCOS Requirements, GCOS No. 143, WMO-TD No. 1530. Available at https://library.wmo.int/opac/doc_num.php?explnum_id=3854.

[9] GCOS (2010b) Implementation plan for the global observing system for climate in support of the UNFCCC (2010 Update), GCOS No. 138, WMO-TD/No. 1523. Available at http://www.wmo.int/pages/prog/gcos/Publications/gcos-138.pdf.

[10] GCOS (2011) Systematic Observation Requirements for Satellite-based Products for Climate, Supplemental details to the satellite-based component of the "Implementation Plan for the Global Observing System for Climate in Support of the UNFCCC – (2010 Update)", GCOS No. 154, December 2011. Available at https://www.wmo.int/pages/prog/gcos/Publications/gcos-154.pdf.

[11] GCOS (2016) The Global Observing System for Climate: Implementation Needs, GCOS No. 200. Available at https://library.wmo.int/opac/doc_num.php?explnum_id=3417.

[12] GUM (2008). Joint Committee for Guides in Metrology (JCGM/WG 1) 100:2008, Evaluation of measurement data – Guide to the expression of uncertainty in a measurement (GUM), ISO/IEC Guide 98-3:2008, http://www.bipm.org/utils/common/documents/jcgm/JCGM_100_2008_E.pdf

[13] Hendrick, F.; Dils, B.; et al., (2016) Historical record of independent reference data for NO2, HCHO, and CO, QA4ECV EC FP7 project deliverable D3.8. Available at http://www.qa4ecv.eu/sites/default/files/QA4ECV_D3.8_v1.0_web.pdf.

[14] Hurtmans, D., Coheur, P.-F., Wespes, C., Clarisse, L., Scharf, O., Clerbaux, C., Hadji-Lazaro, J., George, M., and Turquety, S., (2012). FORLI radiative transfer and retrieval code for IASI. J. Quant. Spectrosc. Radiat. Transfer 113:1391–1408. https://doi.org/10.1016/j.jqsrt.2012.02.036.

[15] ISO 9000:2015(en), Quality management systems - Fundamentals and vocabulary

[16] Keppens, A.; Lambert, J.-C.; Granville, J.; Miles, G.; Siddans, R.; van Peet, J. C. A.; van der R. J., A; Hubert, D.; Verhoelst, T.; Delcloo, A.; Godin-Beekmann, S.; Kivi, R.; Stübi, R. & Zehner, C., (2015) Round-robin evaluation of nadir ozone profile retrievals: methodology and application to MetOp-A GOME-2. Atmos. Meas. Tech., **8**, pp 2093-2120. https://doi.org/10.5194/amt-8-2093-2015.

[17] Larssen, S.; Sluyter, R.; and Helmis, C., (1999) Criteria for EUROAIRNET – The EEA Air Quality Monitoring and Information Network. http://www.eea.europa.eu/publications/TEC12

[18] Lorente, A.; Boersma, K. F.; Yu, H.; Dörner, S.; Hilboll, A.; Richter, A.; Liu, M.; Lamsal, L. N.; Barkley, M.; Smedt, I. D.; Roozendael, M. V.; Wang, Y.; Wagner, T.; Beirle, S.; Lin, J.-T.; Krotkov, N.; Stammes, P.; Wang, P.; Eskes, H. J. & Krol, M., (2017) Structural uncertainty in air mass factor calculation for NO2 and HCHO satellite retrievals. Atmos. Meas. Tech., **3**, 10, pp 759-782. https://doi.org/10.5194/amt-10-759-2017.

[19] Nightingale, J.; Boersma, K. F.; Muller, J.-P.; Compernolle, S.; Lambert, J.-C.; Blessing, S.; Giering, R.; Gobron, N.; De Smedt, I.; Coheur, P.; George, M.; Schulz, J. & Wood, A., (2018) Quality Assurance Framework Development Based on Six New ECV Data Products to Enhance User Confidence for Climate Applications Remote Sensing, **10**, 1254, https://doi.org/10.3390/rs10081254.

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

[21] Thorne, P. W.; Parker, D. E.; Christy, J. R. & Mears, C. A. (2005) Uncertainties in climate trends: lessons from upper-air temperature records. Bull. Am. Meteorol. Soc., **10**, 86, pp 1437-1442.

[22] Verhoelst, T.; Granville, J.; Hendrick, F.; Köhler, U.; Lerot, C.; Pommereau, J.-P.; Redondas, A.; Van Roozendael, M. & Lambert, J.-C., (2015) Metrology of ground-based satellite validation: co-location mismatch and smoothing issues of total ozone comparisons. Atmos. Meas. Tech., **12**, 8, pp 5039-5062, https://doi.org/10.5194/amt-8-5039-2015.

[23] Verhoelst, T. & Lambert, J. C., (2016) Generic metrology aspects of an atmospheric composition measurement and of data comparisons. EC Horizon2020 GAIA-CLIM technical Report / Deliverable D3.2. Available at http://www.gaia-clim.eu/workpackagedocument/d32-generic-metrology-aspects-atmospheric-composition-measurement-and-data.

[24] VIM/ISO:99 (2012). Joint Committee for Guides in Metrology (JCGM/WG 2) 200:2012 & ISO/IEC Guide 99-12:2007, International Vocabulary of Metrology – Basic and General Concepts and Associated Terms (VIM), http://www.bipm.org/en/publications/guides/vim.html

[25] WMO (2012) Guide to Meteorological Instruments and Methods of Observation, 2008 edition, Updated in 2010. WMO-No. 8. https://library.wmo.int/pmb_ged/wmo_8_en-2012.pdf