Precursors_cci+

ESA Climate Change Initiative (CCI)

D4.1 Product Validation and Intercomparison Report

Title	D4.1 Product Validation and Intercomparison Report
Reference	Precursors_cci+_D4.1_PVIR_01_01
Issue	01
Revision	01
Status	Final
Date of issue	26/04/2024
Document type	Deliverable

This work has received funding from the European Space Agency under Contract No 4000138243 and is the result of the cooperation between the partners of the Precursors_cci+ consortium.

	FUNCTION	NAME	DATE
LEAD AUTHORS	Validation Team Leaders	Steven Compernolle Jean-Christopher Lambert	13/03/2024
CONTRIBUTING AUTHORS	Validation Team (VALT)	Tijl Verhoelst (BIRA-IASB) Bavo Langerock (BIRA-IASB) Gaia Pinardi (BIRA-IASB) Mahesh Kumar Sha (BIRA-IASB) Corinne Vigouroux (BIRA-IASB)	08/03/2024
REVIEWED BY	EO Science Team (EOST)	Lieven Clarisse (ULB) Cathy Clerbaux (LATMOS) Isabelle De Smedt (BIRA-IASB) Maya George (LATMOS) Andreas Richter (IUPB) Sora Seo (DLR) Nicolas Theys (BIRA-IASB) Martin Van Damme (ULB)	20/03/2024
	ESA	Claire Macintosh	25/04/2024
REVIEWED AND APPROVED BY	Project Science Leaders	Michel Van Roozendael Folkert Boersma	20/03/2024
	ESA Project Officer	Simon Pinnock	25/04/2024

DISTRIBUTION	ESA Precursors_cci+ Consortium

rev0	2024/03/08	Merged version circulated to validators for checking
rev0.1	2024/03/13	Incorporated feedback validators. Included general sections like executive summary, introduction, …
rev0.2	2024/03/23	Incorporated feedback product providers and project lead.
Rev1	2024/04/26	Addressed RIDs ESA

Table of Contents

List of Figures

List of Tables

Executive Summary

This document is the *Product Validation and Intercomparison Report* (PVIR) Version 1 Revision 0 of the ESA CCI+ Precursors project (hereafter Precursors_cci+). The Validation Team (VALT) has developed a Product Validation Plan (PVP) translating user requirements into validation requirements, in order to ensure independent and traceable validation of the Precursors_cci+ data products and verification of compliance with the user requirements. This Product Validation and Intercomparison Report (PVIR) version 1 reports on the quality of the Climate Research Data Package (CRDP). For each of the Essential Climate Variable (ECV) data records provided by the project, the PVIR provides users with detailed validation results, with a list of quality indicators enabling the verification of fitness-for-purpose of the data for their own application, and with an assessment of the compliance of the CRDP with user requirements outlined in *The 2022 GCOS Implementation Plan* and the Precursors_cci+ *User Requirements Document* (URD).

The ozone and aerosols ECV precursors data products to be produced by the project are: (i) tropospheric nitrogen dioxide (NO_2) column, (ii) tropospheric formaldehyde (HCHO) column, (iii) tropospheric sulphur dioxide (SO_2) column (all 3 based on observations by the satellite instruments GOME, SCIAMACHY, OMI, TROPOMI and (not for SO_2) GOME-2A, -2B and -2C), (iv) carbon monoxide (CO) column (observations by MOPITT and IASI-A, -B and -C), (v) ammonia (NH_3) column (observations by IASI-A, -B and -C), and (vi) glyoxal (CHOCHO) column (observations by GOME-2A, -2B and -2C, OMI and TROPOMI). The glyoxal column product is not yet in scope for the PVIR v1. The final products planned to be released are processed to Level 3. Not all data products are yet in final state. In that case the validation reported here is on the underlying Level 2 data and/or on intermediate prototype products.

It is the aim of the future version 2 of the PVIR to consider the final Level 3 products, including also the glyoxal column product.

1. Introduction

1.1. Purpose and scope

The *Product Validation and Intercomparison Report* (PVIR) describes the validation of the Climate Data Records (CDR) of ozone and aerosol Essential Climate Variable (ECV) precursors data products to be generated by the ESA Precursors_cci+ project, namely:

(i) tropospheric nitrogen dioxide (NO_2) vertical column density (VCD),
(ii) tropospheric formaldehyde (HCHO) VCD,
(iii) tropospheric sulphur dioxide (SO_2) VCD,

all 3 based on observations by the satellite instruments ERS-2 GOME, Envisat SCIAMACHY, EOS-Aura OMI and Sentinel-5p TROPOMI, and (not for SO_2) Metop GOME-2A, -2B and -2C, and

(iv) carbon monoxide (CO) VCD based on observations by EOS-Terra MOPITT and Metop IASI-A, -B and -C,
(v) ammonia (NH_3) VCD based on observations by Metop IASI-A, -B and -C,
(vi) the glyoxal (CHOCHO) VCD based on observations by Metop GOME-2A, -2B and -2C, EOS-Aura OMI and Sentinel-5p TROPOMI.

In this version 1 and as foreseen, the glyoxal product is not yet in scope. Furthermore, the CDRs are not in their final stage for all ECVs.

1.2. Document overview

This Product Validation and Intercomparison Report Version 1 is structured as follows:

- Section 1 contains this introduction describing the purpose and scope of the document.
- Section 2 lists applicable and reference documents.
- Section 3 introduces the CRDP datasets addressed in this report.
- Section 4 introduces general aspects of the validation methodology, with description of the independent reference measurements in section 4.6.
- Sections 5 to 9 describe validation results and compliance assessment for the NO_2, HCHO, SO_2, CO and NH_3 data products. In each of these sections, first the specific validation methodology is explained, followed by an overview of results and ending with an evaluation of compliance with user requirements.
- Section 10 maps the content of this document with the Document Requirement Definitions outlined in Annex A of the Precursors_cci+ Statement of Work.
- Section 11 defines the recommended terminology, abbreviations and acronyms.

2. Applicable and reference documents

2.1. Applicable documents

[RD1:DRD] CCI+ Phase 2 Theme (I) - Precursors ECV SoW - Annex A – Document Requirement Definitions, Issue 1.2, ESA-EOP-SC-SP-2021-38, 06/08/2021.

2.2. Reference documents

2.2.1. Climate data user requirement documents

[RD2:GCOS_IP] The 2022 GCOS Implementation Plan. GCOS-244. GOOS-272. 98 pp. © World Meteorological Organization, 2022. https://gcos.wmo.int/en/publications/gcos-implementation-plan2022

[RD3:GCOS_Req] The 2022 GCOS ECVs Requirements. GCOS-245. (Supplement to the 2022 GCOS Implementation Plan). 244 pp. © World Meteorological Organization, 2022. https://gcos.wmo.int/en/publications/gcos-implementation-plan2022

[RD4:URD] Precursors CCI+ User Requirement Document, deliverable 1.1, issue 1, Precursors_cci+_D1.1_URD_01_01, 25/01/2023.

2.2.2. Satellite data products

2.2.2.1. Multi target products

[RD5:PUG] Precursors CCI+ Product User Guide, deliverable 4.2, Precursors_cci+_D4.2_PUG_01_00, draft version.

2.2.2.2. NO_2 data products

[RD6:vanGeffen2002] van Geffen, J.; Eskes, H.; Compernolle, S.; Pinardi, G.; Verhoelst, T.; Lambert, J.-C.; Sneep, M.; ter Linden, M.; Ludewig, A.; Boersma, K. F. & Veefkind, J. P. Sentinel-5P TROPOMI NO_2 retrieval: impact of version v2.2 improvements and comparisons with OMI and ground-based data Atmos. Meas. Tech., 2022, 15, 2037-2060

[RD7:PRF_S5PPALNO2] S5P Nitrogen Dioxide v02.03.01 intermediate reprocessing on the S5P-PAL system: Readme file, issue 1.0, by Henk Eskes, Jos van Geffen, Maarten Sneep, Pepijn Veefkind, Sander Niemeijer, Claus Zehner, 15 December 2021

[RD8:PSD_QA4ECVNO2] *Product Specification Document for the QA4ECV NO2 ECV precursor product* by F. Boersma et al., in 'Product User Guide for Land ECVs and Product Specification Document for Atmosphere ECV precursors' (D4.6), QA4ECV EC FP7 project deliverable 4.6, March 2017.

[RD9:Boersma2018] 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 A, R. J.; Nightingale, J.; De Rudder, A.; Irie, H.; Pinardi, G.; Lambert, J.-C. & Compernolle, S. C. Improving algorithms and uncertainty estimates for satellite NO_2 retrievals: results from the quality assurance for the essential climate

variables (QA4ECV) project Atmos. Meas. Tech., 2018, 11, 6651-6678, https://doi.org/10.5194/amt-11-6651-2018

2.2.2.3. HCHO data products

[RD10:DeSmedt2018] 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. & Veefkind, P. Algorithm theoretical baseline for formaldehyde retrievals from S5P TROPOMI and from the QA4ECV project Atmos. Meas. Tech., 2018, 11, 2395-2426, https://doi.org/10.5194/amt-11-2395-2018

[RD11:PRF_S5PHCHO] S5P MPC Product Readme Formaldehyde V02.05.00, issue 2.6, by I. De Smedt, F. Romahn, K.-U. Eichmann, 20 July 2023.

[RD12:PSD_QA4ECVHCHO] *Product Specification Document for the QA4ECV HCHO ECV precursor product* by I. De Smedt et al., in 'Product User Guide for Land ECVs and Product Specification Document for Atmosphere ECV precursors' (D4.6), QA4ECV EC FP7 project deliverable 4.6, March 2017.

[RD13:PUM_ACSAFHCHO] Product User Manual for GOME Total Column Products of Ozone, NO_2, BrO, HCHO, SO_2, H2O, OClO and Cloud Properties, GDP 4.8, SAF/AC/DLR/PUM/01, Iss. 3/C, Rev. 1, by Valks, P., et al., (2023). Oct, 2023

2.2.2.4. SO_2 data products

[RD14:Theys2021] Theys, N.; Fioletov, V.; Li, C.; De Smedt, I.; Lerot, C.; McLinden, C.; Krotkov, N.; Griffin, D.; Clarisse, L.; Hedelt, P.; Loyola, D.; Wagner, T.; Kumar, V.; Innes, A.; Ribas, R.; Hendrick, F.; Vlietinck, J.; Brenot, H. & Van Roozendael, M. A sulfur dioxide Covariance-Based Retrieval Algorithm (COBRA): application to TROPOMI reveals new emission sources Atmos. Chem. Phys., 2021, 21, 16727-16744, https://doi.org/10.5194/acp-21-16727-2021

2.2.2.5. CO data products

[RD15:ATBD_ACSAFCO] Algorithm Theoretical Basis Document of "Fast Optimal Retrieval on Layers for IASI (FORLI), IASI Reprocessed CO L2 CDR (O3M-517) Metop-A&B", by Daniel Hurtmans (ULB), Pierre Coheur (ULB), Rosa Astoreca (ULB), Juliette Hadji-Lazaro (LATMOS), Camille Viatte (LATMOS), Cathy Clerbaux (LATMOS), SAF/AC/ULB/ATBD/003, Issue 1.2, 9 March 2023, https://acsaf.org/docs/atbd/Algorithm_Theoretical_Basis_Document_IASI_CO_Mar_2023.pdf.

[RD16:Deeter2022] Deeter, M.; Francis, G.; Gille, J.; Mao, D.; Martnez-Alonso, S.; Worden, H.; Ziskin, D.; Drummond, J.; Commane, R.; Diskin, G. & McKain, K. The MOPITT Version 9 CO product: sampling enhancements and validation Atmos. Meas. Tech., 2022, 15, 2325-2344, https://doi.org/10.5194/amt-15-2325-2022

2.2.2.6. NH_3 data product

[RD17:Clarisse2023] Clarisse, L.; Franco, B.; Van Damme, M.; Di Gioacchino, T.; Hadji-Lazaro, J.; Whitburn, S.; Noppen, L.; Hurtmans, D.; Clerbaux, C. & Coheur, P. The IASI NH_3 version

4 product: averaging kernels and improved consistency Atmos. Meas. Tech., 2023, 16, 5009-5028, https://doi.org/10.5194/amt-16-5009-2023

2.2.3. Ground-based (reference) measurements

2.2.3.1. Zenith-scattered-light UV-Visible DOAS (ZSL-DOAS)

[RD18:Solomon1987] Solomon, S.; Schmeltekopf, A. L. & Sanders, R. W. On the interpretation of zenith sky absorption measurements, *J. Geophys. Res.,* 1987*, 92*, 8311-8319

[RD19:Pommereau1988] Pommereau, J. & Goutail, F. O_3 and NO_2 ground-based measurements by visible spectrometry during Arctic winter and spring 1988 Geophys. Res. Lett., 1988, 15, 891-894

[RD20:Bognar2019] Bognar, K., Zhao, X., Strong, K., Boone, C., Bourassa, A., Degenstein, D., Drummond, J., Duff, A., Goutail, F., Griffin, D., Jeffery, P., Lutsch, E., Manney, G., McElroy, C., McLinden, C., Millán, L., Pazmino, A., Sioris, C., Walker, K., and Zou, J.: Updated validation of ACE and OSIRIS ozone and NO2 measurements in the Arctic using ground-based instruments at Eureka, Canada, J. Quant. Spectrosc. Ra., 238, 106571, https://doi.org/10.1016/j.jqsrt.2019.07.014, 2019.

[RD21:Roscoe1999] Roscoe, H., Johnston, P., Van Roozendael, M., Richter, A., Sarkissian, A., Roscoe, J., Preston, K., Lambert, J.-C., Hermans, C., De Cuyper, W., Dzienus, S., Winterrath, T., Burrows, J., Goutail, F., Pommereau, J.-P., D'Almeida, E., Hottier, J., Coureul, C., Ramon, D., Pundt, I., Bartlett, L., McElroy, C., Kerr, J., Elokhov, A., Giovanelli, G., Ravegnani, F., Premuda, M., Kostadinov, I., Erle, F., Wagner, T., Pfeilsticker, K., Kenntner, M., Marquard, L., Gil, M., Puentedura, O., Yela, M., Arlander, W., Kåstad Høiskar, B., Tellefsen, C., Karlsen Tørnkvist, K., Heese, B., Jones, R., Aliwell, S., and Freshwater, R.: Slant column measurements of O3 and NO2 during the NDSC intercomparison of zenith-sky UV-visible spectrometers in June 1996, J. Atmos. Chem., 32, 281–314, 1999.

[RD22:Hendrick2004] Hendrick, F., Barret, B., Van Roozendael, M., Boesch, H., Butz, A., De Mazière, M., Goutail, F., Hermans, C., Lambert, J.-C., Pfeilsticker, K., and Pommereau, J.-P.: Retrieval of nitrogen dioxide stratospheric profiles from ground-based zenith-sky UV-visible observations: validation of the technique through correlative comparisons, Atmos. Chem. Phys., 4, 2091–2106, https://doi.org/10.5194/acp-4-2091-2004, 2004.

[RD23:Hendrick2011] Hendrick, F.; Pommereau, J.-P.; Goutail, F.; Evans, R. D.; Ionov, D.; Pazmino, A.; Kyrö, E.; Held, G.; Eriksen, P.; Dorokhov, V.; Gil, M. & Van Roozendael, M. NDACC/SAOZ UV-visible total ozone measurements: improved retrieval and comparison with correlative ground-based and satellite observations Atmos. Chem. Phys., 2011, 11, 5975-5995

[RD24:Vandaele2005] Vandaele, A. C., et al. (2005), An intercomparison campaign of ground-based UV-visible measurements of NO_2, BrO, and OClO slant columns: Methods of analysis and results for NO2, J. Geophys. Res., 110, D08305, doi:10.1029/2004JD005423.

[RD25:Yela2017] Yela, M., Gil-Ojeda, M., Navarro-Comas, M., Gonzalez-Bartolomé, D., Puentedura, O., Funke, B., Iglesias, J., Rodríguez, S., García, O., Ochoa, H., and Deferrari, G.: Hemispheric asymmetry in stratospheric NO2 trends, Atmos. Chem. Phys., 17, 13373–13389, https://doi.org/10.5194/acp-17-13373-2017, 2017.

2.2.3.2. Multi-axis UV-Visible DOAS (MAX-DOAS)

[RD26:Wang2017] Wang, Y., Lampel, J., Xie, P., Beirle, S., Li, A., Wu, D., and Wagner, T.: Ground-based MAX-DOAS observations of tropospheric aerosols, NO_2, SO_2 and HCHO in Wuxi, China, from 2011 to 2014, Atmos. Chem. Phys., 17, 2189–2215, https://doi.org/10.5194/acp-17-2189-2017, 2017.

[RD27:Pinardi2013] Pinardi, G., Van Roozendael, M., Abuhassan, N., Adams, C., Cede, A., Clémer, K., Fayt, C., Frieß, U., Gil, M., Herman, J., Hermans, C., Hendrick, F., Irie, H., Merlaud, A., Navarro Comas, M., Peters, E., Piters, A. J. M., Puentedura, O., Richter, A., Schönhardt, A., Shaiganfar, R., Spinei, E., Strong, K., Takashima, H., Vrekoussis, M., Wagner, T., Wittrock, F., and Yilmaz, S.: MAX-DOAS formaldehyde slant column measurements during CINDI: intercomparison and analysis improvement, Atmos. Meas. Tech., 6, 167–185, https://doi.org/10.5194/amt-6-167-2013, 2013.

[RD28:Wang2014a] Wang, T., Hendrick, F., Wang, P., Tang, G., Clémer, K., Yu, H., Fayt, C., Hermans, C., Gielen, C., Müller, J.-F., Pinardi, G., Theys, N., Brenot, H., and Van Roozendael, M.: Evaluation of tropospheric SO_2 retrieved from MAX-DOAS measurements in Xianghe, China, Atmos. Chem. Phys., 14, 11149–11164, https://doi.org/10.5194/acp-14-11149-2014, 2014.

2.2.3.3. Fourier Transform Infra-Red spectrometry (FTIR)

[RD29:Vigouroux2018] Vigouroux, C., Bauer Aquino, C. A., Bauwens, M., Becker, C., Blumenstock, T., De Mazière, M., García, O., Grutter, M., Guarin, C., Hannigan, J., Hase, F., Jones, N., Kivi, R., Koshelev, D., Langerock, B., Lutsch, E., Makarova, M., Metzger, J.-M., Müller, J.-F., Notholt, J., Ortega, I., Palm, M., Paton-Walsh, C., Poberovskii, A., Rettinger, M., Robinson, J., Smale, D., Stavrakou, T., Stremme, W., Strong, K., Sussmann, R., Té, Y., and Toon, G.: NDACC harmonized formaldehyde time series from 21 FTIR stations covering a wide range of column abundances, Atmos. Meas. Tech., 11, 5049–5073, https://doi.org/10.5194/amt-11-5049-2018, 2018.

[RD30:Dammers2015] Dammers, E., Vigouroux, C., Palm, M., Mahieu, E., Warneke, T., Smale, D., Langerock, B., Franco, B., Van Damme, M., Schaap, M., Notholt, J., and Erisman, J. W.: Retrieval of ammonia from ground-based FTIR solar spectra, Atmos. Chem. Phys., 15, 12789–12803, https://doi.org/10.5194/acp-15-12789-2015, 2015.

2.2.4. Geophysical validation

2.2.4.1. Validation protocol

[RD31:PVPv1] Precursors CCI+ Product Validation Plan (PVP), deliverable 1.3, issue 1, Precursors_cci+_D1.3_PVP_01_01, 31/03/2023.

2.2.4.2. Multi target validation reports

[RD32:PVASR] Precursors CCI+ Product Validation and Algorithm Selection Report (PVASR), deliverable 2.3, issue 1, Precursors_cci+_D2.3_PVASR_01_00, draft, 29 Feb 2024.

[RD33:QA4ECV-PVIR] Compernolle, S.; Lambert, J.-C.; Pinardi, G.; Niemeijer, S.; Rino, B.; Granville, J.; Hubert, D.; Keppens, A. & Verhoelst, T., Product Validation and Intercomparison Report for: QA4ECV NO2 OMI 1.0, QA4ECV HCHO OMI 1.0, IASI FORLI CO v20100815+ v20140922 Belgian Institute for Space Aeronomy, EC FP7 QA4ECV deliverable 5.6, 2017.

[RD34:ROCVR] Quarterly Validation Report of the Copernicus Sentinel-5 Precursor Operational Data Products #21: April 2018 – November 2023. Lambert, J.-C., A. Keppens, S. Compernolle, K.-U. Eichmann, M. de Graaf, D. Hubert, B. Langerock, A. Ludewig, M.K. Sha, T. Verhoelst, T. Wagner, C. Ahn, A. Argyrouli, D. Balis, K.L. Chan, M. Coldewey-Egbers, I. De Smedt, H. Eskes, A.M. Fjæraa, K. Garane, J.F. Gleason, F. Goutail, J. Granville, P. Hedelt, K.-P. Heue, G. Jaross, Q. Kleipool, ML. Koukouli, R. Lutz, M.C Martinez Velarte, K. Michailidis, S. Nanda, S. Niemeijer, A. Pazmiño, G. Pinardi, A. Richter, N. Rozemeijer, M. Sneep, D. Stein Zweers, N. Theys, G. Tilstra, O. Torres, P. Valks, J. van Geffen, C. Vigouroux, P. Wang, and M. Weber. S5P MPC Routine Operations Consolidated Validation Report series, Issue #21, Version 21.01.00, 194 pp., 18 December 2023.

2.2.4.3. NO_2 column validation

[RD35:Verhoelst2021] Verhoelst, T., Compernolle, S., Pinardi, G., Lambert, J.-C., Eskes, H. J., Eichmann, K.-U., Fjæraa, A. M., Granville, J., Niemeijer, S., Cede, A., Tiefengraber, M., Hendrick, F., Pazmiño, A., Bais, A., Bazureau, A., Boersma, K. F., Bognar, K., Dehn, A., Donner, S., Elokhov, A., Gebetsberger, M., Goutail, F., Grutter de la Mora, M., Gruzdev, A., Gratsea, M., Hansen, G. H., Irie, H., Jepsen, N., Kanaya, Y., Karagkiozidis, D., Kivi, R., Kreher, K., Levelt, P. F., Liu, C., Müller, M., Navarro Comas, M., Piters, A. J. M., Pommereau, J.-P., Portafaix, T., Prados-Roman, C., Puentedura, O., Querel, R., Remmers, J., Richter, A., Rimmer, J., Rivera Cárdenas, C., Saavedra de Miguel, L., Sinyakov, V. P., Stremme, W., Strong, K., Van Roozendael, M., Veefkind, J. P., Wagner, T., Wittrock, F., Yela González, M., and Zehner, C.: Ground-based validation of the Copernicus Sentinel-5P TROPOMI NO_2 measurements with the NDACC ZSL-DOAS, MAX-DOAS and Pandonia global networks, Atmos. Meas. Tech., 14, 481–510, https://doi.org/10.5194/amt-14-481-2021, 2021.

[RD36:Compernolle2020] Compernolle, S., Verhoelst, T., Pinardi, G., Granville, J., Hubert, D., Keppens, A., Niemeijer, S., Rino, B., Bais, A., Beirle, S., Boersma, F., Burrows, J. P., De Smedt, I., Eskes, H., Goutail, F., Hendrick, F., Lorente, A., Pazmino, A., Piters, A., Peters, E., Pommereau, J.-P., Remmers, J., Richter, A., van Geffen, J., Van Roozendael, M., Wagner, T., and Lambert, J.-C.: Validation of Aura-OMI QA4ECV NO_2 climate data records with ground-based DOAS networks: the role of measurement and comparison uncertainties, Atmos. Chem. Phys., 20, 8017–8045, https://doi.org/10.5194/acp-20-8017-2020, 2020.

2.2.4.4. HCHO column validation

[RD37:ACSAF-VAL]. Gaia Pinardi, Isabelle De Smedt, Klaus-Peter Heue, Pieter Valks, AC SAF Validation Report of Reprocessed L2 Total HCHO, GOME-2/MetopA&MetopB, doc ref: SAF/AC/IASB/VR/HCHO, October 2023. Draft.

[RD38:Vigouroux2020] Vigouroux, C., Langerock, B., Bauer Aquino, C. A., Blumenstock, T., Cheng, Z., De Mazière, M., De Smedt, I., Grutter, M., Hannigan, J. W., Jones, N., Kivi, R., Loyola, D., Lutsch, E., Mahieu, E., Makarova, M., Metzger, J.-M., Morino, I., Murata, I., Nagahama, T., Notholt, J., Ortega, I., Palm, M., Pinardi, G., Röhling, A., Smale, D., Stremme, W., Strong, K., Sussmann, R., Té, Y., van Roozendael, M., Wang, P., and Winkler, H.: TROPOMI–Sentinel-5 Precursor formaldehyde validation using an extensive network of ground-based Fourier-transform infrared stations, Atmos. Meas. Tech., 13, 3751–3767, https://doi.org/10.5194/amt-13-3751-2020, 2020.

[RD39:DeSmedt2021] De Smedt, I., Pinardi, G., Vigouroux, C., Compernolle, S., Bais, A., Benavent, N., Boersma, F., Chan, K.-L., Donner, S., Eichmann, K.-U., Hedelt, P., Hendrick, F., Irie, H., Kumar, V., Lambert, J.-C., Langerock, B., Lerot, C., Liu, C., Loyola, D., Piters, A., Richter, A., Rivera Cárdenas, C., Romahn, F., Ryan, R. G., Sinha, V., Theys, N., Vlietinck, J., Wagner, T., Wang, T., Yu, H., and Van Roozendael, M.: Comparative assessment of TROPOMI and OMI formaldehyde observations and validation against MAX-DOAS network column measurements, Atmos. Chem. Phys., 21, 12561–12593, https://doi.org/10.5194/acp-21-12561-2021, 2021.

[RD40:Mueller2024] Müller, J.-F.; Stavrakou, T.; Oomen, G.-M.; Opacka, B.; De Smedt, I.; Guenther, A.; Vigouroux, C.; Langerock, B.; Aquino, C. A. B.; Grutter, M.; Hannigan, J.; Hase, F.; Kivi, R.; Lutsch, E.; Mahieu, E.; Makarova, M.; Metzger, J.-M.; Morino, I.; Murata, I.; Nagahama, T.; Notholt, J.; Ortega, I.; Palm, M.; Röhling, A.; Stremme, W.; Strong, K.; Sussmann, R.; Té, Y. & Fried, A. Bias correction of OMI HCHO columns based on FTIR and aircraft measurements and impact on top-down emission estimates Atmos. Chem. Phys., 2024, 24, 2207-2237, https://doi.org/10.5194/acp-24-2207-2024

[RD41:Oomen2024] Oomen, G.-M.; Müller, J.-F.; Stavrakou, T.; De Smedt, I.; Blumenstock, T.; Kivi, R.; Makarova, M.; Palm, M.; Röhling, A.; Té, Y.; Vigouroux, C.; Friedrich, M. M.; Frieß, U.; Hendrick, F.; Merlaud, A.; Piters, A.; Richter, A.; Van Roozendael, M. & Wagner, T. Weekly derived top-down volatile-organic-compound fluxes over Europe from TROPOMI HCHO data from 2018 to 2021 Atmos. Chem. Phys., 2024, 24, 449-474, https://doi.org/10.5194/acp-24-449-2024

2.2.4.5. SO_2 column validation

[RD42:Theys2015] Theys, N.; De Smedt, I.; van Gent, J.; Danckaert, T.; Wang, T.; Hendrick, F.; Stavrakou, T.; Bauduin, S.; Clarisse, L.; Li, C.; Krotkov, N.; Yu, H.; Bernot, H. & Van Roozendael, M. Sulfur dioxide vertical column DOAS retrievals from the Ozone Monitoring Instrument: Global observations and comparison to ground-based and satellite data J. Geophys. Res. Atmos., Wiley-Blackwell, 2015, 120, 1-22

2.2.4.6. CO column validation

[RD43:VAL_FORLICO] AC SAF FORLI CO validation report, by B. Langerock et al. Submitted.

2.2.4.7. Methods, models, tools and miscellaneous

[RD44:Keppens2019] Keppens, A., Compernolle, S., Verhoelst, T., Hubert, D., and Lambert, J.-C.: Harmonization and comparison of vertically resolved atmospheric state observations: methods, effects, and uncertainty budget, Atmos. Meas. Tech., 12, 4379–4391, https://doi.org/10.5194/amt-12-4379-2019, 2019.

[RD45:Verhoelst2015] 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, https://doi.org/10.5194/amt-8-5039-2015, 2015.

[RD46:Langerock2015] Langerock, B.; De Mazière, M.; Hendrick, F.; Vigouroux, C.; Desmet, F.; Dils, B. & Niemeijer, S. Description of algorithms for co-locating and comparing gridded model data with remote-sensing observations Geosci. Model Dev., Copernicus GmbH, 2015, 8, 911-921

[RD47:Rodgers2003] Rodgers, C. D. & Connor, B. J. Intercomparison of remote sounding instruments, *J. Geophys. Res.*, **2003**, *108*, 4116

[RD48:Errera2001] Errera, Q., and Fonteyn, D. (2001), Four-dimensional variational chemical assimilation of CRISTA stratospheric measurements, *J. Geophys. Res.*, 106(D11), 12253– 12265, doi:10.1029/2001JD900010.

[RD49:Eskes2003] Eskes, H. J. & Boersma, K. F. Averaging kernels for DOAS total-column satellite retrievals Atmos. Chem. Phys., 2003, 3, 1285-1291, doi: 10.5194/acp-3-1285-2003

[RD50:HARP] HARP is a toolset for ingesting, processing and inter-comparing satellite or model data against correlative data. https://github.com/stcorp/harp; manual at https://stcorp.github.io/harp/doc/html/index.html Copyright (C) 2015-2022 S[&]T, The Netherlands

3. Climate Research Data Package (CRDP)

3.1. NO_2 column data sets

Table 3-1 presents an overview of data products input to or produced within the CCI+precursors project. The QA4ECV L2 products [RD9:Boersma2018][RD8:PSD_QA4ECVNO2] and the S5P PAL L2 product [RD7:PRF_S5PPALNO2][RD6:vanGeffen2002] have been publicly released. The GOME-2A AC SAF L2 product validated here is an intermediate prototype product.

Table 3-1. NO_2 column data sets input to, or produced within, the CCI+precursors project. In analogy to [RD5:PUG], product ID is given as {species}_{processing level}_{platform}_{sensor}.

Product ID	Processor	Provider	Time range	VAL Section[4]
L2 data sets				
NO2_L2_ERS2_GOME1	QA4ECV v1.1[1]	QA4ECV/KNMI	1995-2003	5[5]
NO2_L2_ENVISAT_SCIAMACHY	QA4ECV v1.1[1]	QA4ECV/KNMI	2002-2012	5[5]
NO2_L2_GOME2_A	AC SAF proto[2]	AC SAF/DLR	2007-2018	5
NO2_L2_GOME2_B	AC SAF proto[2]	AC SAF/DLR	2013-now	N/A
NO2_L2_GOME2_C	AC SAF proto[2]	AC SAF/DLR	2018-now	N/A
NO2_L2_AURA_OMI	QA4ECV v1.1/ L1v4[3]	QA4ECV/KNMI	Planned: 2004-now Not yet available	N/A
NO2_L2_S5P_TROPOMI	PAL RPRO v2.3.1[1]	EU/ESA/KNMI	2018-2021	5
L3 data sets: not yet available				

1: publicly available product, input to this project. 2: intermediate non-public product. 3: Reprocessed OMI/EOS Aura data (based on L1 v4) were not yet available at the time of the validation analysis. 4: Validation analysis targets only stratospheric VCD in this document. 5: Both default QA4ECV data assimilation and alternative QA4ECV STREAM processors are analysed.

3.2. HCHO column data sets

Table 3-2 presents an overview of L2 and L3 data products input to or produced within the CCI+precursors project. The L2 S5P and QA4ECV products [RD10:DeSmedt2018][RD12:PSD_QA4ECVHCHO][RD11:PRF_S5PHCHO], and AC SAF GOME2 product [RD13:PUM_ACSAFHCHO] are publicly available products and are input to the L3 products. The L3 products validated here are prototype versions: the L3 processor version is fv0100 for GOME1, SCIAMACHY and GOME2, and fv0200 (with e.g., exclusion of edge rows and inclusion of averaging kernels) for OMI and TROPOMI.

Table 3-2. HCHO column data sets input to, or produced within, the CCI+precursors project. In analogy to [RD5:PUG], product ID is given as {species}_{processing level}_{platform}_{sensor}.

Product ID	L2 Processor {L3 processor}	Provider	Time range	VAL: section or ref
L2 data sets				
HCHO_L2_ERS2_GOME1	QA4ECV v1.2[1]	QA4ECV/ BIRA-IASB	1995-2003	N/A
HCHO_L2_ENVISAT_SCIAMACHY	QA4ECV v1.2[1]	QA4ECV/ BIRA-IASB	2003-2012[4]	N/A
HCHO_L2_GOME2_A	GDP 4.8/4.9 RPRO[1]	AC SAF/DLR	2007-2018	6.3.1, [RD37:ACSAF-VAL]
HCHO_L2_GOME2_B	GDP 4.8/4.9 RPRO[1]	AC SAF/DLR	2013-now	6.3.1, [RD37:ACSAF-VAL]
HCHO_L2_GOME2_C	GDP 4.8/4.9 RPRO[1]	AC SAF/DLR	2018-now	6.3.1, [RD37:ACSAF-VAL]
HCHO_L2_AURA_OMI	QA4ECV v1.2[1]	QA4ECV/ BIRA-IASB	2004-2021	6.3.1, [RD40:Mueller2024]
HCHO_L2_S5P_TROPOMI	Operational[1], collection 3, ≥v2.4	EU/ESA/BIRA-IASB/DLR	2018-now	6.3.1, [RD34:ROCVR]
L3 data sets[2]				
HCHO_L3_ERS2_GOME1	QA4ECV v1.2 {fv0100}	BIRA-IASB/QA4ECV	1996-2002[3]	6.3.2
HCHO_L3_ENVISAT_SCIAMACHY	QA4ECV v1.2 {fv0100}	BIRA-IASB/QA4ECV	2003-2012	6.3.2
HCHO_L3_METOP_GOME2_A/B/C	GDP4.8/4.9 {fv0100}	BIRA-IASB/DLR/EUMETSAT	2007-2023	6.3.2
HCHO_L3_AURA_OMI	QA4ECV v1.2 {fv0200}	BIRA-IASB/QA4ECV	2005-2021	6.3.2
HCHO_L3_S5P_TROPOMI	OFFL v0.3_02 {fv0200}	BIRA-IASB/DLR/ESA	2018-2023	6.3.2

1: publicly available product, input to this project. 2: intermediate non-public product. 3: Although GOME did still observations in 2003, global coverage of GOME ends in early 2003. After only Europe is covered, therefore background correction can no longer be performed. 4: Although there are SCIAMACHY observations from 08/2002, QA4ECV processing only starts in 2003.

We furthermore note that in [RD32:PVASR] comparisons are made with GOME-2 AC SAF variants, using alternative priors (TM5 or CAMS reanalysis instead of the default IMAGES) and with/without cloud correction applied.

3.3. SO_2 column data sets

The SO_2 L2 data products have to be produced during the CCI+precursors project. The retrieval is based on the COBRA algorithm [RD14:Theys2021]. At the time of writing this document, only a single year of OMI data (intermediate prototype, with CAMS reanalysis as prior and OMI O2-O2 for cloud correction) was available for validation (Table 3-3) and also considered in [RD32:PVASR].

Table 3-3. SO_2 column data sets (to be) produced within, the CCI+precursors project. In analogy to [RD5:PUG], product ID is given as {species}_{processing level}_{platform}_{sensor}.

Product ID	Processor	Provider	Time range	VAL: section
L2 data sets				
SO2_L2_ERS2_GOME1	Tbd	BIRA-IASB		N/A
SO2_L2_ENVISAT_SCIAMACHY	Tbd	BIRA-IASB		N/A
SO2_L2_AURA_OMI	CCI+p prototype	BIRA-IASB	Planned: 2004-now Available: 2012	7.3
SO2_L2_S5P_TROPOMI	OFFL v0.3_02/ CCI+p v1.0	BIRA-IASB	Not yet available	N/A
L3 data sets: not yet available				

In [RD32:PVASR] comparisons are made with OMI COBRA variant products, using TM5 as prior, and also including not-cloud corrected (“clear-sky”) products.

3.4. CO column data sets

The L3 CO IASI multi-platform product validated here (see Table 3-4) is a non-public intermediate product, based on IASI L2 FORLI CO data reprocessed by EUMETSAT in 2021 (see [RD15:ATBD_ACSAFCO]). The L2 CO total columns from IASI A, B and C are filtered with an aposteriori cloud product and averaged in monthly grids (1°x1°). This resulted then in a new product, generated during cycle 1 of the project, called the L3 CO IASI multi-platform product. The final product, that will be based on a combination of IASI and MOPITT v9T [RD16:Deeter2022] data, is being generated during cycle 2 of the project and will be delivered in February 2025.

Validation results of IASI L2 data (reprocessed by EUMETSAT in 2021) took place within the AC SAF project and are presented in [RD43:VAL_FORLICO]. Only a very brief summary is given in Section 8.3.1.

Table 3-4. CO column data sets (to be) produced within the CCI+precursors project. In analogy to [RD5:PUG], product ID is given as {species}_{processing level}_{platform}_{sensor}.

Product ID	Processor	Provider	Time range	VAL: section
L3 data sets				
CO_L3_IASI_A/B/C[1]	L2: EUMET CO	LATMOS-ULB	2008-2022	8.3
CO_L3_MRG	L2: EUMET CO /MOPITT v9T	LATMOS-ULB	Not yet available	N/A

1. IASI multi-platform product. Intermediate product provided to BIRA, not for end users.

3.5. NH_3 column data sets

The L3 NH_3 IASI data validated here (see Table 3-5) is based on the L2 ANNI v4.0.0 product, described in [RD17:Clarisse2023].

Table 3-5. NH_3 column data sets produced within the CCI+precursors project. In analogy to [RD5:PUG], product ID is given as {species}_{processing level}_{platform}_{sensor}.

Product ID	Processor	Provider	Time range	VAL: section
L3 data sets				
NH3_L3_IASI_A/B/C	ANNI v4.0.0	ULB-LATMOS	2007-2023	9.3
NH3_L3_MRG	ANNI v4.0.0	ULB-LATMOS	Not yet available	

4. ECV Validation Methodology

4.1. General principles of the validation process

The Precursors_cci+ Product Validation Plan [RD31:PVPv1] describes the validation protocol applied in this assessment.

Every ECV data set produced by the project needs to be validated against the official user requirements formulated in the Precursors_cci+ User Requirement Document [RD4:URD]. In the following section, we summarize the user requirements applicable to the present validation study. The translation of these user requirements into validation requirements is described in the Precursors_cci+ Product Validation Plan [RD31:PVPv1]. The geophysical validation of ECV data products delivered in the CRDP relies primarily on comparisons with ground-based reference measurements. These comparisons are reported in Sections 5 to 9. The reference measurements used in this study are summarised in Section 4.6. In preparation of the comparisons, the data sets must undergo a suite of data manipulations, including data filtering based on, e.g., quality flags, harmonisation of coordinate systems and of units, reduction of differences in vertical and horizontal smoothing, selection of co-locations meeting appropriate criteria… These operations depend on the ECV data product and associated retrieval algorithms; therefore, they are described in the respective sections reporting the comparisons.

4.2. Compliance with user requirements

To report the level of compliance with numerical requirements from [RD3:GCOS_Req][RD4:URD], tables with the following color code will be used.

Color code:

x>Threshold	Threshold≥x>Breakthrough	Breakthrough≥x>Goal	Goal≥x

Hereafter we reproduce the user requirements as described in [RD3:GCOS_Req][RD4:URD], against which Precursors_cci+ ECV data products have to be verified and/or validated.

4.2.1. NO_2 data product requirements

NO_2 data product requirements are reproduced here from [RD3:GCOS_Req]. These also served as input to [RD4:URD]. It is usual practice to express uncertainty in 1-sigma, therefore the table is adapted accordingly.

Table 4-1. 2022 GCOS IP requirements for the ECV tropospheric NO_2 column, adapted from [RD3:GCOS_Req]. Specifically, 2-sigma uncertainty is adapted to 1-sigma.

QUANTITY	THRESHOLD	BREAKTHROUGH	GOAL
Horizontal resolution	100 km	30 km	10 km
Vertical resolution	N/A	N/A	N/A
Temporal resolution	30 days	1 day	1 hour
Timeliness	30 days	7 days	1 day
Measurement	Max(50%,	Max(20%,	Max(10%,

uncertainty (1-sigma)	2.5E15 molec cm^{-2})	1E15 molec cm^{-2})	0.5E15 molec cm^{-2})
Stability	Max(20%, 1E15 molec cm^{-2}/decade)	Max(8%, 1E15 molec cm^{-2}/decade)	Max(4%, 1E15 molec cm^{-2}/decade)

The absolute (relative) requirement on measurement uncertainty is applicable below (above) measurand values of 5E15 molec cm^{-2}.

4.2.2. **HCHO data product requirements**

HCHO data product requirements are reproduced here from [RD3:GCOS_Req]. These also served as input to [RD4:URD]. It is usual practice to express uncertainty in 1-sigma, therefore the table is adapted accordingly.

Table 4-2. 2022 GCOS IP requirements for the ECV tropospheric HCHO column, adapted from [RD3:GCOS_Req]. Specifically, 2-sigma uncertainty is adapted to 1-sigma.

QUANTITY	THRESHOLD	BREAKTHROUGH	GOAL
Horizontal resolution	100 km	30 km	10 km
Vertical resolution	N/A	N/A	N/A
Temporal resolution	30 days	1 day	1 hour
Timeliness	30 days	7 days	1 day
Measurement uncertainty (1-sigma)	Max(50%, 20E15 molec cm^{-2})	Max(20%, 8E15 molec cm^{-2})	Max(10%, 4E15 molec cm^{-2})
Stability	Max(20%, 8E15 molec cm^{-2}/decade)	Max(8%, 8E15 molec cm^{-2}/decade)	Max(4%, 8E15 molec cm^{-2}/decade)

The absolute (relative) requirement on measurement uncertainty is applicable below (above) measurand values of 40E15 molec cm^{-2}.

4.2.3. **SO_2 data product requirements**

SO_2 data product requirements are reproduced here from [RD31:PVPv1] and are taken originally from [GCOS]. These also served as input to [RD4:URD]. It is usual practice to express uncertainty in 1-sigma, and specifically for SO_2, to express VCD in DU instead of molecule cm^{-2}, therefore the table is adapted accordingly.

Table 4-3. 2022 GCOS IP requirements for the ECV tropospheric SO_2 column, adapted from [RD3:GCOS_Req]. Specifically, molec cm-2 is adapted to DU and 2-sigma uncertainty to 1-sigma.

QUANTITY	THRESHOLD	BREAKTHROUGH	GOAL
Horizontal resolution	100 km	30 km	10 km
Vertical resolution	N/A	N/A	N/A
Temporal resolution	30 days	1 day	1 hour
Timeliness	30 days	7 days	1 day
Measurement uncertainty (1-sigma)	Max(50%, 0.37 DU)	Max(30%, 0.22 DU)	Max(15%, 0.11 DU)
Stability	Max(20%, 0.15 DU/decade)	Max(12%, 0.089 DU/decade)	Max(6%, 0.045 molec cm^{-2}/decade)

The absolute (relative) requirement on measurement uncertainty is applicable below (above) measurand values of 0.73 DU.

4.2.4. CO data product requirements

CO data product requirements are reproduced here from [RD3:GCOS_Req]. These also served as input to [RD4:URD]. It is usual practice to express uncertainty in 1-sigma, therefore the table is adapted accordingly.

Table 4-4. 2022 GCOS IP requirements for the ECV tropospheric CO column, adapted from [RD3:GCOS_Req]. Specifically, 2-sigma uncertainty is adapted to 1-sigma.

QUANTITY	THRESHOLD	BREAKTHROUGH	GOAL
Horizontal resolution	100 km	30 km	10 km
Vertical resolution	N/A	N/A	N/A
Temporal resolution	30 days	1 day	1 hour
Timeliness	30 days	7 days	1 day
Measurement uncertainty (1-sigma)	5 ppb	2.5 ppb	0.5 ppb
Stability	2 ppb/decade	1 ppb/decade	<1 ppb/decade

4.2.5. NH_3 data product requirements

At the time of writing, no GCOS requirements were formulated for NH_3 column products, therefore requirements were formulated in [RD4:URD]. NH_3 data product requirements are reproduced here from [RD4:URD]. It is usual practice to express uncertainty in 1-sigma, therefore the table is adapted accordingly.

Table 4-5. Consortium specified user requirements for NH_3 tropospheric column, adapted from the URD [RD4:URD]. Specifically, 2-sigma uncertainty is adapted to 1-sigma.

QUANTITY	THRESHOLD	GOAL
Horizontal resolution	100 km	10 km
Vertical resolution	N/A	N/A
Temporal resolution	1 week	1 hour
Measurement uncertainty (1-sigma)	Max(50%, 5E15 molec cm^{-2})	Max(20%, 1.25E15 molec cm^{-2})
Stability	Max(10%, 2e+15 molec. cm^{-2} /decade)	Max(2%, 1e+15 molec. cm^{-2} /decade)

The absolute (relative) threshold requirement on measurement uncertainty is applicable below (above) measurand values of 10E15 molec cm^{-2}. The absolute (relative) goal requirement on measurement uncertainty holds below (above) measurand values of 6.25E15 molec cm^{-2}.

4.3. Validation metrics

In [RD31:PVPv1] validation metrics were specified for bias, dispersion and stability. See Table 4-6. To keep the number of reporting metrics manageable, we will focus on robust metrics in the compliance tables.

Table 4-6. Specific validation metrics to be derived from the comparison to reference data. We focus on robust metrics in this work.

QUALITY INDICATOR	METRIC
bias	**median(Δ), median(Δr)** mean(Δ)
dispersion	**½IP68(Δ), ½IP68(Δr)** std(Δ)
total uncertainty estimate (encompassing systematic and random error)	RMS(Δ)=√[mean(Δ)^2+std(Δ)^2] **√[median(Δ)^2+½IP68(Δ)^2], √[median(Δr)^2+½IP68(Δr)^2]**
stability	temporal trend of the bias, in percent per decade, using robust method
proportional and constant bias	**Theil-Sen slope and intercept**

We also specify in more detail the drift assessment method as used for HCHO, CO and NH_3 in this work. See section 4.4.

4.4. Drift estimation methodology

To study the drift, a multiple linear regression (MLR) is applied on the time-series ***Y(t)*** of the absolute differences satellite-groundbased at each site. We prefer a multiple linear regression instead of a simple linear one, because for CO and NH_3 (and to a lesser extent for HCHO) the biases show a strong seasonal dependency (see Sect. 8.3.3). So, a seasonal cycle (4 harmonics) is fitted together with the trend (or "drift") **B**:

$$Y(t)= a + \mathbf{B} * t + C_1 * \cos(2\pi * t) + C_2 * \sin(2\pi * t) + C_3 * \cos(4\pi * t) + C_4 * \sin(4\pi * t) \quad \text{(Eq.1)}$$

Note that to reduce the influence of outliers on the drift calculation, a robust multiple linear regression is used (bisquare method).

4.5. Validation of individual components of ECV processing chain

ECV line components are the individual processing blocks by which ECV data products are generated in their interim or final version. For complex processing chains, international standards require to validate or at least verify the good performance of every component and the accuracy of its output.

Limiting validation to the final data product only is not sufficient. The validation – or at least some relevant quality check – of intermediate data products is highly desirable to avoid, e.g., that the apparently good behaviour of the final data product at the end of the chain hides large compensating errors affecting separate components of the data retrieval.

For the round-robin studies, this is accomplished in a limited way by comparing the impact of including/not including the cloud correction of the HCHO and SO_2 product, and by comparison of the prior HCHO and SO_2 profiles of the candidate models (TM5, CAMS, MAGRITTE) with those of MAX-DOAS. See [RD32:PVASR], from which conclusions are taken over in this document (sections 6.3.1.4 (HCHO) and 7.3.1.6 (SO_2)).

For the comprehensive validation of the final CDRs, this will be done for the NO_2 product, by the validation of the stratospheric NO_2 column, known as an important intermediate step in the retrieval of the tropospheric NO_2 column. See section 5.3 and especially the conclusions in section 5.3.1.4.

Similarly, the validation of the L2 products will be the grounding step of the validation of the L3 products (see Section 2.2.4.2) which are a main data deliverable of this project. See section 6.3.2.4 (for HCHO) and 8.3.2.5 (for CO).

4.6. Comparison to independent reference measurements

4.6.1. NO_2 column validation data sources

4.6.1.1. Zenith scattered light DOAS

Serving as a reference for the validation of stratospheric NO_2 from all nadir UV–visible satellites to date since GOME (1995-2011), ZSL-DOAS measurements [RD19:Pommereau1988][RD23:Hendrick2011] are particularly useful for the validation of satellite climate data records and of the stratospheric column subtracted to the total column to obtain the tropospheric column (e.g., the validation of the QA4ECV NO_2 ECV precursor data product [RD36:Compernolle2020]). A key property of these zenith-sky measurements at twilight is the geometrical enhancement of the optical path in the stratosphere [RD18:Solomon1987], which offers high sensitivity to stratospheric absorbers of UV-visible radiation and lower sensitivity to clouds and tropospheric species.

The instruments perform network operations in the context of the Network for the Detection of Atmospheric Composition Change (NDACC). Their geographical distribution is visualized in Figure 4-1. Several of these instruments have data records dating back to the nineties, and even the eighties for a few of them (see Figure 4-2 for the availability of stratospheric NO_2 reference measurements obtained with the network of SAOZ instruments).

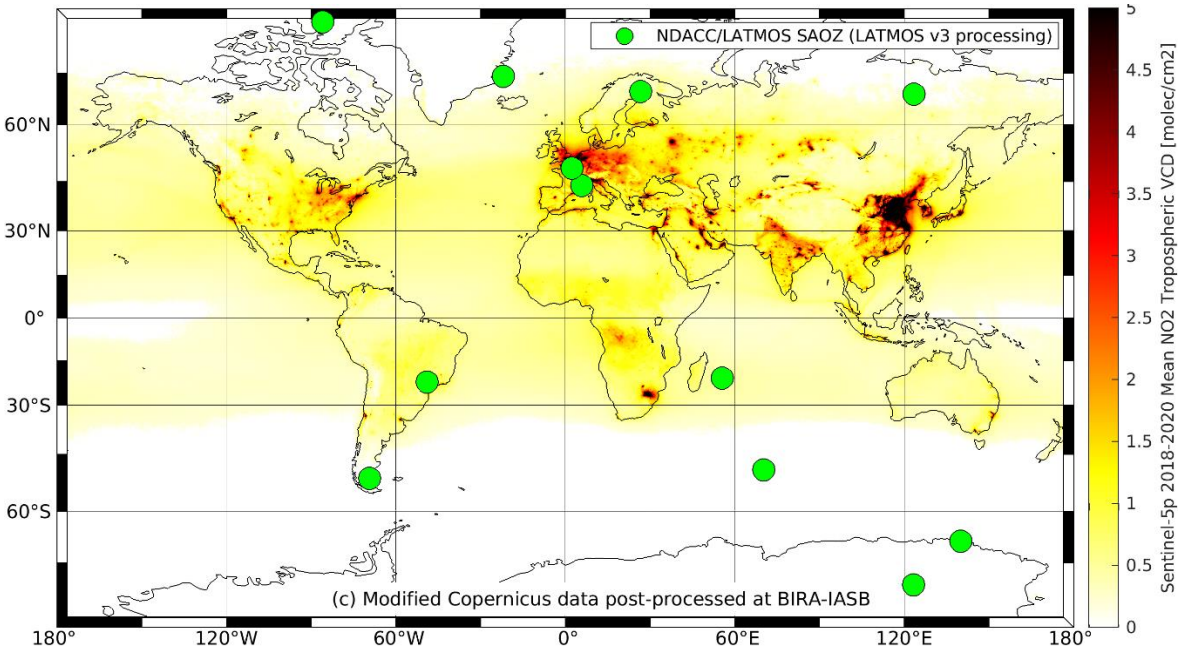


Figure 4-1: Geographical distribution of the NDACC ZSL-DOAS spectrometers as used for the validation of stratospheric NO_2 column data. Background is a multi-annual mean tropospheric vertical column density (VCD) based on S5P NO_2 observations.

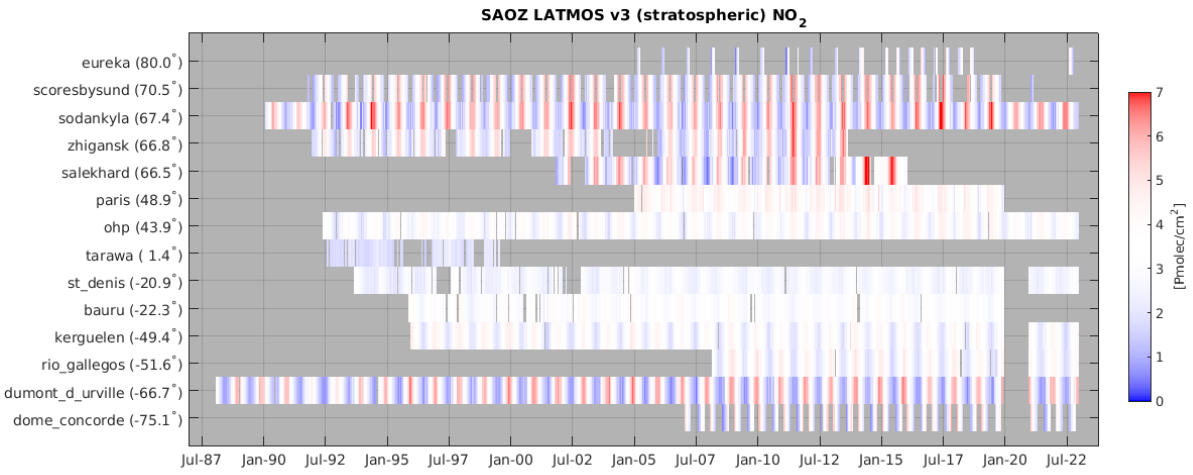


Figure 4-2: Data availability from the SAOZ network of ZSL-DOAS instruments (in consolidated LATMOSv3 processing), as used for the validation of stratospheric NO_2 column data.

NDACC field intercomparison campaigns [RD21:Roscoe1999][RD24:Vandaele2005] conclude to an uncertainty of about 4-7% on the slant column density. Converting the slant column into a vertical column using a zenith-sky AMF, the uncertainty on the vertical column is estimated to be about 10-14 % for the latest data processing version [RD25:Yela2017][RD20:Bognar2019]. A limiting factor comes from the temperature dependence of the NO_2 absorption cross-sections used in the DOAS retrieval of the slant column density. Most of the NDACC instruments use cross-sections at a single temperature of 220 K, which introduces a seasonal error of up to a few percent at middle and high latitudes. See also [RD31:PVPv1] for a discussion on the ZSL-DOAS data.

4.6.2. **HCHO column validation data sources**

Regarding the L2 GOME-2 validation, 24 MAX-DOAS and 27 FTIR sites have been used in [RD37:ACSAF-VAL] over the 2007-2023 time-period, covering a wide range of HCHO levels, including Arctic, oceanic, mountainous remote levels up to polluted city and Amazonian conditions. We refer to that document figures and tables in appendix for the detailed list. Two additional FTIR stations, and no MAX-DOAS, were used for the L3 validation. Their geographical distribution is visualized in Figure 4-3.
More detail about the FTIR network can be found in [RD29:Vigouroux2018]. See also [RD31:PVPv1] for a discussion on the MAX-DOAS and FTIR data [RD31:PVPv1].

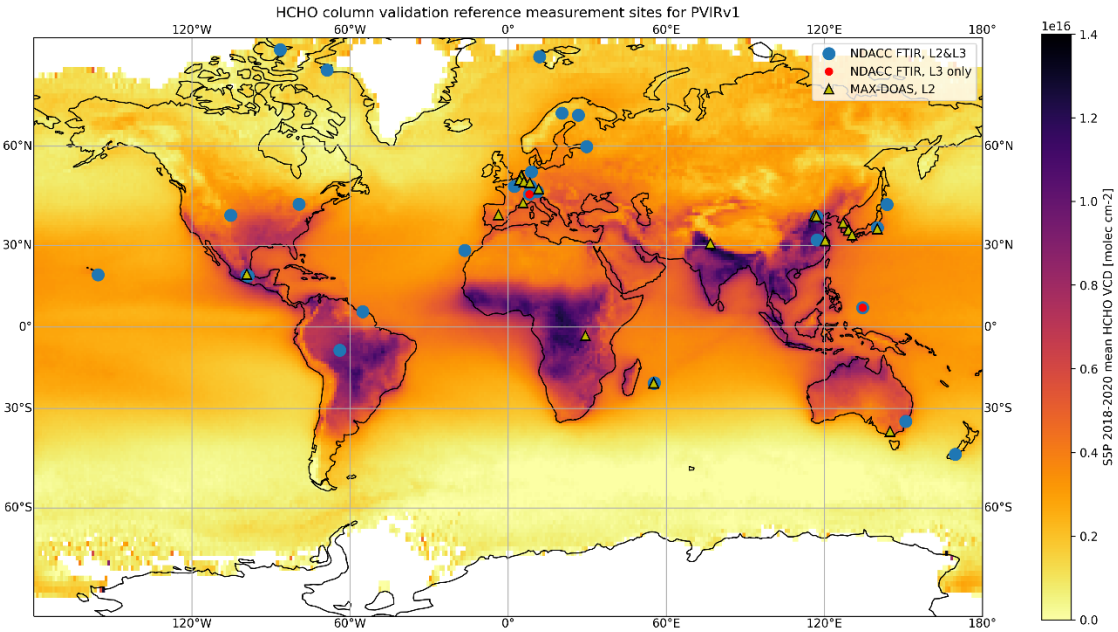


Figure 4-3: Geographical distribution of the NDACC FTIR and MAX-DOAS spectrometers as used for the validation of L2 and or L3 tropospheric HCHO satellite column data in this report. Background is a multi-annual mean tropospheric vertical column density (VCD) based on S5P HCHO observations.

4.6.3. SO_2 column validation data sources

The FRM data used for the current validation is limited to BIRA-IASB MAX-DOAS data from Xianghe (see Figure 4-4) (2010/03/07-2013/12/26, 2014/05/31-2018/7/11, 2019/11/8-2021/10/31). This is a research-grade product, available in GEOMS format. Profiles ranging up to 3-4 km of altitude and with a DOF between 0.7 and 3 are retrieved with the bePRO optimal estimation method. The total column uncertainty is estimated to be 23% [RD31:PVPv1], with the leading term being the uncertainty in the prior profile. Horizontal representativeness of MAX-DOAS depends on aerosol load and can range from a few to tens of km. More information about this MAX-DOAS data can be found in [RD28:Wang2014a].

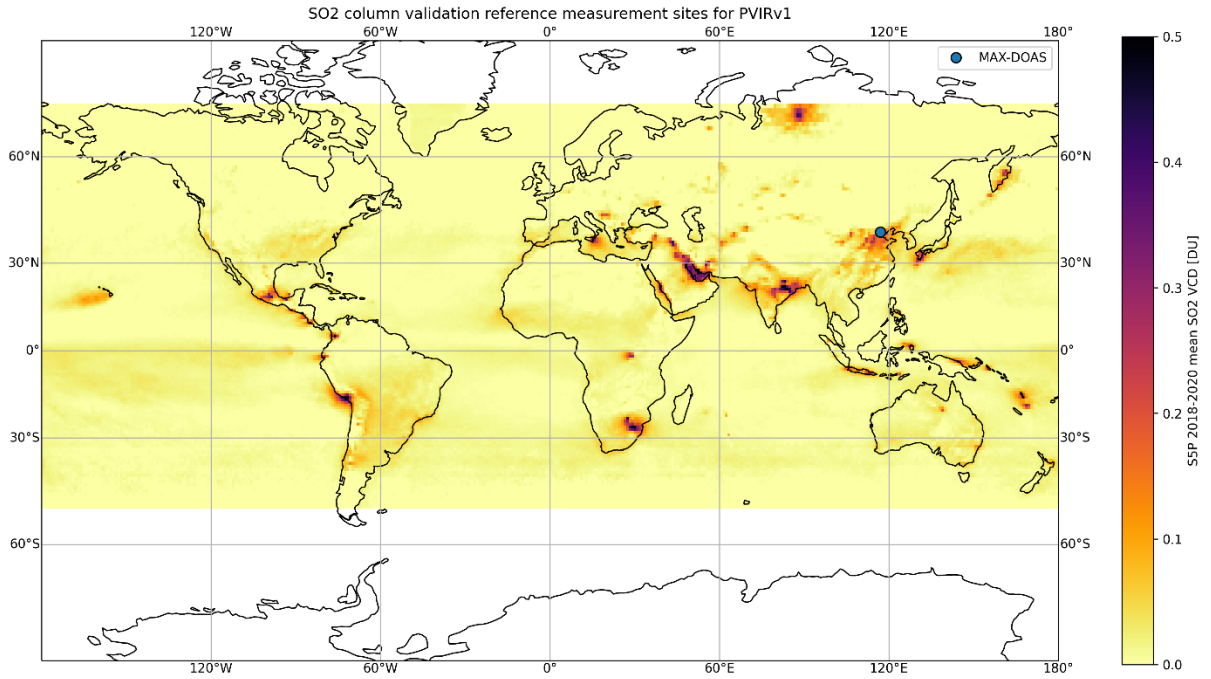


Figure 4-4: The MAX-DOAS spectrometer at Xianghe as used for the validation of L2 SO_2 satellite column data in this report. Background is a multi-annual mean tropospheric vertical column density (VCD) based on S5P SO_2 observations.

4.6.4. **CO column validation data sources**

Many of the NDACC FTIR stations were already active at the beginning of IASI operation in 2007, which allows for the validation of the long-term satellite time-series. These mid-infrared measurements provide CO profiles with about 3 Degree of Freedom of the Signal, in addition to total columns. See also [RD31:PVPv1] for a discussion on the NDACC FTIR CO data. The geographical distribution is displayed in Figure 4-5.

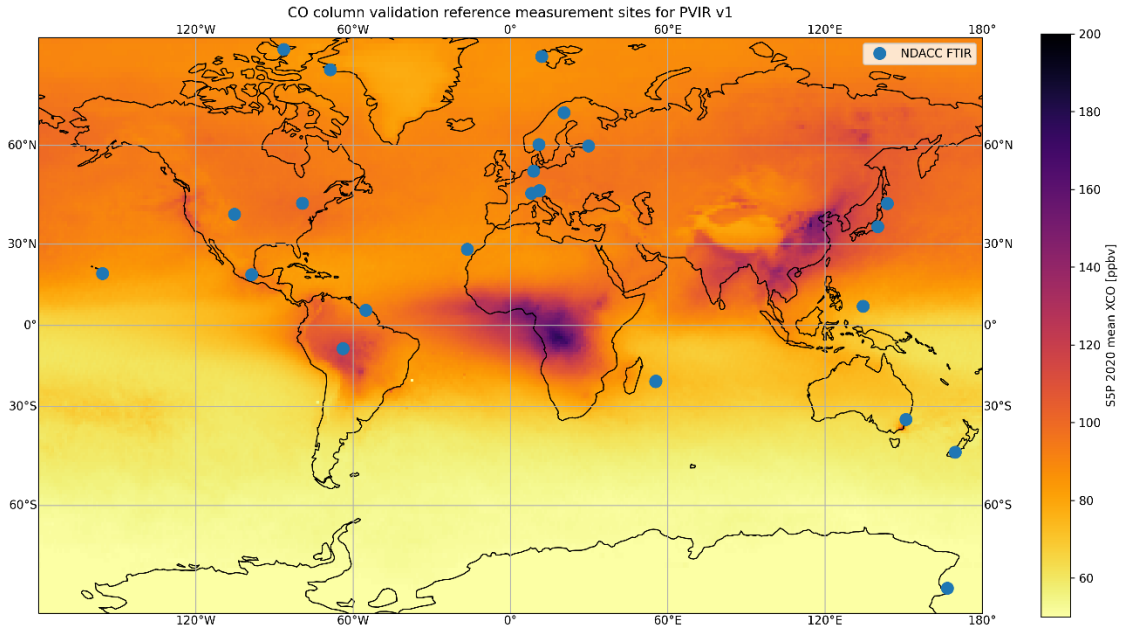


Figure 4-5: Geographical distribution of the NDACC FTIR spectrometers as used for the validation of L3 CO satellite column data in this report. Background is the 2020 annual mean dry-air column VMR based on S5P CO observations.

4.6.5. NH_3 column validation data sources

Regarding NH_3 measurements by FTIR [RD30:Dammers2015], although there is in principle data from over 16 FTIR stations [RD31:PVPv1], only 12 could be used here because the L3 NH_3 files do not provide the necessary information for correcting for the different altitude of the ground-based sites and satellite pixel. It is therefore irrelevant to include the high-altitude stations. The geographical distribution is in Figure 4-6.

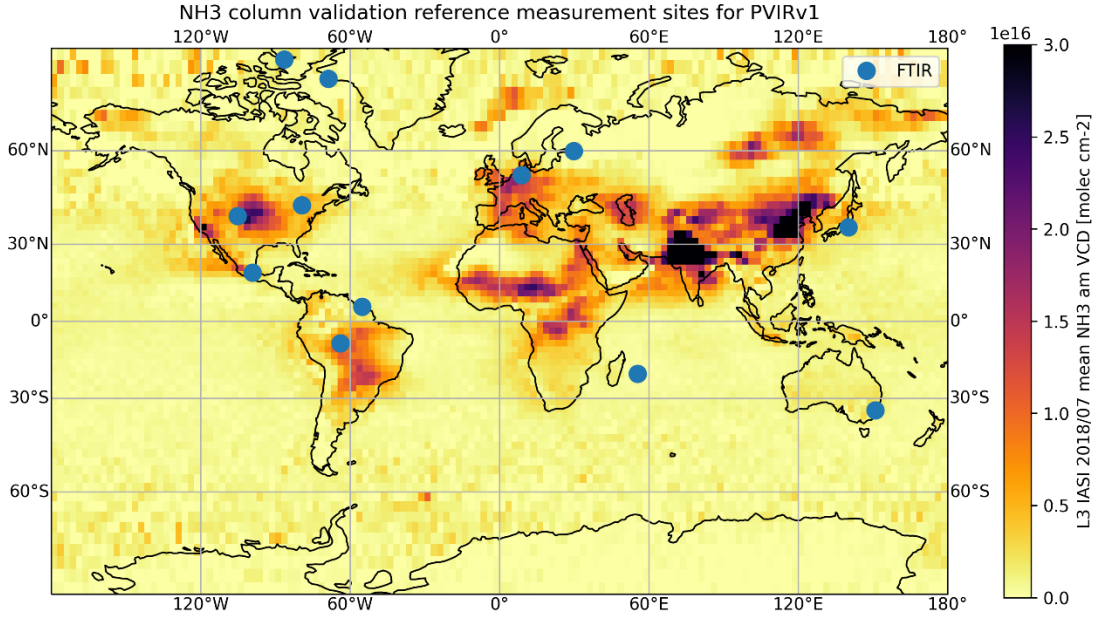


Figure 4-6: Geographical distribution of the NDACC FTIR spectrometers as used for the validation of L3 NH_3 satellite column data in this report. Background is the 2018/07 mean VCD based on the IASI L3 product.

4.6.6. **Error budget of the comparison of atmospheric data**

A major objective of the ground-based validation is to assess the validity of the theoretical (ex-ante, or prognostic) uncertainty estimates provided in the satellite data product. However, the discrepancy between the satellite dataset being validated and the reference dataset combines uncertainties associated with each individual measurement system, plus uncertainties associated with the methodology of data comparison. Uncertainties associated with the comparison methodology include the following sources of co-location mismatch:

(i) Comparison uncertainties associated with the difference in sampling of atmospheric variability and structures: e.g., geographical mismatch, diurnal cycle effects, assumptions related to the area of representativeness.
(ii) Comparison uncertainties associated with the difference in smoothing of atmospheric variability and structures: e.g., MAX-DOAS measurements sampling an air mass of a few km long, compared with GOME ground pixels of 40 x 320 km^2 or TROPOMI ground pixels of 5.5 x 3.5 km^2.

As much as possible, comparison uncertainties due to co-location mismatch will be reduced by a cautious design of the selection of data sets to be compared, and by considering that a multivariate analysis of the comparison results taking into account the specifics of the data being compared (modelling data or remote sensing data, atmospheric variability and gradients etc.) might be required and preferred over entirely statistical approaches. For traceability purposes it is essential to document, for each validation exercise, the selection method applied to the data sets (temporal and spatial co-location criteria, how differences in vertical and horizontal smoothing are handled etc.) [RD44:Keppens2019].

Part of discrepancy between satellite and ground-based data lie in their different vertical sensitivity. This can also cause differences between comparisons with different types of reference data (e.g., MAX-DOAS and FTIR). In this work therefore, if the ground-based data are profiles, both *direct comparisons* and *comparisons smoothed using satellite averaging kernel* [RD47:Rodgers2003][RD49:Eskes2003] are presented.

Obviously, for a detailed uncertainty budget to be made, prognostic uncertainties have to be included in the data sets, at least with distinction between random and systematic error contributions.

Although essential, the closure of a complete error budget for each data comparison is still a matter of research at the time being [RD45:Verhoelst2015] and it falls partly beyond the scope of the Precursors_cci+ precursor project. Validation researchers as well as data producers are aware that neglecting uncertainties linked to the comparison method can yield erroneous conclusions on the quality of the compared data product. This awareness must be transmitted to the reader of Precursors_cci+ PVIR for a proper use of the validation results and, in fine, of the ozone precursor ECV data records. When misinterpretation is possible, common statements like “the discrepancy between the two data sets ranges within their individual

error bars" will be suitably annexed with a provision on the – actual calculated or simply expected – contribution of the selection and comparison methods to this discrepancy. Provisions like "temporal and spatial mismatches exist but their contribution to the discrepancy between the two data sets has not been assessed; nevertheless, this contribution is assumed to be small…" or "the selection method has been optimised to reduce apparent discrepancies between the data products, that would be generated actually by temporal and spatial mismatches and by differences in smoothing of atmospheric variability" are acceptable examples.

5. Validation of NO_2 Data Products

5.1. Scope and generalities

For the current version of this document, the ground-based validation of the NO_2 products is limited to the L2 stratospheric vertical column densities (VCD) of the sensors listed in Table 3-1: GOME, SCIAMACHY, GOME-2A, TROPOMI. Reprocessed OMI/EOS Aura data (based on L1 v4) were not yet available at the time of the validation analysis. The QA4ECV processors provide both default TM5 data assimilation and alternative STREAM stratospheric NO_2, which are both considered. The AC SAF processor for GOME-2A considered here is an intermediate prototype. The processors considered are those that are most likely to be used for the L3 CDR creation.

5.2. Validation methodology

5.2.1. Stratospheric NO_2 column validation by ZSL-DOAS

To account for effects of the photochemical diurnal cycle of stratospheric NO_2, the ZSL-DOAS measurements, which are obtained two times a day at twilight, are adjusted to the satellite overpass time using a model-based factor. This is calculated with the PSCBOX 1D stacked-box photochemical model [RD48:Errera2001][RD22:Hendrick2004], initiated with daily fields from the SLIMCAT chemistry-transport model (CTM). The amplitude of the adjustment depends strongly on the effective SZA assigned to the ZSL-DOAS measurements, which is here taken to be 89°. The uncertainty related to this adjustment is in the order of 10%. To reduce mismatch errors due to the horizontal smoothing differences between satellite and ZSL-DOAS measurements, satellite NO_2 values (from ground pixels at high resolution) are averaged over the air mass footprint where ground-based zenith-sky measurements are sensitive.

The validation of the stratospheric NO_2 products follows the methodology already published in [RD36:Compernolle2020] and [RD35:Verhoelst2021], which builds upon previous developments in the AC SAF, H2020 QA4ECV, and the S5P MPC VDAF.

Filtering. Satellite data were only filtered on general quality flags (e.g., requiring that no processing error flag is raised) and on solar zenith angle (SZA<90 degrees). The QA4ECV data were additionally filtered on the snow-ice flag for some of the analyses, as recommended for tropospheric VCD[RD8:PSD_QA4ECVNO2]. No filtering was done on the ZSL-DOAS data.

Co-location

- Satellite data are co-located with sunrise and sunset ZSL-DOAS measurements separately, allowing at most 12 hours time difference. Analysis of the sounders with a morning equatorial overpass time are preferentially based on the sunrise ZSL-DOAS data, and vice versa for the afternoon sounders.

- Spatial co-location is based on so-called observation operators: 2-D (latitude, longitude) polygons that represent a parametrization of the area of sensitivity of a ZSL-DOAS measurements, towards the rising or setting sun. For the relatively low spatial resolution GOME and SCIAMACHY data, only an intersection between satellite pixel and this observation operator is required. For higher-resolution data sets, the center of the satellite pixel is required to fall within the observation operator.
- The satellite pixels satisfying this spatial and temporal constraint are averaged, per ground-based reference measurement such that a ground-based measurement is only used once (and compared to an averaged satellite measurement satisfying the co-location criteria).

Harmonization. ZSL-DOAS measurements are adjusted for the diurnal photochemistry to the satellite overpass time (this is the averaged time of the pixels averaged in the co-location procedure) using an interpolation of a look-up table of photochemical adjustment factors (in latitude, day-of-year, and time of day) based on a 1-D photochemical box model (PSC-BOX).

Analysis. Further analysis consists of statistical analysis of the set of the differences derived from these co-located and harmonized measurement pairs, treating sunrise and sunset comparisons separately.

5.3. Validation results

5.3.1. Level-2 products

The validation analysis at individual stations is illustrated for the QA4ECV DA GOME product in Figure 5-1Figure 5-2. Both the excellent temporal correlation and the (ubiquitous) small negative bias at moderate latitudes are evident in these illustrations.

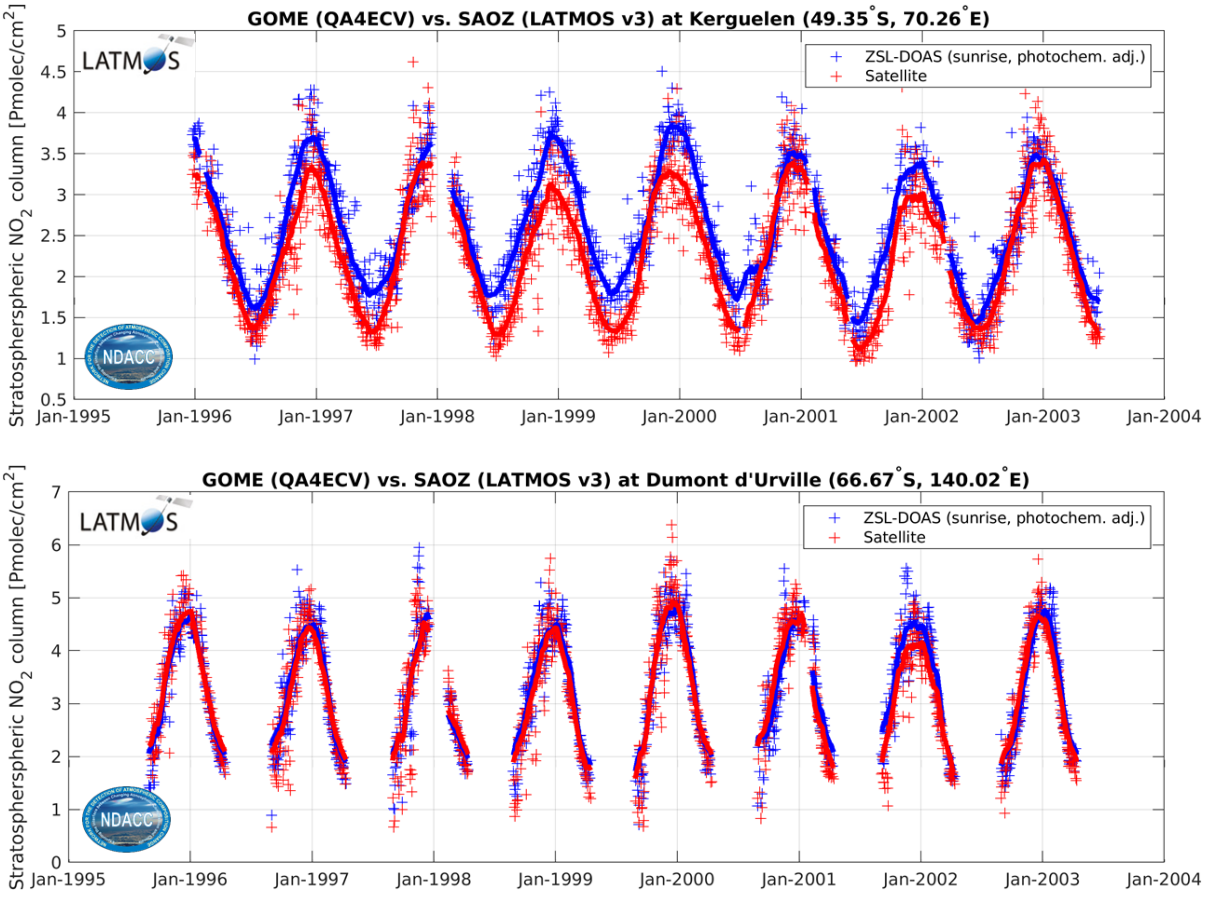


Figure 5-1: Illustration of typical comparison time series, here for GOME on ERS-2 versus the SAOZ on the Kerguelen Archipelago in the Southern Indian Ocean (top panel) and versus the SAOZ at Dumont d'Urville on the East coast of Antactica (bottom panel). Data gaps in the latter correspond to polar night.

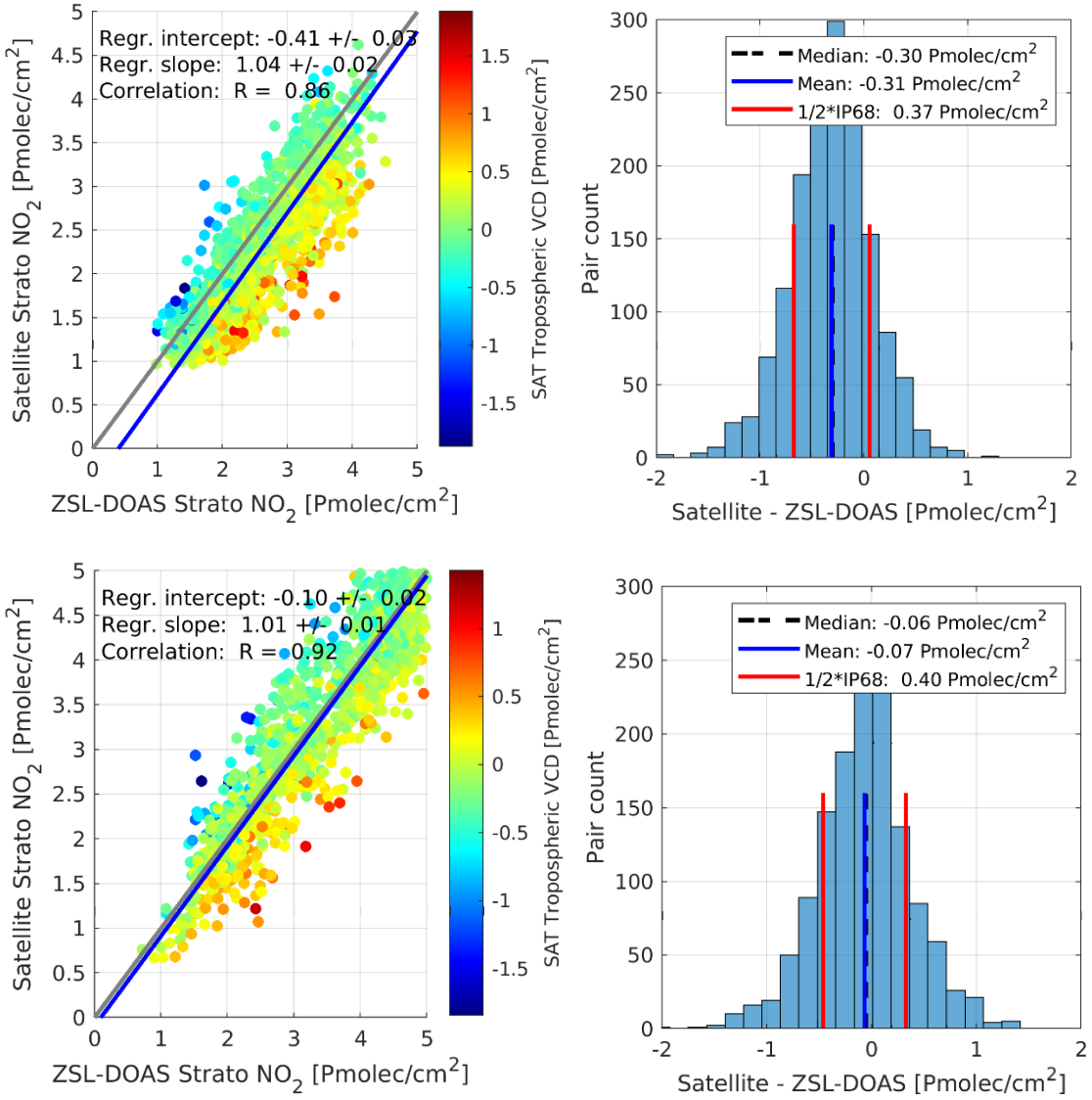


Figure 5-2: Scatter plots with regression analysis (left-hand panels) and histograms of differences (right-hand panels) for the comparisons at Kerguelen (top row) and Dumont d'Urville (bottom row) already visualized in Figure 5-1.

The histograms in Figure 5-2 demonstrate the Normal distribution of the differences with almost equal means and medians, and symmetrical locations of the 16 and 84% percentiles. The scatter plots demonstrate the high correlation (typically around 0.9) and the near-unity slope of the regression. The small negative bias observed at Kerguelen (and at many other mid-latitude stations) appears to be of additive nature. The colour scale in these scatter plots refers to the tropospheric columns derived from these satellite measurements. As overestimated satellite stratospheric columns correspond to negative tropospheric columns and vice versa, it appears that the total column is in fact more accurate than the stratospheric columns, and the noise on the - essentially zero - tropospheric columns is dominated by the uncertainties in the stratosphere-troposphere separation. While shown here only for the QA4ECV GOME data, these results hold for all products at low to mid-latitudes and clean

conditions. More complex behavior for some products at high latitudes is discussed in the following sections.

5.3.1.1. Bias, dispersion and total uncertainty

The results for all considered comparisons to SAOZ instruments are summarized as per-station box-whisker plots in Figure 5-3Figure 5-5, for GOME, SCIAMACHY, and GOME-2A respectively. The general picture that emerges is one of small negative mean differences, of the order of -0.2 to -0.3 Pmolec/cm2, at low and middle latitudes, albeit with a station-to-station scatter of similar magnitude. Also, the dispersion is of this magnitude. Results at higher latitudes do not always follow these common findings, with stronger differences between products.
In Figure 5-3, sunrise comparisons are compared to sunset comparisons for GOME, revealing excellent agreement, except at Bauru (Brazil) where tropospheric anthropogenic pollution is known to affect the sunrise SAOZ measurements. This is reflected in slightly different network-wide mean biases, but the median difference is unaffected.

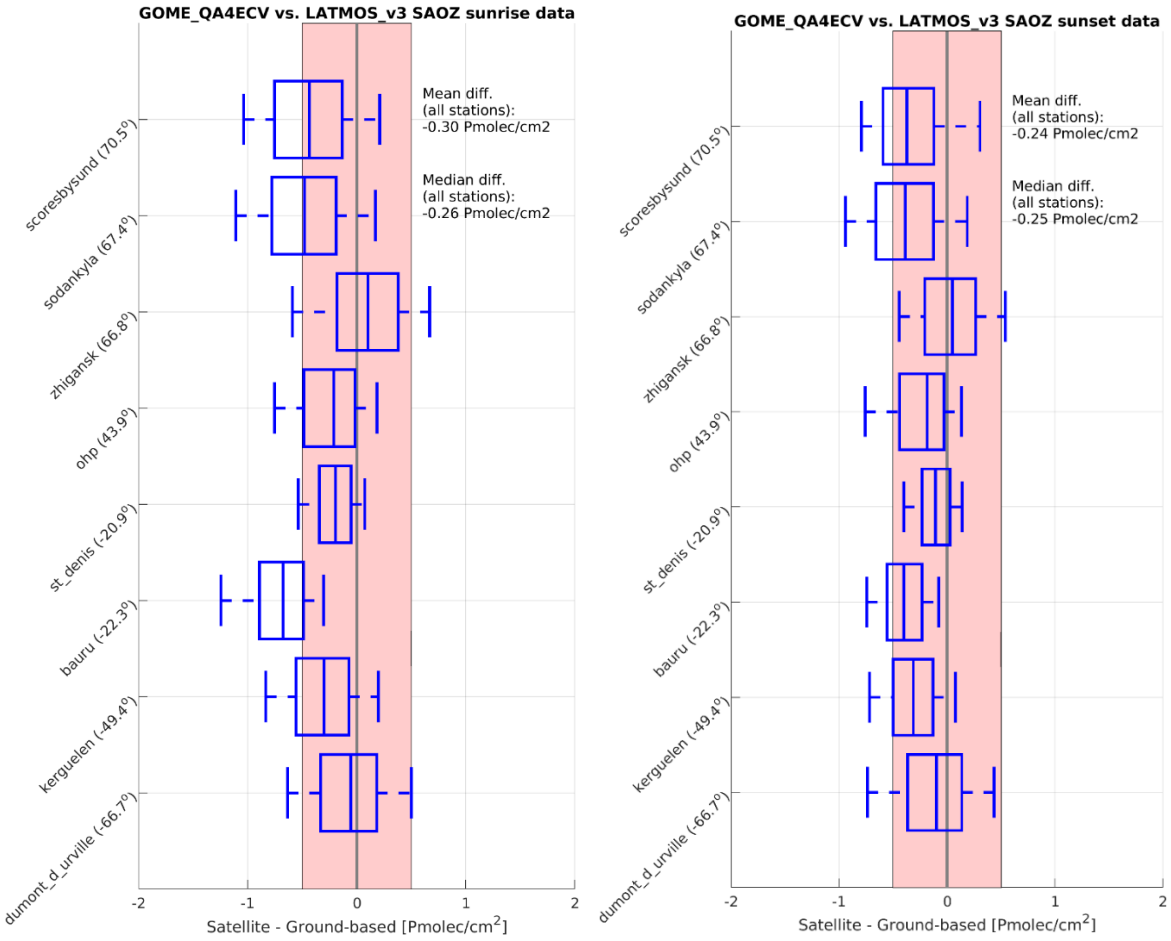


Figure 5-3: Network-wide statistics for the QA4ECV DA product of GOME on ERS-2, either versus sunrise SAOZ data (left-hand panel) or versus sunset SAOZ data (right-hand panel). Stations are ordered from South (bottom) to North (top). The boxes cover 25 – 75% of the differences, the whiskers extend from 9 to 91%.

Figure 5-4 is similar to Figure 5-3 but it compares QA4ECV DA retrievals from SCIAMACHY observations with QA4ECV STREAM retrievals. As for GOME, some strongly negative station mean differences can be understood from a tropospheric pollution signal in the SAOZ data (Bauru and Paris). More concerning are the large differences between DA and STREAM at Southern high latitudes (Dumont d'Urville and Dome Concorde), and at the Northernmost site (Eureka). These differences between both processors are also observed (even to a larger extent) for GOME-2A (not shown here, but details are available in [RD32:PVASR]).

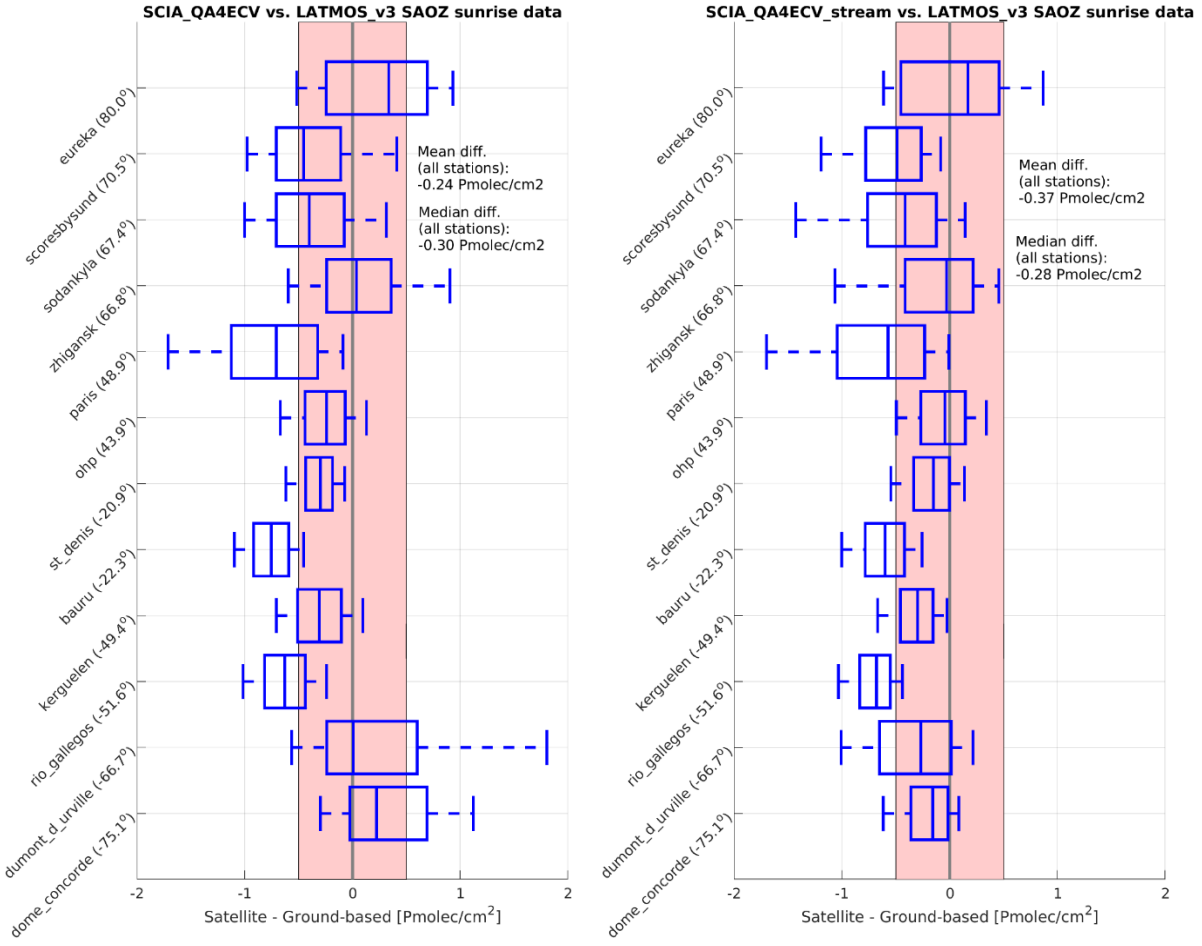


Figure 5-4: Network-wide statistics for the QA4ECV DA (left-hand panel) and QA4ECV STREAM (right-hand panel) products of SCIAMACHY on ENVISAT.

These positive biases and large dispersion (w.r.t. the SAOZ measurements) in the SCIAMACHY and GOME-2(A) QA4ECV DA data at these high latitude sites are caused by a strong overestimations of the stratospheric column in polar summer, and seems to be confined to pixels with a raised snow-ice flag. The total column in these conditions is of normal magnitude and the overestimated stratospheric column therefor results in strongly negative tropospheric columns. This issue is believed to be related to an assimilation of data in the descending part of the orbit for these products and is the topic of further investigation, and potentially a reprocessing.

Figure 5-5 is similar to Figure 5-3Figure 5-4 but compares the AC SAF prototype GOME-2A product to the QA4ECV STREAM product. The AC SAF product is slightly more negatively biased w.r.t. the SAOZ data than the QA4ECV STREAM product, across all latitudes (median difference of -0.27 Pmolec/cm2 versus -0.14 Pmolec/cm2).

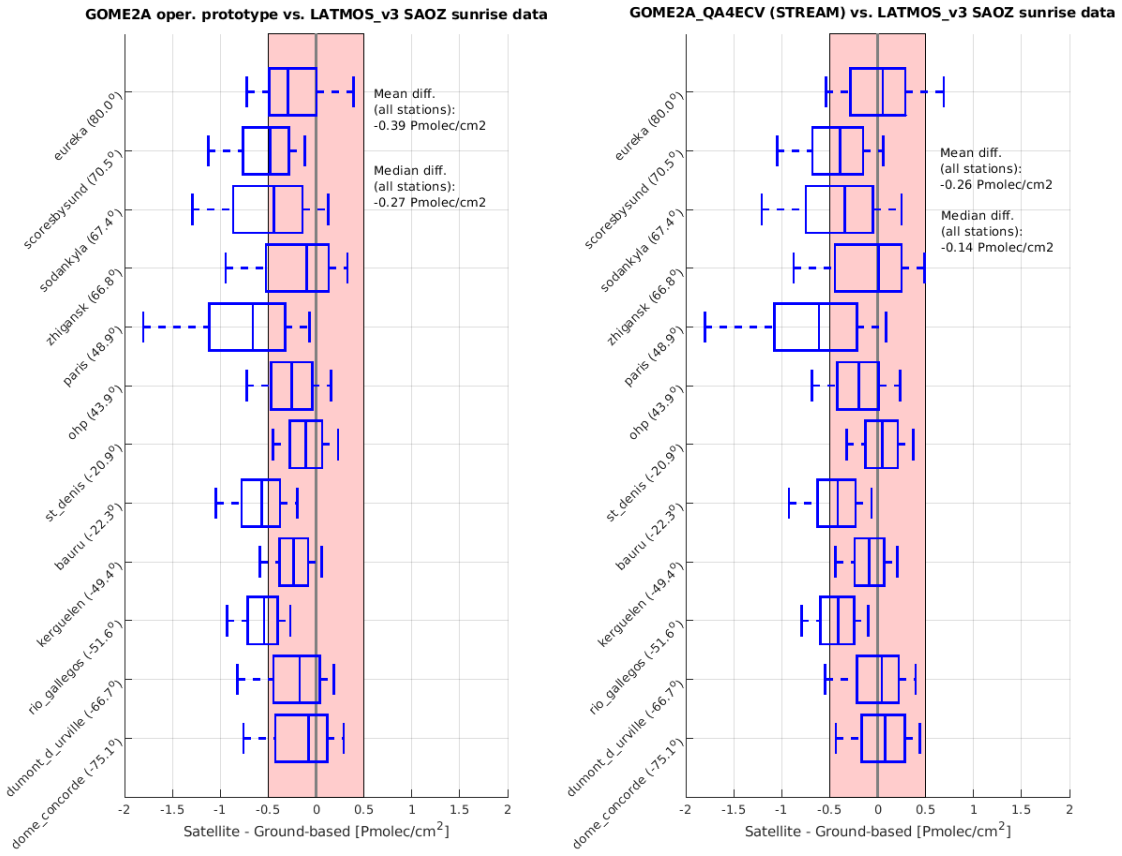


Figure 5-5: Network-wide statistics for the AC SAF prototype (left-hand panel) and QA4ECV STREAM (right-hand panel) products of GOME-2 on Metop-A.

Table 5-1 summarizes the quantitative quality indicators discussed above for each of the products validated hitherto.

Table 5-1: Summary of the Quality Indicators derived from the comparison of various stratospheric NO_2 VCD data sets to ground-based ZSL-DOAS measurements. The bias is computed as the median of the per-station median differences. The dispersion is the median of the per station 1/2IP68 (i.e., half of the 68% interpercentile). The total uncertainty is calculated as the square root of the squared sum of bias and dispersion. Note that all these numbers include ground-based measurement and comparison uncertainties.

Product	Bias (Pmolec/cm2)	Dispersion (Pmolec/cm2)	Total uncert. (Pmolec/cm2)	Comments
GOME – QA4ECV DA	-0.26	0.39	0.47	Only 8 sites
SCIAMACHY – QA4ECV DA	-0.30	0.48	0.57	Issues at high latitudes
SCIAMACHY – QA4ECV STREAM	-0.28	0.37	0.46	
GOME-2A – AC SAF prototype	-0.27	0.37	0.46	
GOME-2A – QA4ECV DA	-0.24	0.41	0.48	Issues at high latitudes
GOME-2A – QA4ECV STREAM	-0.14	0.34	0.37	

TROPOMI – RPRO PAL	-0.14	0.26	0.30	Compared to LATMOS NRT sunset data

5.3.1.2. Stability

For this version of the PVIR, the stability of the records was not yet formally quantified. Perusal of the time series of differences does not suggest the presence of strong drifts or other long-term temporally varying biases, besides the seasonal cycle observed in most of the differences which is caused by the fixed cross sections (as opposed to cross sections varying with the effective temperature) in the ZSL-DOAS retrievals, and at the polar latitudes also by the DA issues for SCIAMACHY and GOME-2 described in 5.3.1.1.

5.3.1.3. Influence quantities

Three influence quantities are of prime interest for stratospheric NO_2 VCD retrievals: (1) the solar zenith angle (SZA), (2) the surface albedo, and (3) the cloud fraction (CF). An initial analysis of a potential dependence of the bias and dispersion on these influence quantities did not reveal any strong features. Some dependence on SZA is observed at mid-latitude sites, but this is believed to be related to the fixed cross sections in the ZSL-DOAS retrievals, leading to a seasonal bias that consequently causes a dependence on SZA. For SCIAMACHY (and GOME-2A) QA4ECV DA data, a strong positive bias is observed at the highest surface albedo, which is in agreement with the issues described for these products at high latitudes correlating with the snow-ice flag.

5.3.1.4. Compliance with user requirements

No user requirements are formulated for the stratospheric NO_2 VCD. Nevertheless, the total uncertainty on the stratospheric columns derived from the ground-based validation can be compared to the requirements formulated for the tropospheric NO_2 VCD and is found to be mostly smaller than the “goal” (i.e., the most stringent) requirement, even without accounting for ground-based measurement and comparison errors. As such, the accuracy of the stratospheric columns (i.e., of the stratosphere-troposphere separation) should not be a critical factor in the compliance of the tropospheric VCD with the user requirements.
One exception to this compliance concerns some of the data assimilation results at snow/ice covered polar latitudes for SCIAMACHY and GOME-2, for which overestimated stratospheric columns lead to strongly negative tropospheric columns that do not satisfy the user requirements. Therefore, for GOME-2A and SCIAMACHY QA4ECV DA data, it is recommended to exclude data using the snow-ice flag.

5.3.2. Level-3 products

No level-3 data was available yet for validation.

5.3.3. Level-3 merged product

No level-3 merged product was available yet for validation.

6. Validation of HCHO Data Products

6.1. Scope and generalities

L2 products

HCHO L2 validation results were previously reported for TROPOMI operational product [RD38:Vigouroux2020][RD39:DeSmedt2021][RD34:ROCVR][RD41:Oomen2024]; QA4ECV OMI [RD40:Mueller2024], and GOME-2(A,B,C) AC SAF RPRO L2 GDP 4.8 [RD37:ACSAF-VAL]. Regarding GOME-2(A,B,C), the main findings with respect to MAX-DOAS and FTIR instruments are summarized in section 6.3.1. For OMI and TROPOMI the coherence with the L3 data is demonstrated.

L3 products

We show in this report the validation of a first version of L3 datasets for most of the instruments involved in the project (SCIAMACHY, GOME-2 A and B, OMI, and TROPOMI). Only GOME, and GOME-2 C L3 are not yet validated. The L3 datasets are validated using the FTIR network [RD29:Vigouroux2018].

Due to the high number of satellite products provided (all cited satellites with both L3 daily and monthly files), choices had to be made, especially knowing that the L3 provided up to now are not the final products.

Another complexity with the HCHO products is that if we want to check the consistency between OMI and TROPOMI, it is advised [RD39:DeSmedt2021] to use the variable "tropospheric_HCHO_column_number_density_clear" (column calculated without applying a cloud correction). So, even if validation has been performed for cloud-corrected products and "clear-sky" ones, we have to limit ourselves in this document, and we show validation results for:

- Daily L3 files for SCIAMACHY, GOME-2A and GOME2-B, TROPOMI; cloud-corrected products.
- Monthly L3 files for TROPOMI and OMI; "clear-sky" products.

6.2. Validation methodology

L2 GOME-2 validation

For the L2 GOME-2 validation shown in Sec. 6.3.1, valid HCHO data within 150km of the ground-based sites are averaged every day and are compared to daily means ground-based columns. Direct VCD comparison is performed every day, and then from the daily coincidences, monthly means are calculated. The difference in vertical sensitivity between the measurement types is also taken into account by applying the satellite column averaging kernels to the ground-based HCHO profiles when available [RD49:Eskes2003]. These comparisons are referred as "smoothed ground-based HCHO VCDs ($VCD_{GB,smoothed}$)" below. We refer to [RD37:ACSAF-VAL] for the detailed methodology.

L3 validation
Regarding the L3 validation, the following steps are applied:

Satellite pixel filtering
The prototype daily L3 files validated at this stage did not yet have a quality filtering available, so no additional filtering is made.
The monthly L3 OMI and TROPOMI files contain a QA value indicating the coverage of the grid cell: we use only L3 data with QA=1.

Co-location
- Daily SCIAMACHY and GOME 2A and B: 150 km collocation criteria in accordance with L2 validation. (see above)
- Daily TROPOMI: we chose the nearest pixel due to the better spatial resolution of TROPOMI. This is rather similar to the 20km criteria used in L2 validation.
- Monthly OMI and TROPOMI: 20 km collocation criteria is used.

Vertical harmonization
Daily SCIAMACHY and GOME 2A and B:
1. The satellite averaging kernels are not yet provided in the L3 files (version fv0100). So only direct comparisons can be performed (without smoothing).
2. The altitude difference between the ground-based measurement and the satellite pixel cannot be corrected (no surface pressure/altitude is provided in the L3 files). Therefore, we must limit the validation to non-elevated stations. Note that even with non -elevated sites, this is not optimal, the altitude correction has proven to provide better validation results for CO also at low altitude sites (e.g. for sites at sea-level but close to mountains).

Daily TROPOMI and monthly OMI and TROPOMI:
1. The satellite averaging kernels are provided in the L3 files (version fv0200). So, the formalism of [RD47:Rodgers2003] is applied (with smoothing and a priori correction; see also [RD38:Vigouroux2020]).
2. The altitude difference between the ground-based measurement and the satellite pixel is corrected by scaling the satellite column with a ratio of the satellite prior partial column between the 2 surface altitudes (pixel and station) and the satellite prior total column.

Comparison pair handling
Averages of satellite pixels are used (for 150km or 20km criteria) or the nearest pixel is used (for TROPOMI daily files) and compared to an average of the FTIR columns measured within the same day (for daily L3) or month (for monthly L3). The regridding/smoothing process for OMI and TROPOMI is done using each FTIR observation before averaging.

Drift estimation methodology: see section 4.4.

6.3. Validation results

6.3.1. Level-2 products

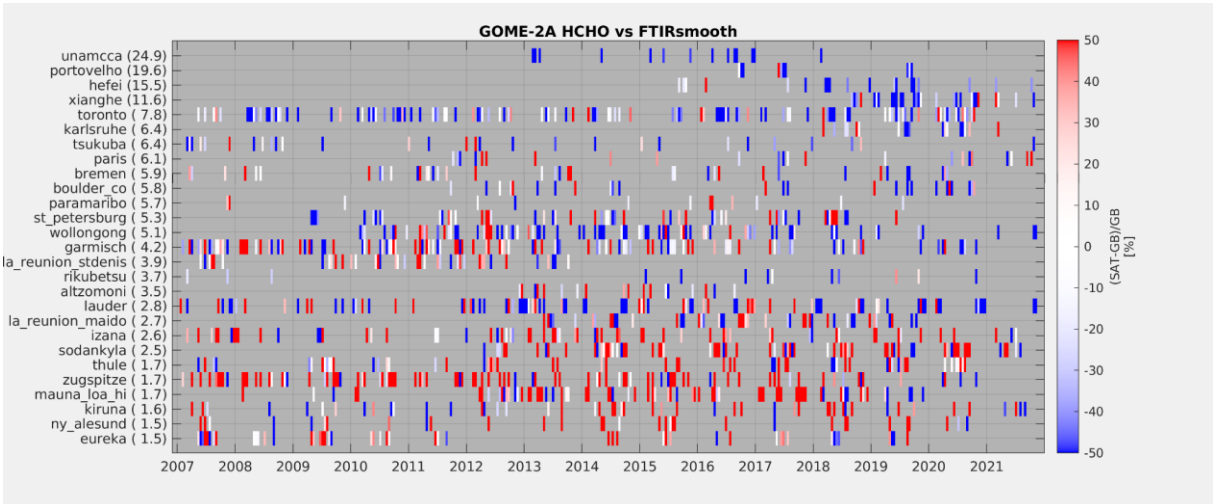
Regarding **L2 S5P/TROPOMI collection 3 ≥v2.4 HCHO validation results** using FTIR, MAX-DOAS and Pandora, we refer the reader to [RD34:ROCVR]. Analyzing daily comparisons of 20 km radius spatial averages with the FTIR network (29 stations), a bias of +32% and -30% is found in clean (<2.5 Pmolec cm^{-2}) and polluted (>8 Pmolec cm^{-2}) conditions respectively. A dispersion (calculated with scaled median absolute deviation from the median (MAD)) of 1.5 Pmolec cm^{-2} and 4 Pmolec cm^{-2} is found in clean and polluted conditions respectively. Theil-Sen regression provides y=0.62x+1.14 Pmolec cm^{-2}, while the Pearson correlation coefficient=0.85.

Regarding **L2 QA4ECV OMI HCHO validation results** using FTIR, we refer the reader to [RD40:Mueller2024]. In this work, monthly means of coincident pairs are analyzed to reduce the noise. Theil-Sen regression provides y=0.659x+2.02 Pmolec cm^{-2}, while the Pearson correlation coefficient=0.67. Similarly as for TROPOMI, a positive bias is reported for clean sites and a negative bias for polluted sites.

Regarding **L2 AC SAF RPRO GOME-2A, -B, -C HCHO**, we summarize below results from the AC SAF validation report [RD37:ACSAF-VAL] and refer to results from [RD32:PVASR].

6.3.1.1. Bias and dispersion

Figure 6-1 presents the relative (100*(SAT-GB))/GB in %) differences for all the available FTIR (a) and MAX-DOAS (b) stations as mosaic time-series plots for RPRO AC SAF GOME-2A L2 data, while the VCD absolute (SAT-GB in $x10^{15}$ molec/cm^2) biases and dispersion are shown as box and whisker plots in Figure 6-2 to Figure 6-3 for GOME-2A, GOME-2B and GOME-2C datasets vs FTIR and MAXDOAS, respectively.



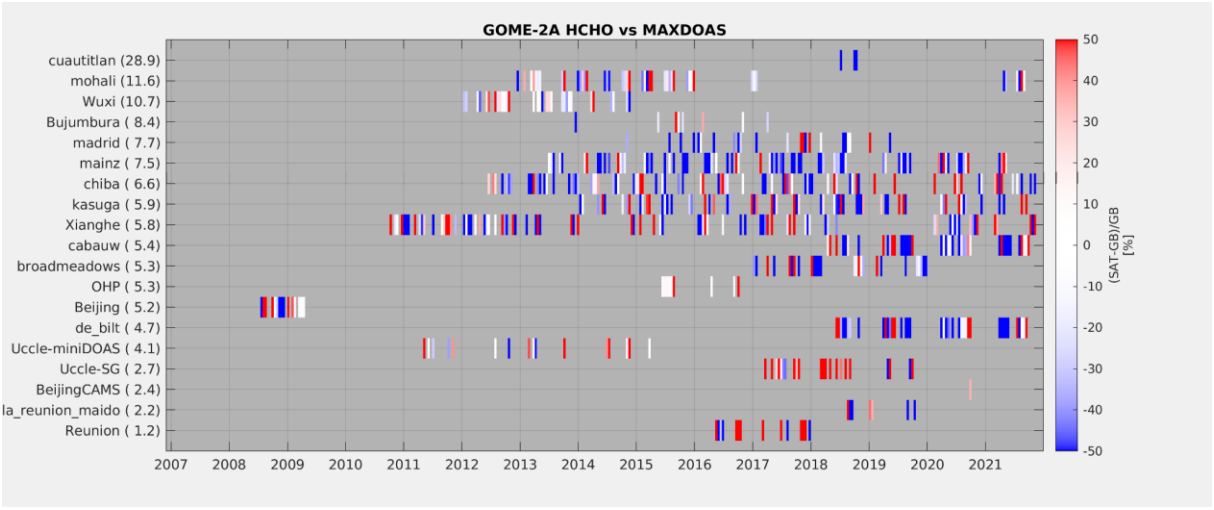


Figure 6-1. Mosaic plot of the bi-weekly averages (SAT-GB)/GB relative differences [%] between GOME-2A AC SAF L2 and FTIR HCHO smoothed column data (upper row), and MAXDOAS columns (bottom row). Stations are ordered from clean to polluted (bottom to top) and their median HCHO content in Pmolec/cm² is given between brackets.

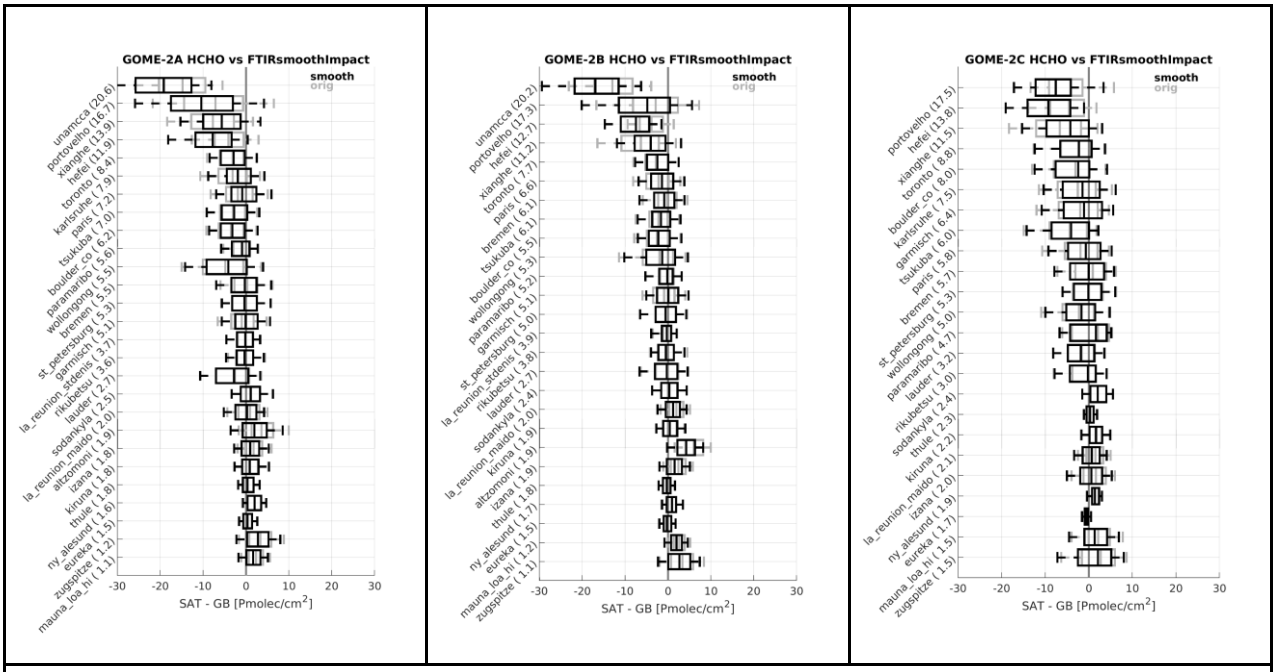


Figure 6-2: Box-and-whisker plot of the absolute GOME-2 minus FTIR ground-based data for the validation of GOME-2A (left), B (middle) and C (right) HCHO L2 AC SAF datasets. Original column comparisons are shown in grey and the smoothed comparisons are overlaid in black. The median biases are shown as horizontal lines, the 25th and 75th percentiles as boxes and 9th and 91th percentiles as whiskers. Stations are ordered from clean to polluted (bottom to top) and their median HCHO content in Pmolec/cm² is given between brackets.

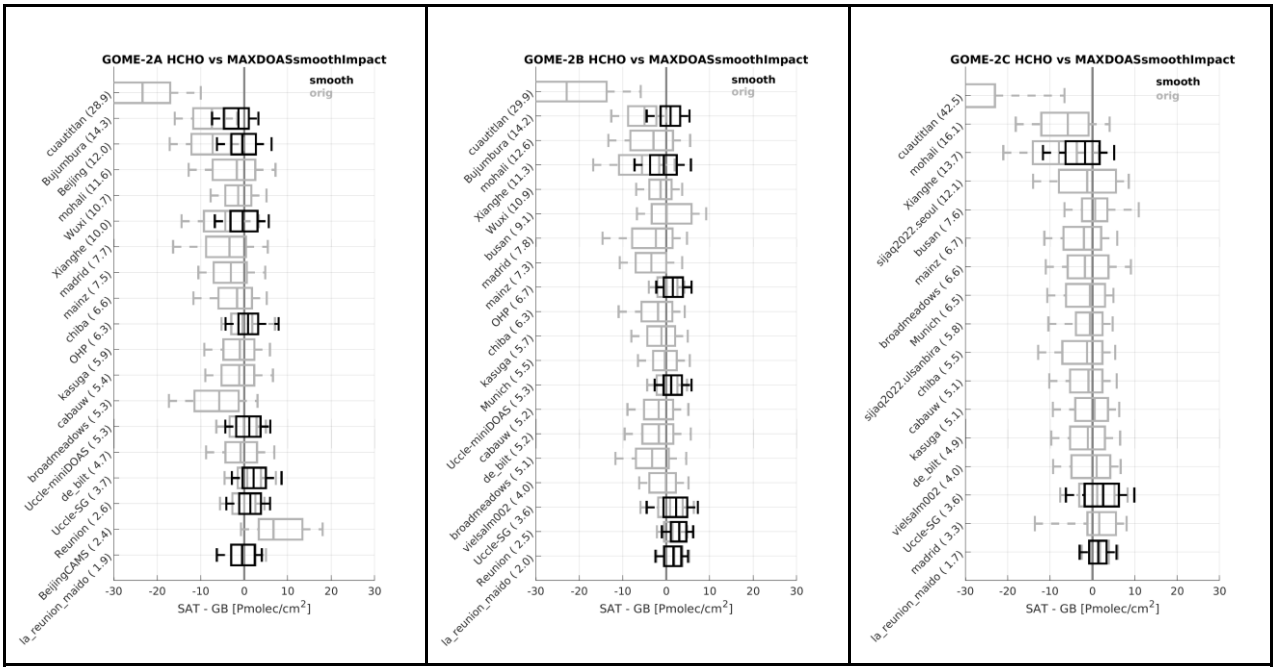


Figure 6-3: Box-and-whisker plot of the absolute GOME-2 minus MAXDOAS ground-based data for the validation of GOME-2A (left), B (middle) and C (right) HCHO L2 AC SAF datasets. Original column comparisons are shown in grey and the smoothed MAXDOAS comparisons are overlaid in black when these are possibles. The median biases are shown as horizontal lines, the 25th and 75th percentiles as boxes and 9th and 91th percentiles as whiskers. Stations are ordered from clean to polluted (bottom to top) and their median HCHO content in Pmolec/cm² is given between brackets.

As seen from the above figures, there is a tendency for the 3 GOME-2 instruments to have positive biases for relatively clean sites and negative biases for more polluted sites, as also seen for TROPOMI with MAX-DOAS [RD39:DeSmedt2021] and with FTIR [RD38:Vigouroux2020]. All the robust and non-robust statistics can be found for each site in [RD37:ACSAF-VAL]. We quantify this further here by separating all the comparisons daily pairs from all the stations into 2 categories: high emission conditions (> 8 x10^{15} molec/cm^2) and clean conditions (< 2.5 x10^{15} molec/cm^2), looking at the ground-based columns values for the separation. This is done for the original MAX-DOAS and for both the original and the smoothed FTIR. Results are summarised below for robust statistics:

Bias

- Clean cases (<2.5e15): 2e15, 2.1e15 and 2.6e15 molec/cm^2 median absolute biases for MAXDOAS wrt GOME-2A/B/C, i.e. 89.3%, 97% and 114.2%. Values of 1.3e15, 1.5e15 and 1e15 are found for the original FTIR (i.e., 83.4%, 99.7%, 60.6%) and of 0.91e15, 0.9e15 and 0.8e15 for the smoothed FTIR (i.e., 58.5%, 53%, 48.7%).
- Polluted cases (>8e15): -4.6e15, -4.5e15 and -4.9e15 median absolute biases for MAXDOAS, i.e. -37.1%, -36.7% and -36%. Values of -5.3e15, -5.1e15, -6.4e15 are found for the original FTIR (ie -42%, -37.7%, -43%) and of -5.3e15, -5.2e15, -6.1e15 for the smoothed FTIR (i.e., -40.5%, -38.03%, -40.3%).

Dispersion

- Clean cases: 4.49e15, 3.94e15 and 5.1e15 ½IP68 for MAXDOAS wrt GOME-2A/B/C, i.e. 321%, 272% and 373%. Values of 3.1e15, 3e15 and 3.7e15 are found for the original FTIR (i.e., 240%, 220%, 220%) and of 2.9e15, 2.9e15 and 3.9e15 for the smoothed FTIR (i.e., 200%, 180%, 230%).
- Polluted cases: 6e15, 5.9e15 and 6.7e15 ½IP68 for MAXDOAS wrt GOME-2A/B/C, i.e. 41.8%, 42% and 47%. Values of 6e15, 5.8e15 and 6.6e15 are found for the original FTIR (i.e., 38%, 34% and 42%) and of 6.2e15, 6e15 and 6.7e15 for the smoothed FTIR (i.e., 38%, 33% and 41%).

Biases are generally coherent between the 3 GOME-2 instruments, with slightly larger values for GOME-2C. Differences with respect to the MAX-DOAS are slightly larger than those wrt FTIR for the clean sites (where the impact of smoothing the FTIR is the largest) and slightly smaller for the polluted sites, where the original and smoothed FTIR results are very similar. The biases are of typically ~-40% in polluted conditions and between 50 to 100% in clean conditions.
Dispersion is also generally coherent between the 3 GOME-2 instruments, with slightly larger values for GOME-2C. Dispersion for the MAX-DOAS cases is larger than for the FTIR in clean conditions, while for polluted conditions values are similar ~6e15 molec/cm^2.
In polluted conditions, the numbers on bias and dispersion are very coherent for both techniques, and when smoothing the FTIR using the GOME-2 averaging kernel.

Table 6-1: Overview statistics of the GOME-2 L2 vs MAX-DOAS ground-based data. For bias and dispersion, the daily pairs of all the stations are separated by pollution level (see text) while monthly means are used for the regression statistics. PMC=E15 molec/cm^2

L2 AC SAF vs MAXDOAS	Bias (median) Daily pairs	Dispersion (1/2IP68) Daily pairs	Total uncertainty √(med^2+1/2IP68^2) Daily pairs	Regression (TS, Pearson) Monthly means
GOME2A 150 km av	Clean: 2PMC; 89.3% Polluted: -4.6PMC;-37.1%	Clean: 4.5 PMC Polluted: 6 PMC; 42%	Clean: 5PMC Polluted: 7.5PMC; 56%	y = 0.37x+2.4PMC R = 0.054
GOME2B 150 km av	Clean: 2.1PMC; 97% Polluted: -4.5PMC; -36.7%	Clean: 4 PMC Polluted: 6 PMC; 42%	Clean: 4.5PMC Polluted: 7.5PMC; 56%	y = 0.43x+2.4PMC R = 0.68
GOME2C 150 km av	Clean: 2.6PMC; 114.2% Polluted: -4.9PMC; -36%	Clean: 5.1 PMC Polluted: 6.7PMC; 47%	Clean: 5.7PMC Polluted: 8.3PMC; 59%	y = 0.4x+1.9PMC, R = 0.6

Table 6-2: Overview statistics of the GOME-2, OMI and S5P L2 vs FTIR ground-based smoothed data. For bias and dispersion, the daily pairs of all the stations are separated by pollution level (see text) while monthly means are used for the regression statistics. PMC=E15 molec/cm^2

L2 vs smoothed FTIR	Bias (median)	Dispersion (1/2IP68)	Total uncertainty √(med^2+1/2IP68^2)	Regression (TS, Pearson) Monthly means

AC SAF GOME2A 150 km av Daily pairs	Clean: 0.91PMC; 58.3% Polluted: -5.3PMC; -40.5%	Clean: 2.9PMC Polluted: 6.2PMC; 38%	Clean: 3 PMC Polluted: 8.2 PMC; 55.5%	y = 0.42x+2.2 PMC R = 0.055
AC SAF GOME2B 150 km av Daily pairs	Clean: 0.9PMC;53% Polluted: -5.2PMC; -38%	Clean: 2.9PMC Polluted: 6e15; 33%	Clean: 3 PMC Polluted: 8 PMC; 50.3%	y = 0.51x+2 PMC R = 0.7
AC SAF GOME2C 150 km av Daily pairs	Clean: 0.8PMC; 49% Polluted: -6.1PMC; -40.3%	Clean: 3.9PMC Polluted: 6.7PMC; 41%	Clean: 4 PMC Polluted: 9 PMC; 57.5%	y = 0.42x+1.8 PMC R = 0.64
QA4ECV OMI Monthly means	Not provided	Not provided	Not provided	y=0.659x+2.02PMC R=0.67
TROPOMI ≥v2.4 20 km av Daily pairs	Clean: +32% Polluted: -30%	Clean: 1.5PMC[a] Polluted: 4PMC[a]	Not provided	y=0.62x+1.14 PMC R=0.85

a. In [RD34:ROCVR] the dispersion is not calculated with ½ IP68, but with the scaled MAD: 1.4826 med$|x_i$-med($x_j)|$. Both return 1 standard deviation for a normal distribution.

6.3.1.2. Stability

This has not been quantitatively assessed here, but the quality of GOME-2A is clearly degraded in the last years of operation. This is often clear in the time-series, see e.g. Figure 6-4 for the comparisons in Lauder, where the effect is very strong. This is also leading to very small values of the overall Pearson correlation coefficient R (see Table 6-1 and Table 6-2), which is strongly influenced by outliers. The Theil-Sen regression is less affected by outliers and result in regression values coherent with GOME-2 B and GOME-C.

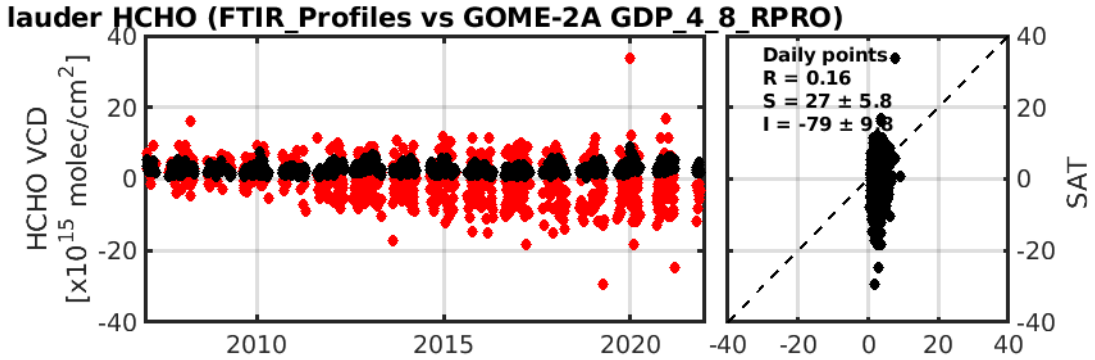


Figure 6-4: Time-series of the daily FTIR (black) and GOME-2A (red) HCHO L2 AC SAF datasets.

The degradation is mentioned in [RD37:ACSAF-VAL]: "However, the GOME-2A HCHO VCDs present clear signs of degradation after 6 years of operation, with a noticeable negative bias compared to GOME-2B results. This is more pronounced in regions like Australia, Africa and Central and South America. On the contrary, the GOME-2B and C time series present a good level of agreement in mostly all the regions, even after 8 years of GOME-2B operations."

The OCRA cloud fraction, which is input to the GOME-2 A/B/C HCHO product, also exhibits a stability problem. Again from [RD37:ACSAF-VAL]: "The cloud fractions present an artificial increase from 2018 in the three datasets. This anomaly will have an impact on the data selection, to the least."

6.3.1.3. Influence quantities

This has not been assessed.

6.3.1.4. Intercomparison with alternative EO data sets

In [RD32:PVASR] validation results with MAX-DOAS or FTIR of the GOME-2A AC SAF OFFL default product (very close to the RPRO product) are intercompared with the alternative "clear-sky" product (info available in the same data files). The conclusion is that for this particular product (based on the OCRA-ROCCIN cloud product), the impact of cloud correction on validation results is negligible. Note that this is not always true: for the GOME-2A HCHO QA4ECV product (based on the Fresco cloud product), cloud correction has a significant impact.

6.3.1.5. Compliance with user requirements

Table 7-2 provides an overview of the compliance with GCOS requirements, based on the results of the FTIR smoothed results, considered as the representative with its more extensive network. Using the MAX-DOAS network, mostly the same compliance conclusions would be drawn, except that dispersion and total uncertainty would be 'near goal' instead of 'better than goal.

Compliance with GCOS requirements for GOME-2 A/B/C AC SAF HCHO L2 and S5P data. No requirements were provided for bias and dispersion, but they have been each compared with the GCOS total uncertainty requirement. PMC=E15 molec cm^{-2}.

Quantity	Compliance/evaluation	Requirement T|B|G
Horizontal resolution	GOME2:40x80-40x40km^2, reduced swath period	100 |30 | 10 km
	TROPOMI: 5x3.5 km^2, nadir	
Temporal resolution	Daily, <daily at high lat	30days |1day|1h
Bias	Clean: 0.9 PMC (GOME-2), <0.9 PMC[a] (TROPOMI)	Abs: 20|8|4 PMC Rel: 50%|20%|10%
	Polluted: -40% (GOME-2), -30% (TROPOMI)	
Dispersion	Clean: 3-4 PMC (GOME-2), 1.5 PMC (TROPOMI)	
	Polluted: 33-41% (GOME-2), <33%[a] (TROPOMI)	
Total uncertainty	Clean: 3-4 PMC (GOME-2), 1.5 PMC (TROPOMI)	
	Polluted: 50-57% (GOME-2) <33% [a] (TROPOMI)	
Dependencies	Not assessed	
Stability	Clear degradation during last years of GOME2A operation. Not recommended to use.	

	GOME-2 OCRA cloud fraction artificial increase after 2018.	

Color code:

x>Threshold	Threshold≥x>Breakthrough	Breakthrough≥x>Goal	Goal≥x

a. Absolute bias at clean conditions not directly provided in [RD34:ROCVR], but the relative bias is smaller than that of GOME-2 (Table 6-2). Likewise, relative dispersion at polluted conditions not directly provided in [RD34:ROCVR], but the absolute dispersion is smaller than that of GOME-2 (Table 6-2).

6.3.2. **Level-3 products**

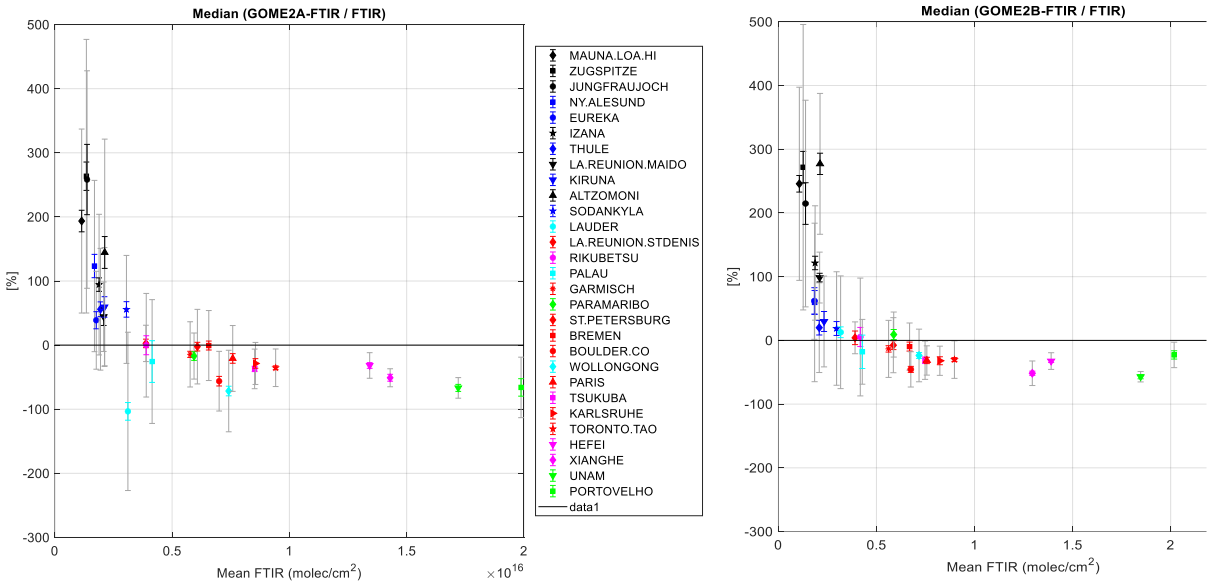
6.3.2.1. **Bias and dispersion**

Bias

Similar results are found with L3 files of all sensors as for L2 (Sect. 6.3.1; [RD38:Vigouroux2020][RD39:DeSmedt2021]): the bias is positive for clean sites and negative for polluted sites. They are summarized in Figure 6-5, Figure 6-6 and Table 6-3. This is also represented by the scatter plots where we see positive intercepts and slopes lower than unity for all sensors. These plots are not shown here but the robust regression slope and intercept, as well as the correlations are provided in Table 6-3.

Dispersion

The 1/2IP68 dispersion values are provided in Table 6-4.



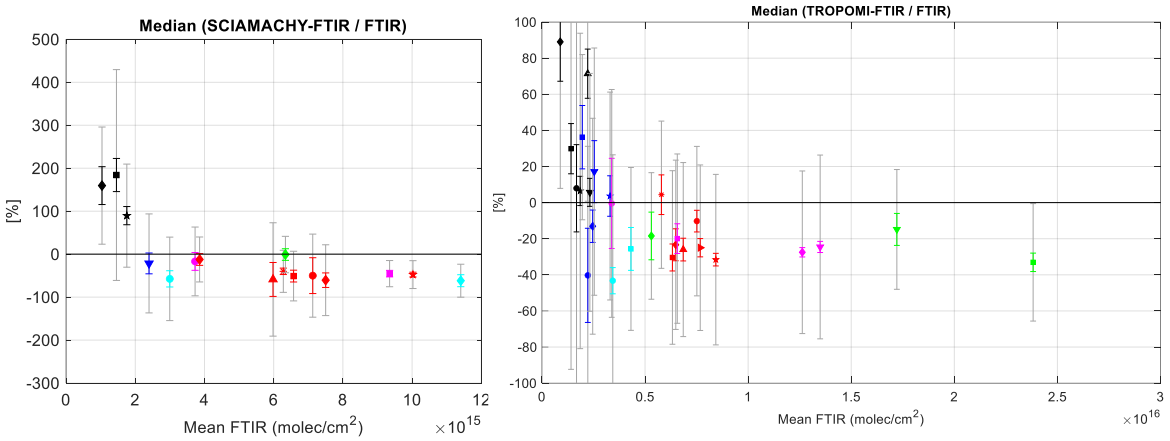


Figure 6-5: Biases for L3 daily cloud-corrected products GOME2 A and B, SCIAMACHY and TROPOMI. Note that for the three former, altitude correction can not be performed, so the high-altitude sites (in black) show higher biases and will not be taken into account into the statistics of Table 6-3.

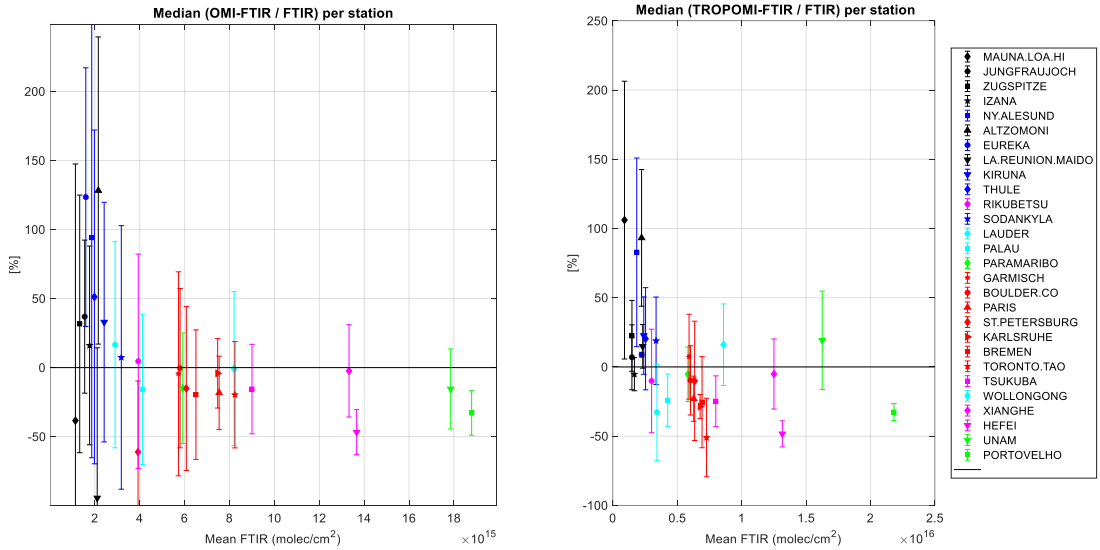


Figure 6-6: Biases for L3 monthly "clear-sky" (i.e., no cloud correction applied) products OMI and TROPOMI.

Table 6-3: Overview statistics of the HCHO L3 files vs FTIR ground-based data. For bias and dispersion, the daily (for SCIAMACHY, GOME2A, B, and TROPOMI) and monthly (for TROPOMI and OMI) pairs of all the stations are separated by pollution level: Clean<2.5E15 molec/cm^2; Polluted: >8E15molec/cm^2). Monthly means are used for the regression statistics. PMC=E15 molec cm^{-2}.

L3 HCHO vs FTIR	Bias (median)	Dispersion (1/2IP68)	Total uncertainty √(med^2+1/2IP68^2)	Regression (TS, Pearson) Monthly means
SCIAMACHY Daily 150km No alt.corr.	Clean: -0.2PMC; -11.3% Polluted: -6.7PMC; -54.3%	Clean: 4.4PMC; 258% Polluted: 5.3PMC; 36%	Clean: 4.4PMC; 258% Polluted: 8.5PMC; 65%	y = 0.41x+1.1PMC R = 0.43
GOME2A Daily 150km	Clean: 1.0PMC; 57% Polluted:	Clean: 4.2PMC; 279%	Clean: 4.4PMC; 285%	y = 0.44x+2.1PMC R = 0.26

No alt.corr.	-5.2PMC; -41.0%	Polluted: 6.1PMC; 43%	Polluted: 8.0PMC; 59%	
GOME2B Daily 150km No alt.corr.	Clean: 0.5PMC; 29% Polluted: -4.9PMC; -37%	Clean: 4.0PMC; 254% Polluted: 5.5PMC; 37%	Clean: 4.0PMC; 255% Polluted: 7.4PMC; 52%	y = 0.57x+1.3PMC R = 0.67
TROPOMI Daily Nearest	Clean: 0.4PMC; +25% Polluted: -3.7PMC; -27%	Clean: 2.1PMC; 151% Polluted: 4.4PMC; 27%	Clean: 2.2PMC; 153% Polluted: 5.8PMC; 38%	y = 0.65x+0.9PMC R = 0.87
TROPOMI monthly; clear products 20km	Clean: 0.5PMC; +31% Polluted: -3.6 PMC; -28%	Clean: 1.0PMC; 69% Polluted: 2.9PMC; 22%	Clean: 1.1PMC; 76% Polluted: 4.6PMC; 36%	y = 0.63x+1.2PMC R = 0.91
OMI Monthly; clear products 20km	Clean: 0.6PMC; +34% Polluted: -2.7PMC; -22%	Clean: 2.1PMC; 134% Polluted: 3.8PMC; 28%	Clean: 2.1PMC; 138% Polluted: 4.6PMC; 35%	y = 0.71x+1.0PMC R = 0.74

6.3.2.2. **Stability**

We apply the MLR model described above (Eq. 1) to the time-series of absolute differences SAT-FTIR at all sites. The trends (drifts) are given in Figure 6-7 in percent by dividing the absolute trends by the mean of FTIR values at each site. Only time-series with more than 70 coincidences are use for drift calculations (so no drift evaluation yet of TROPOMI).
The drifts for all sensors are usually within 50%/decade, and non-significant. Except for:

- There are significant and high negative drifts in the Southern Hemisphere (Lauder and Wollongong), in GOME2 A (which is expected to degrade after 2018), but also in GOME2 B. Looking at the time-series at Lauder (Figure 6-8) and Wollongong (not shown), we observe that the drift is present before the degradation of the noise.
- The drift of GOME2 A seems to increase with latitude, with significant positive drifts observed at 2 mid-latitude sites and all Arctic sites. These positive drifts are not observed in GOME2 B, which shows on the contrary negative drifts at 4 (over 8) mid-latitudes stations. If we allow for less than 70 coincidences to check the Arctic sites as well, the drifts are negative, but not significant (large uncertainties due to low number of coincidences and high variability of the differences; so not shown).
- Drifts in the Southern Hemisphere are better for OMI, although still significant (negative at Lauder, positive at Wollongong). The drifts for OMI are mainly negative, usually below 30%/decade.

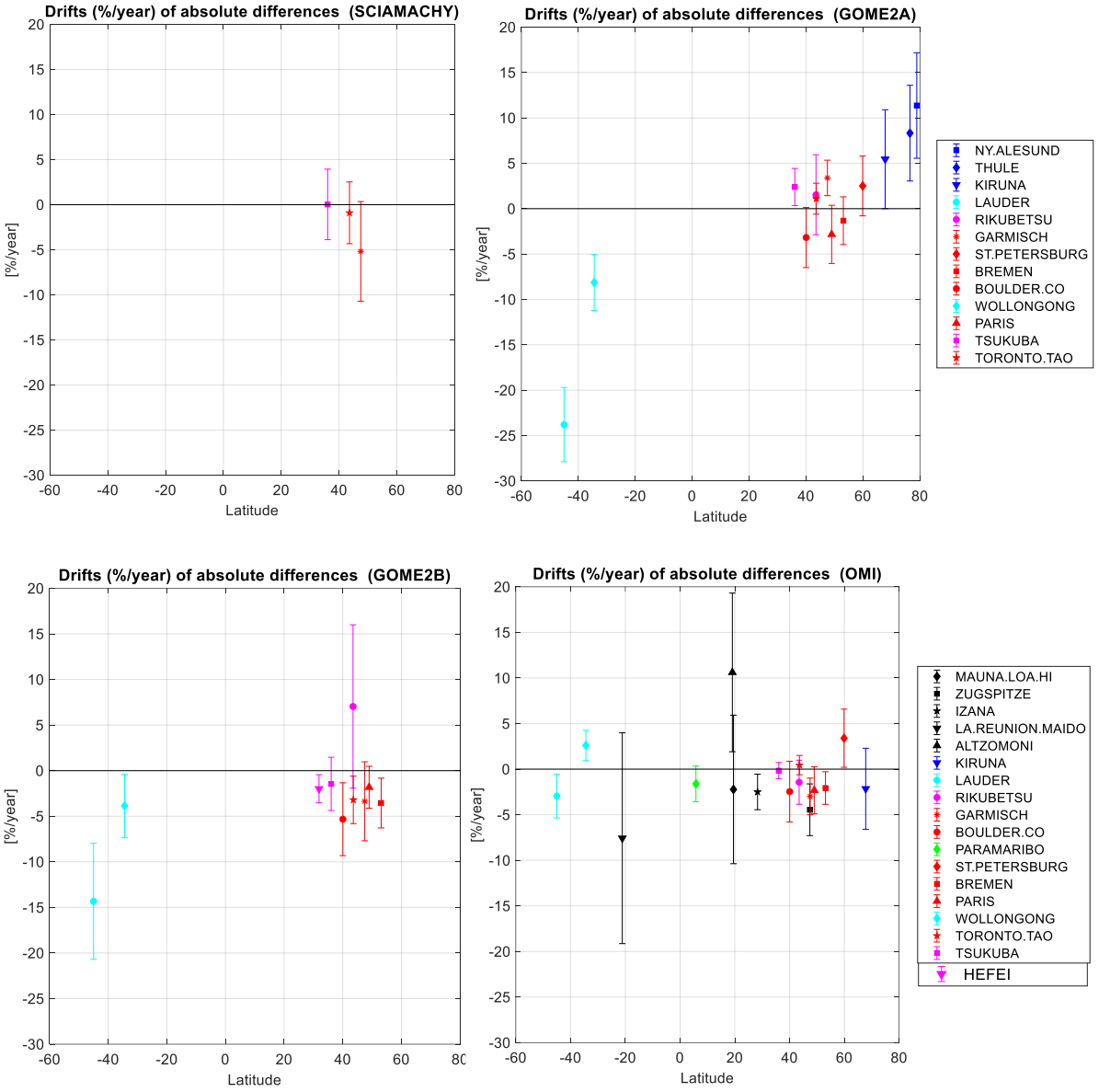


Figure 6-7: Drift (%/year) of the HCHO satellite L3 – FTIR absolute differences (SCIAMACHY, GOME2-A, GOME2-B, OMI). The error bars are equivalent to 2-sigma uncertainty on the drifts. For the three former satellites, altitude correction could not be made, so all high-altitude sites are not used.

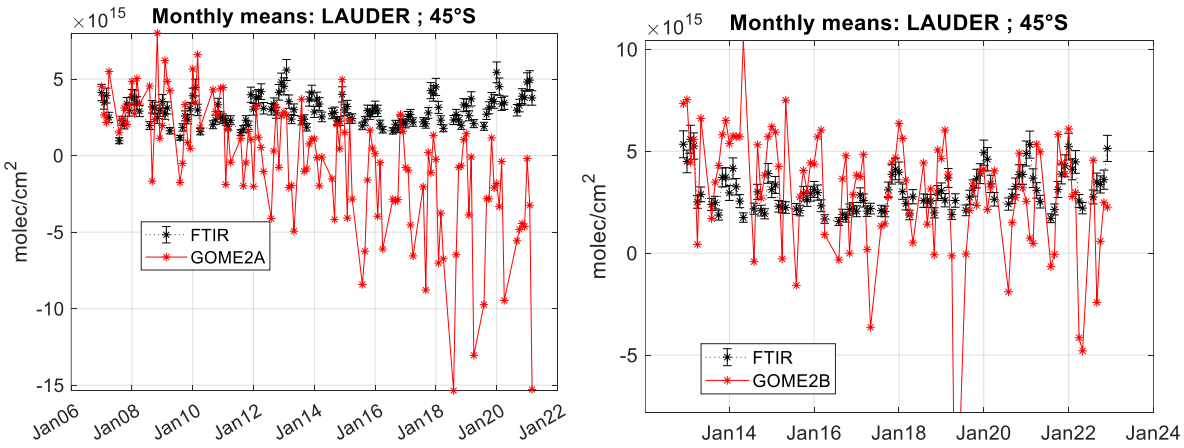


Figure 6-8: Time-series at Lauder, showing the negative drifts of GOME2 A and B compared to FTIR observations.

6.3.2.3. Influence quantities

Not checked here.

6.3.2.4. Impact of L2-to-L3 conversion on validation results

The validation results of L2 and L3 are very consistent as can be seen in the GOME2 A and B Table 6-2 (L2) and Table 6-3 (L3). The validation results of OMI and TROPOMI L3 data are also very consistent with past studies on L2 ([RD40:Mueller2024], [RD38:Vigouroux2020], respectively).

6.3.2.5. Compliance with user requirements

Table 6-4. Compliance with GCOS requirements for HCHO L3 data.

Quantity	Requirements T|B|G	Compliance/evaluation		Remark
Horizontal resolution	100|30|10 km	L3: 0.2°x0.2° grid		The spatial resolution breakthrough requirement is achieved.
Temporal resolution	30days |1day|1h	L3: Monthly	L3: Daily	User requirements T or B are reached.
Bias	Not specified	Check summary in Table 6-3.		Bias is positive for clean sites; negative for polluted sites Bias and dispersion will be used in the next PVIR to validate random and systematic uncertainties.
Dispersion	Not specified	Check summary in Table 6-3		
Total uncertainty (**1-sigma**)	Absolute: 20e15|8e15|4e15 *molec/cm^2	**Clean (< 2.5E15 molec/cm^2)**: Morning sat D: 4-4.4E15 molec.cm^{-2} (>250%) Afternoon sat M: 1.1-2.1E15 molec.cm^{-2} (>76-138%)		The "goal" is reached in absolute values for all conditions, and the "threshold" is reached in % for polluted conditions.

Morning daily; 150km Afternoon monthly; 20km	Relative: 50%|20%|10%	**Polluted (> 8E15 molec/cm^2)** : Morning sat D: 7.4-8.5E15 molec.cm^{-2} (52-65%) Afternoon sat M: 4.6E15 molec.cm-2 (35%)	Afternoon satellites show better precision (but here Morning tested files are daily ones) Next PVIR: monthly files only for morning satellites !
Stability	Absolute: 8e15|8e15|8e15 molec cm^{-2}/decad Relative: 20%|8%|4% /decade	Drifts are usually within 50%/decade (or ±3E15molec.cm^{-2}), but usually non significant (large uncertainties) so when non significant = goal is assumed to be achieved. Significant drifts for: GOME2 A Southern Hemisphere and Artic sites. Some significant drifts (~30%/dec.) at mid-latitudes (positive for GOME2 A; negative for GOME2 B and OMI).	The requirement is always achieved in absolute value, even with strongest drift at Lauder in GOME 2A (-7E15 molec.cm-2). The "threshold" is sometimes not achieved in %/decade.

Color code:

x>Threshold	Threshold≥x>Breakthrough	Breakthrough≥x>Goal	Goal≥x

7. Validation of SO_2 Data Products

7.1. Scope and generalities

Only a limited prototype L2 dataset was available for validation: the SO_2 column retrieved with the COBRA algorithm, for the sensor Aura/OMI, for the year 2012 (see Table 3-3). Two variant products were generated: with prior profile taken from CAMS reanalysis (CAMS-REA) and with prior profile taken from TM5. As CAMS-REA was chosen in the round robin phase to be used for general processing [RD32:PVASR], only results from that variant product are reported here. Column retrievals with and without cloud correction (based on OMI O_2-O_2) are available in the data product. Here, only validation results using the cloud correction are reported; in [RD32:PVASR] both are compared.
As this is a prototype dataset, some variables are missing, like the qa_value normally used for quality flagging (but quality filtering could proceed thanks to specific recommendations of the data provider), and the column uncertainty. No L3 processing is yet available for the SO_2 product. As there is only 1 year of satellite data, and only one station contributing (section 4.6.3), the validation is geographically and temporally limited in scope.

7.2. Validation methodology

Overpasses in HARP-compatible format [RD50:HARP] were generated from the L2 OMI SO_2 data. Two types of comparison were performed: (i) a direct comparison, comparing directly the tropospheric columns of satellite and MAX-DOAS, and (ii) comparison with smoothing, where the averaging kernel (AVK) of the satellite is applied to the MAX-DOAS profile to calculate a smoothed column. The specific steps are provided below.

Satellite pixel filtering. Based on quality filter recommendations from the product provider, pixels were kept only if AMF>0.15, 10 DU>VCD>-3.5 DU, SZA<65, cloud fraction<0.3 and xtrack_flag equal to zero.
Co-location. Only satellite pixels within 50 km of the MAX-DOAS sensor location* are kept. Ground-based data within ±1 h of a satellite pixel are kept, and per satellite pixel, only the best matching ground-based measurement (in distance), is kept.
Vertical harmonization.
1/ Direct comparison: No action
2/ Comparison with smoothing.
The MAX-DOAS profile is vertically regridded with conservation of mass [RD46:Langerock2015] to the satellite pressure grid, and extended above the MAXDOAS range with the satellite prior,**. Then, the MAX-DOAS profile is multiplied with the satellite averaging kernel to obtain a smoothed column*** [RD49:Eskes2003].
Comparison pair handling
All satellite columns co-located with the same MAX-DOAS measurement are averaged. Comparison pairs where less than 5 satellite pixels contribute are rejected, to reduce cases with high random error and low representativeness.

Other tests

*We tested the MAX-DOAS 'probed air mass' location (instead of sensor location), also available in the data file, as well, but agreement in slope and correlation was less good, probably because the estimated distance of the effective air mass from the sensor is overestimated.
**A test with zero extrapolation above the MAX-DOAS vertical range, giving a lower limit on the smoothed column, was performed, and had only 10% impact on bias, and negligible effect on dispersion and slope.
***A substitution of the MAX-DOAS prior by that of the satellite was attempted ([RD47:Rodgers2003], Eq. 10), but was found to worsen the agreement. This is tentatively attributed to a too large difference between both priors, violating the assumption of linearity.

Robust and non-robust validation metrics are calculated as defined in [RD31:PVPv1] using the original and the smoothed MAX-DOAS. The robust validation metrics are used for discussion and to assess the compliance with user requirements.

7.3. Validation results

7.3.1. Level-2 products

7.3.1.1. Bias and dispersion

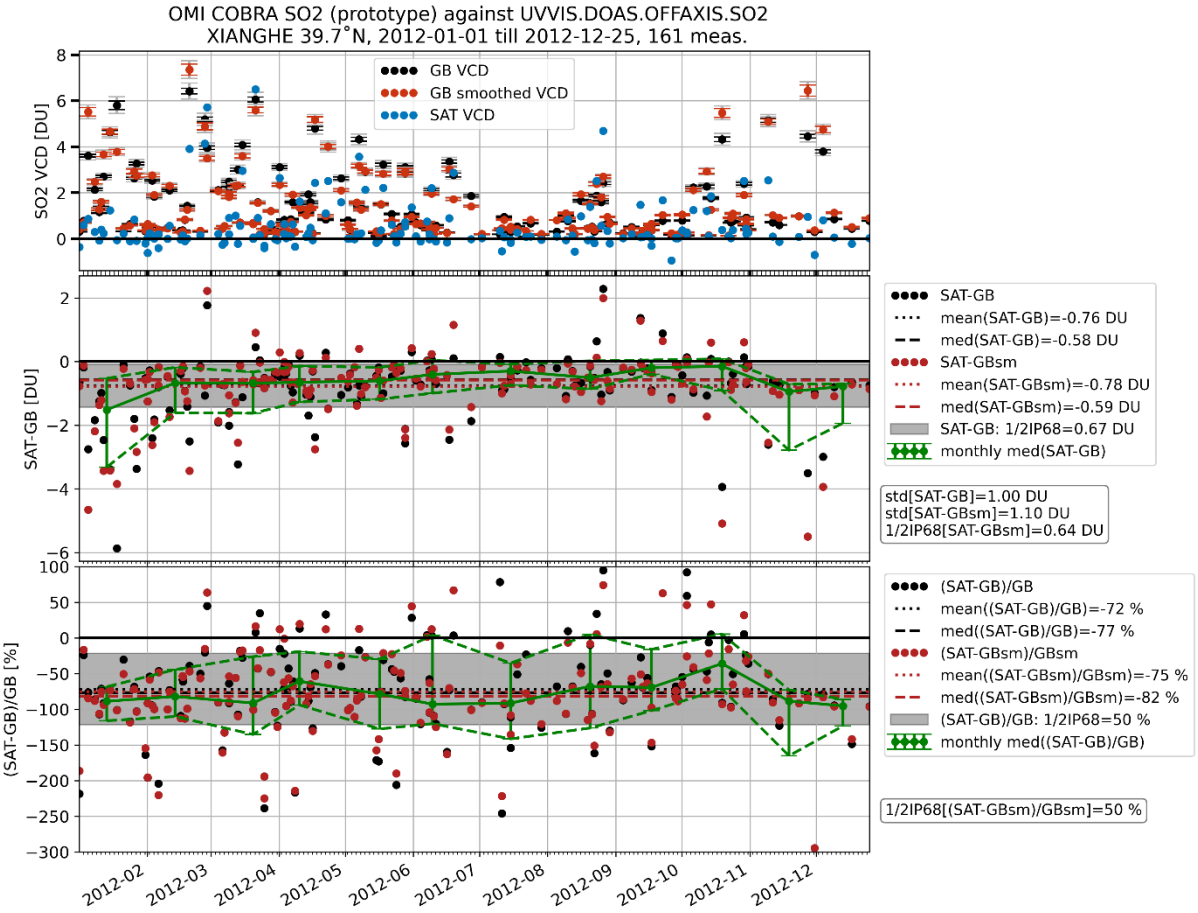
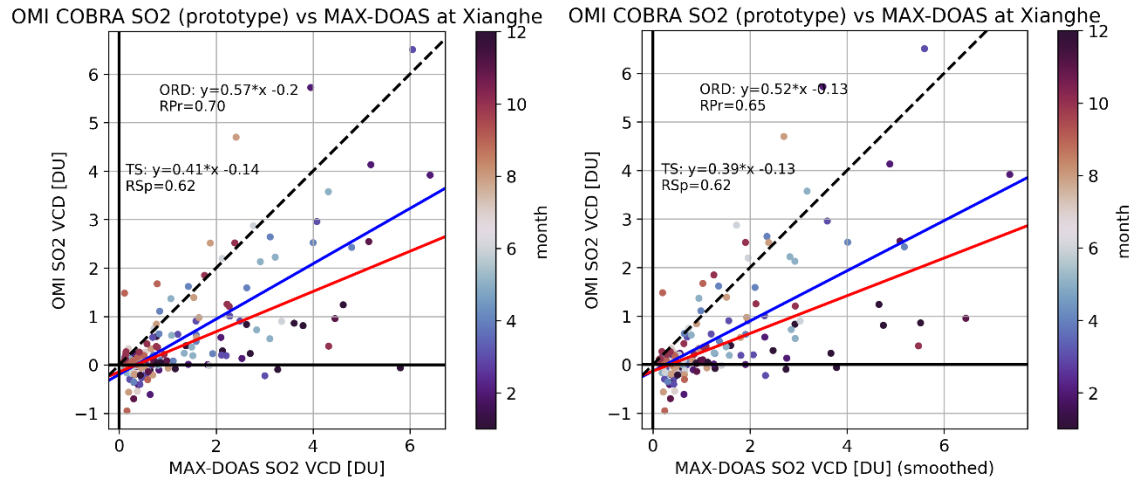


Figure 7-1. Top: time series of OMI COBRA SO_2 (prototype product) and MAX-DOAS at Xianghe (original data and smoothed by the OMI averaging kernel). Middle and bottom: difference and relative difference. Also indicated on the last two plots are mean and median (relative) difference, the [P16,P84] range and monthly medians and monthly [P16,P84] ranges.



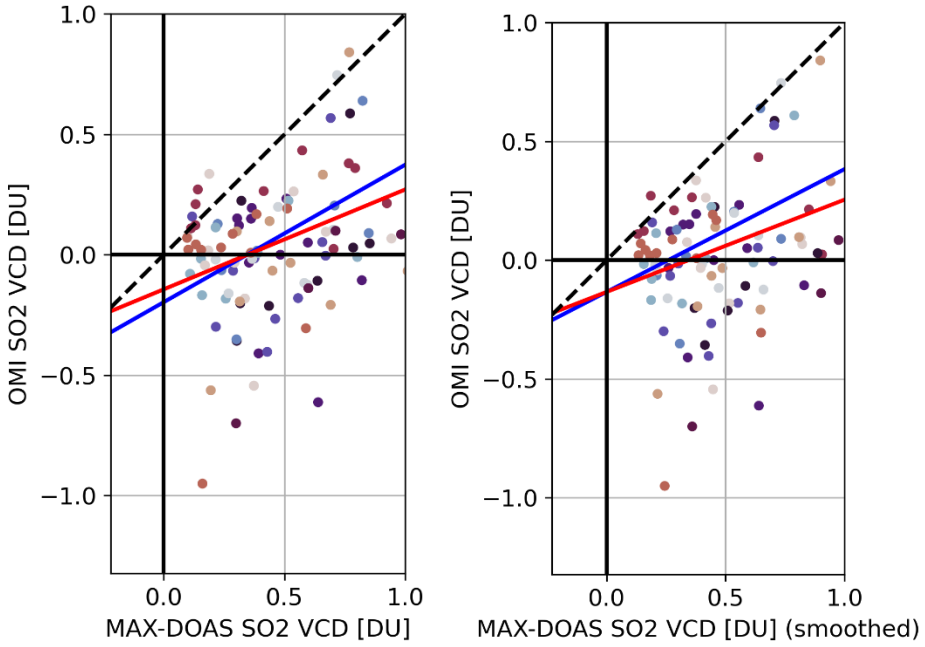


Figure 7-2. Upper left. Correlation plot of OMI COBRA SO_2 (prototype product) vs MAX-DOAS at Xianghe. Upper right. Same but with MAX-DOAS profiles smoothed using the OMI averaging kernel. Indicated are ordinary linear regression and robust Theil-Sen linear regression, and Pearson and Spearman correlation coefficients. Bottom. Zoomed-in panels of the region <1DU.

We discuss here bias and dispersion using robust statistics. Non-robust statistics can be found in Figure 7-1 and Figure 7-2. Note that generally, standard deviation of relative difference is not well behaved e.g., for low reference measurement values, and is therefore not reported here.

Bias. The overall median difference is -0.6 DU and the overall median relative difference is -77%. The monthly median difference varies from -0.15 DU to -1.5 DU, with larger biases when VCD is higher. The monthly median relative difference varies between -60% and -95% in most cases. Theil-Sen regression reveals a slope of 0.41 and an intercept of -0.14 DU.

Dispersion. The overall dispersion (1/2IP68) 0.6 DU and 50% for difference and relative difference respectively. Monthly values range from 0.2 DU to 1.4 DU and from 20% to 65%.

The numbers on bias and dispersion hardly change when the MAX-DOAS profile is smoothed using the OMI averaging kernel.

A significant part of the data is tightly clustered at low SO_2 values, therefore we provide zoomed-in panels in Figure 7-2. Due to scatter in the satellite data values up to -1 DU are reached. The characteristics of this cluster hardly change when smoothing the data.

7.3.1.2. Stability

Not possible to assess, as there is only 1 year of data available.

7.3.1.3. Influence quantities

Difference and relative difference in function of cloud fraction, cloud pressure, solar zenith angle were investigated, without revealing clear dependencies.

7.3.1.4. Summary table of validation results

Table 7-1 presents an overview of validation results. To limit the number of entries, only the robust statistics defined in [RD31:PVPv1] are shown here. Note the very minor difference between unsmoothed and smoothed results.

Table 7-1. Summary of comparison results between OMI COBRA SO_2 (prototype) and MAX-DOAS at Xianghe.

OMI COBRA vs	Bias (median)	Dispersion (1/2IP68)	Total uncertainty $\sqrt{(med^2+1/2IP68^2)}$	T-S regr	Rsp
MAX-DOAS	-0.6 DU,-77%	0.6 DU,50%	0.9 DU,92%	y=0.41*x-0.14	0.62
Smoothed MAX-DOAS	-0.6 DU,-82%	0.6 DU,50%	0.9 DU,96%	y=0.39*x-0.13	0.62

IP68: 68 interpercentile, calculated as P84-P16. T-S regr: Theil-Sen regression. Rsp: Spearman correlation coefficient.

7.3.1.5. Uncertainty budget of the comparison

As there are no prognostic uncertainties provided in this prototype data set, a detailed uncertainty budget is not possible.

We remark that bias, dispersion and their combination are significantly larger than the uncertainty associated with the MAX-DOAS at Xianghe (23%). The fact that smoothing the MAX-DOAS profile using the satellite AVK does not improve the results, suggests that the cause of the discrepancy is not with the choice of the satellite prior (see also [RD32:PVASR] indicating that CAMS is a good prior choice). Finally, validation of another OMI SO_2 data set with Xianghe data [RD42:Theys2015] showed much lower discrepancy after smoothing. This could indicate that the currently found discrepancy is not due to comparison error, but it has to be kept in mind that [RD42:Theys2015] the validation is performed for a wider time range and with monthly averaging.

7.3.1.6. Intercomparison with alternative EO data sets

The validation results provided above are for the cloud-corrected product using CAMS reanalysis for the SO_2 prior profile. This is planned to be used for the full processing. In [RD32:PVASR] validation was also performed for variant products: using TM5 as prior profile, and/or not applying cloud correction. Using TM5 and/or not applying the cloud correction leads to a larger proportional error [RD32:PVASR].

7.3.1.7. Compliance with user requirements

Table 7-2 provides an overview of the compliance with user requirements. OMI horizontal resolution was assessed in [RD33:QA4ECV-PVIR]. The size of the pixels varies from worse than threshold to better than breakthrough, but the median pixel size is close to the breakthrough value. Also, the temporal resolution meets breakthrough requirement.

Bias and total uncertainty exceeds the GCOS threshold requirement on uncertainty, while the dispersion just meets this requirement.

Table 7-2. Compliance with GCOS user requirements (Table 4-3) for L2 OMI COBRA SO_2.

Quantity	Compliance/evaluation	Remark
Horizontal resolution	24x13 to 165x13 km^2	T|B|G : 100|30|10 km Median across track dist=33km [RD33:QA4ECV-PVIR]
Temporal resolution	Daily, <daily at high lat	B: 1 day, G: 1 hour
Bias	-0.6 DU,-77%	T|B|G: 0.37|0.22|0.11 DU, 50%|30%|15% Limited in scope: only 1 site and 1 year of data.
Dispersion	0.6 DU,50%	
Total uncertainty	0.9 DU, 92%	
Dependencies	None found.	
Stability	Not assessed	Only 1 year of data, so not possible to assess.

Color code:

x>Threshold	Threshold≥x>Breakthrough	Breakthrough≥x>Goal	Goal≥x

8. Validation of CO Data Products

8.1. Scope and generalities

L2 IASI products are validated within ACSAF. Only a very brief summary of their validation results is reported here. The validation results of the non-public L3 multi-platform IASI product (see Table 3-4) are reported here.

8.2. Validation methodology

L2 products: see [RD43:VAL_FORLICO].

L3 products:

The CO IASI total columns are provided for day and night separately, but the reference data being solar measurements, we validate only the day variable.

Satellite pixel filtering.
The quality filtering is made by the satellite data provider, and only quality checked L3 CO data are provided in the multi-platform product that we validate. Therefore, no additional filtering is needed.

Co-location.
We have tested 150km collocation criteria which corresponds to about 5-6 pixels collocated to low-mid latitude sites and up to 26 pixels for the high latitude sites. But the better comparisons results (that will be shown in this report) are obtained using the nearest pixel.

Vertical harmonization.
1/ The satellite averaging kernels are not provided in L3 files. So only direct comparisons can be performed (without smoothing).
2/ The altitude difference between the ground-based measurement and the satellite pixel is corrected by scaling the satellite column with a ratio of the IASI prior partial column between the 2 surface altitudes (pixel and station) and the IASI prior total column.

Comparison pair handling
The nearest pixel satellite CO day column co-located is compared to all FTIR CO columns measured within the same month.

Drift estimation methodology: see section 4.4.

8.3. Validation results

8.3.1. Level-2 products

The L2 CO IASI products are validated within ACSAF (EUMETSAT Atmosphere Composition Satellite Application Facility). The validation reports of L2 products are found here: cdop.aeronomie.be [RD43:VAL_FORLICO]. The validation of IASI B and C using FTIR stations shows biases ranging from -3% up to 23% depending on the site, with an average bias using all sites of +7% and +6%, respectively. The dispersion (calculated as the standard deviation of the relative differences) is about 7% for both platforms. The Pearson correlation coefficient is 0.90.

8.3.2. Level-3 product

8.3.2.1. Product specification and file content checks

The intermediate multi-platform (A, B, and C) CO L3 product has a monthly temporal resolution and a 1°x1° spatial resolution. It covers the 2008-2022 period. See Table 3-4.

For each L3 file, CO total columns are provided for the day and for the night. For these 2 variables, associated number of averaged observations and total uncertainties are provided. The mean surface altitude for day and night have been added according to the validation team suggestion, in order to be able to correct for the altitude of the ground-based sites.

However, the IASI CO apriori profile had to be harcoded in order to perform this altitude correction. Providing it in the L3 file might be better for other potential users of the files. Also, no influence quantities are provided in the L3 files.

8.3.2.2. Bias and dispersion

Bias. The overall (all stations together) bias (median of relative difference) is +2.72%. The biases at individual sites vary from -5.2% to +12% (See Fig.1a). The Southern Hemisphere sites show usually higher positive biases than Northern mid-latitudes ones. We do not see a dependence of the bias on the CO mean concentrations. This is confirmed (as well as the small overall bias) by the Theil-Sen regression (see Figure 8-1b) giving a slope of 0.963 and an intercept of only 0.9E17 molec/cm2.
An important point to mention is that, although the median biases at individual sites are usually within 5% (10% for -80S to 20°N latitudes), the individual relative differences are up to -25 / + 25%, and show clear seasonal cycles (see the example in Figure 8-2).

Dispersion. The overall dispersion (1/2IP68) is 1.1E17 molec/cm^2 (or 7.4%). It varies from 0.3E17 molec/cm2 (5.4%) at the cleanest site Arrival Heights up to 4.5E17 molec/cm^2 (15.6%) at the polluted site Xianghe. Note that for an estimation of the precision of the satellite, polluted sites are less relevant since the collocation uncertainty is probably larger there, contributing more to the observed dispersion.

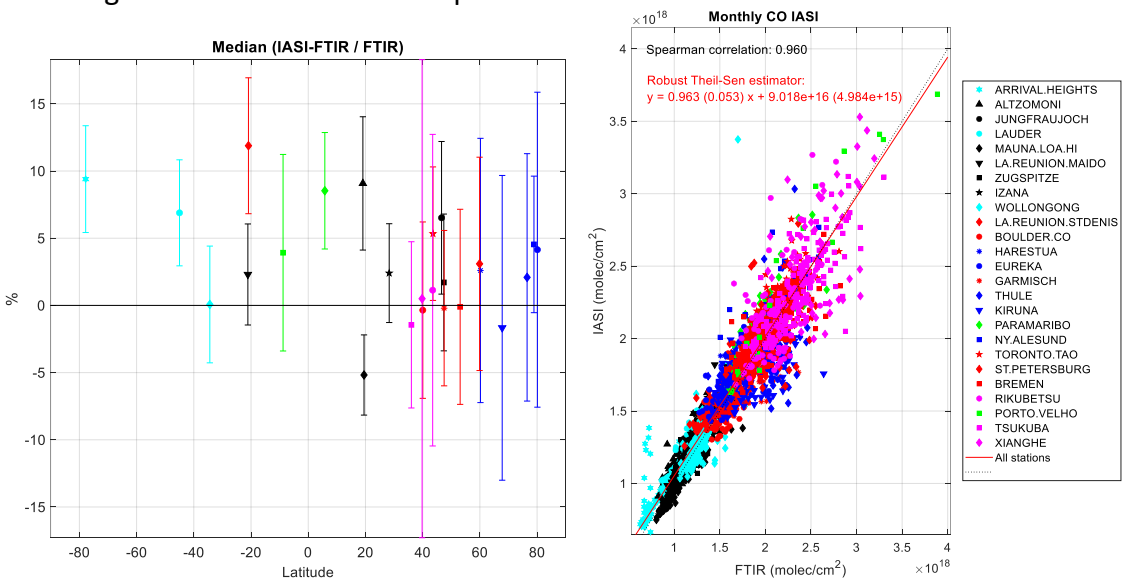


Figure 8-1: a) Left: Individual biases at all sites as a function of latitude; the error bars correspond to a dispersion of 1-sigma. b) Right: the scatter plot of the comparisons.

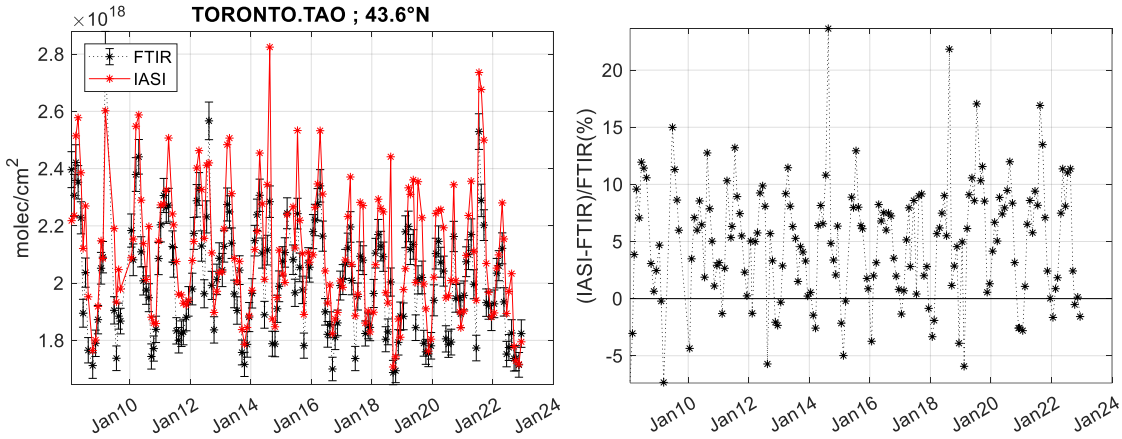


Figure 8-2: Left: CO time-series from IASI L3 multi-platform data set and FTIR at Toronto. Right: their relative differences.

8.3.2.3. **Stability**

We apply the MLR model described above (Eq. 1) to the time-series of absolute differences IASI-FTIR at all sites. The trends (drifts) are given in Figure 8-3 in percent by dividing the absolute trends by the mean of FTIR values at each site. Only time-series with more than 70 coincidences are use for drift calculations.

Usually, IASI CO L3 show no significant drift, except at 5 sites (positive drifts at Lauder, Bremen and 2 Arctic sites (Thule and Ny-Alesund; negative drift at Rikubetsu). Except for these sites, the drifts are within 2%/decade.

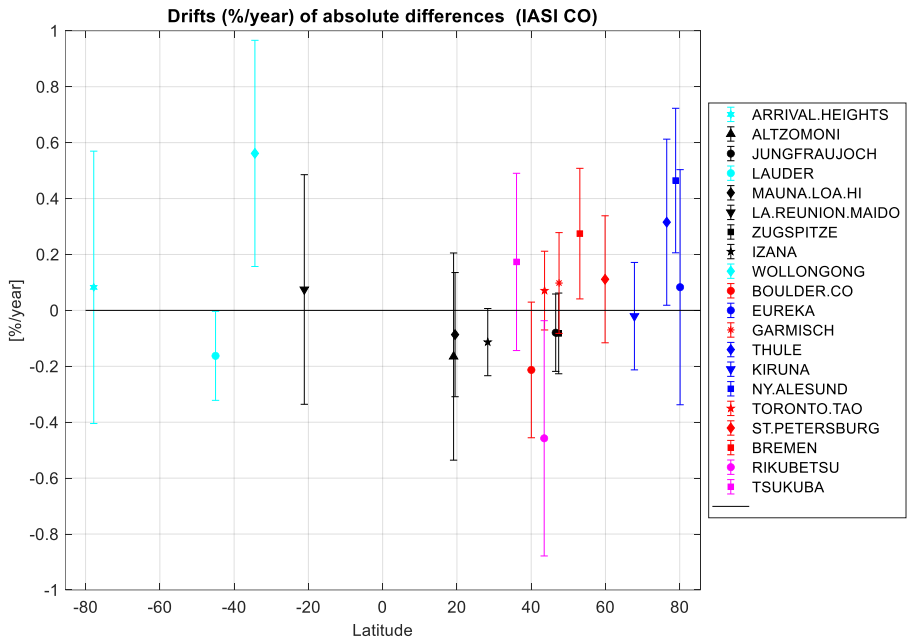


Figure 8-3: Drift (%/year) of the IASI CO L3 – FTIR absolute differences. The error bars are equivalent to 2-sigma uncertainty on the drifts.

8.3.2.4. Influence quantities

As no influence quantities are provided in the L3 files, we cannot study their impact on the validation. The latitude might have an impact on the bias (see Sect. 8.3.3.2) but this needs to be confirmed in the future (e.g. there is only 17 coincidences at the PortoVelho site).

8.3.2.5. Impact of L2-to-L3 conversion on validation results

The overall and individual biases seem improved from L2 (EUMETSAT validation work [RD43:VAL_FORLICO]) to the L3 data provided in CCI+ precursors, but that might be expected due to the improvement of the quality filtering applied here (see [RD5:PUG]).

8.3.2.6. Compliance with user requirements

Table 8-1. Evaluation, compliance with GCOS user requirements (Table 4-4) for L3 IASI CO multi-platform product and recommendations.

Quantity	Requirements T|B|G	Compliance/evaluation	Remark

Horizontal resolution	100|30|10 km	L3: 1°x1° grid ~ 100x100km	Ok: nearest pixel validation gives the best results so **spatial resolution could probably be smaller**
Temporal resolution	30 | 7 | 1 day	L3: Monthly	User "threshold" requirements are reached.
Bias	Not specified for bias and dispersion separately. We compare with total uncertainty requirement	Overall: +2.72%. Individual sites: from -5.2% to +12%	- Large seasonal cycles on the relative differences (can be as large as -25 / + 25%). - Bias and dispersion give an estimate of systematic and random uncertainties separately; but not available in L3; advise to **provide random/systematic uncertainties separately**. - Recommend to provide **prior profiles in datafiles** for altitude correction.
Dispersion		Overall: 1.1E17 molec/cm^2 (7.4%).	
Total uncertainty (**1-sigma**)	5|2.5|0.5 ppb → ~ 5|2.5|0.5 %	Sqrt(median^2+1/2IP68^2)= 1.16E17 molec/cm^2 (7.85%).	Satellite and GB data are obtained in molec.cm^{-2}, therefore the requirements in ppb are not estimated. We roughly estimate their correspondence in % in this Table. Total uncertainty exceeds threshold, mainly due to dispersion (not the bias).
Dependencies		Maybe bias dependency on latitude.	Advise to **provide influence quantities** in L3 files
Stability	2|1|<1 ppb/dec → ~ 2|1|<1 %/dec	Drift are within 6%/dec. for all the sites; and usually within 2%/dec.	Same remark on ppb units as above. The "threshold" is usually achieved for stability.

Color code:

x>Threshold	Threshold≥x>Breakthrough	Breakthrough≥x>Goal	Goal≥x

8.3.3. **Level-3 merged product**

IASI-MOPITT merged product, generated during cycle 2, will be delivered in February 2025 and will be validated in the future v2 of this document.

9. Validation of NH_3 Data Products

9.1. Scope and generalities

Only the L3 NH_3 IASI data is validated (see Table 3-5) here. Validation of the underlying L2 ANNI v4.0.0 product [RD17:Clarisse2023] is out of scope.

9.2. Validation methodology

The NH_3 IASI L3 total columns are provided for am+pm; am-only and pm-only. According to the [RD5:PUG], we validate the **am-only** variable.

Satellite pixel filtering.
The NH_3 L3 files contain the number of observations for each variable, which can be below 1 if only a fraction of the pixel is covered per month. We have tested to filter the data using a threshold on this number: no filtering; filtering number<0.25; <5; and <1.
The best validation results according to bias/dispersion/correlation is obtained with the **<0.5 filtering**. Only these validation results are shown in this report.

Co-location.
We have tested 150km collocation criteria, 50km, and the nearest pixel. Unlike CO, the best results are obtain using 50km and not using the nearest pixel but that might be expected due to the much smaller grid of NH_3 L3 (0.125°x0.125°). Only results using the **50km collocation criteria** are shown in this report. It corresponds to about 50 pixels collocated to low-mid latitude sites and up to 170 pixels for the high latitude sites.

Vertical harmonization.
1/ The satellite averaging kernels are not provided in L3 files. So only direct comparisons can be performed (without smoothing).
2/ The altitude difference between the ground-based measurement and the satellite pixel can not be corrected (no surface pressure/altitude is provided in L3 files). Therefore, we must **limit the validation to non-elevated stations**. Note that even with non -eleveted sites, this is not optimal, the altitude correction has proven to provide better validation results for CO also at low altitude sites (e.g., for sites at sea-level but close to mountains).

Comparison pair handling
We have tested to compare the average of satellite pixels within 50km to all FTIR NH_3 columns measured within the same month. But also to limit the average of FTIR data to +/- 2 hours from the IASI overpass time, according to the short lifetime of NH_3 and its diurnal cycle that can be large. Indeed, the **+/-2 hours collocation time criterium** improves the validation, and is therefore used in the results shown in this report.

Drift estimation methodology: see section 4.4.

9.3. Validation results

9.3.1. Level-2 products

The v4 of IASI NH_3 L2 products [RD17:Clarisse2023] have not been validated yet, and their validation is beyond scope of the current project.

9.3.2. Level-3 products

9.3.2.1. Product specification and file content checks

Product identifier	Source	Time periods	Spatial resolution	Temporal resolution	Algorithm version
NH3_L3_IASI_A/B/C	ULB-LATMOS	2007→ 2023	0.125° x 0.125°	Monthly	ANNI v4.0.0

There are three different L3 products to validate: IASI A (2007-2021), B (2013-2023), and C (2019-2023). They all have a monthly temporal resolution and 0.125°x0.125° spatial resolution.
For each product, three data sets are provided: an averaged column for am+pm observations, and an am-only and a pm-only averaged columns. For these 3 variables, associated number of averaged observations and random and systematic uncertainties are provided. Note that the number of observations can be below 1 if only a fraction of the pixel is covered per month.

9.3.2.2. Bias and dispersion

Bias. The individual and overall (all stations together) biases (median of relative difference) for the three sensors IASI A, B, and C are shown in Fig. 4a for the 9 – out of 12 – stations that contribute to the 3 sensors validation). We observe a similar behavior as for HCHO: a large (in %) positive bias for very clean sites and a negative bias for polluted sites (except PortoVelho but there are only 5 coincidences only at this site).
We therefore give bias for clean and polluted conditions separately, an overall bias for all sites being unrelevant for NH_3 (as for HCHO):

- **Below 2E15 molec/cm^2**: the respective biases for clean conditions are +77%, + 90%, and + 126% for Metop A, B, and C; (about 6 to 7E14 molec.cm^{-2})
- **Above 4E15 molec/cm^2**: the respective biases fot these conditions are -32%, -25%, and -24% for Metop A, B, and C; (about 2.1 to 2.5E15 molec.cm^{-2})

A better way to summarize this is to look at the scatter plots showing positive intercepts and slope smaller to one, for the three sensors (see Fig. 5). It reflects the two different biases of L3 NH_3 IASI: a small positive constant bias (offset) and a negative proportional bias.

There are small differences in the slopes (e.g. Metop C has a better slope than Metop A), but there are much less coincidences with Metop C, especially for polluted sites which are expected to have negative bias. More data would be needed to have a more robust slope/intercept for Metop C.

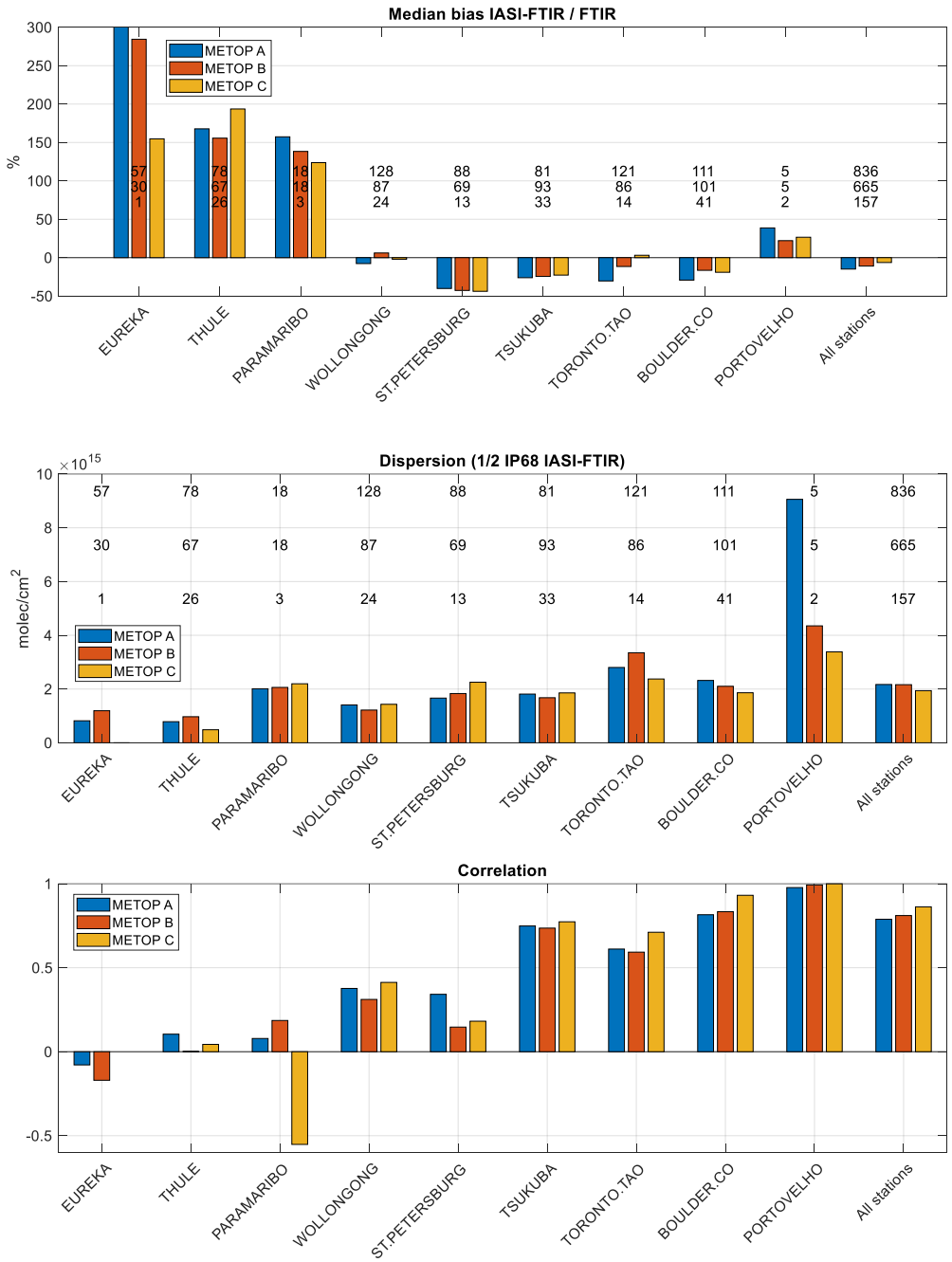


Figure 9-1. For the three IASI sensors: a) The bias (median of the relative differences, in %; b) the dispersion (1/2IP68), in molec/cm^2; c) the correlation between IASI and FTIR. The number of coincidences is provided inside b): from top (Metop A) to bottom (Metop C). Note the smaller number for the more recent IASI C and the lack of recent FTIR data. Note that the bias in a) for Eureka is up to 600% but the axis goes up to 300% for clarity.

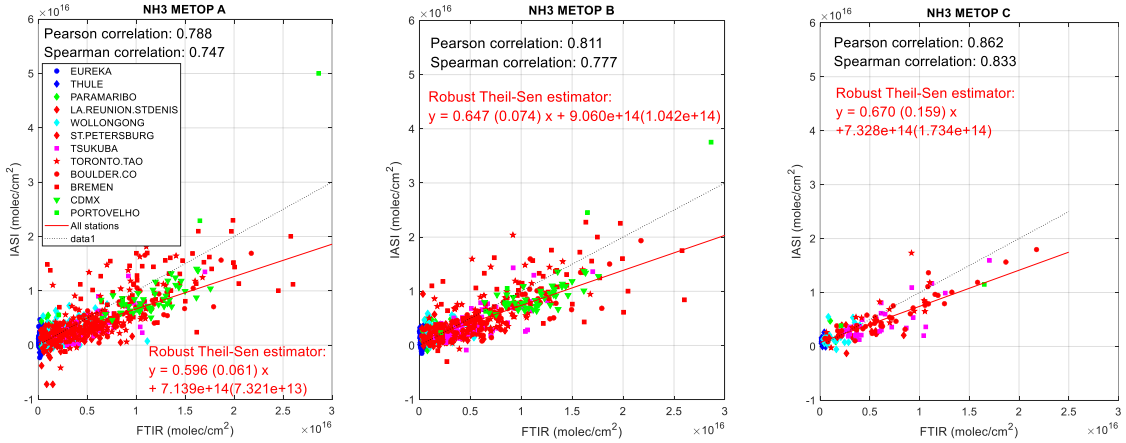


Figure 9-2: Scatter plots between IASI A, B, and C, and the FTIR stations. Correlation coefficients are provided as well as the regression slope and intercept.

Dispersion. The individual and overall dispersion (1/2IP68) are shown in Fig.4b.
The overall dispersion is about 2.2E15 molec.cm^{-2} for Metop A and B, and 1.9E15 molec.cm^{-2} for Metop C. If we separate for pollution conditions, we obtain:

- **Below 2E15 molec/cm^2**: the dispersion is 1.1 and 1.3 E15 molec.cm^{-2} for Metop A and B (270% to 242%) and seems very much improved in Metop C for clean conditions: 0.7 E15 molec.cm^{-2} (193%).
- **Above 4E15 molec/cm^2**: the dispersion is 2.4 E15 molec.cm^{-2} for Metop A and B (30%) and seems improved in Metop C: 2.1 E15 molec.cm^{-2} (27%), but more data would help to confirm this.

9.3.2.3. Stability

We apply the MLR model described above (Eq. 1) to the time-series of absolute differences IASI-FTIR at all sites. The trends (drifts) are given in Figure 6 in percent by dividing the absolute trends by the mean of FTIR values at each site. Only time-series with more than 65 coincidences are use for drift calculations. Only a few sites can be used for this stability validation, due to the poorer state of the art of ground-based FTIR NH_3 network compared to HCHO and CO. The stability of Metop C can not be assessed at present due to the too short time-series.
Usually, IASI NH_3 L3 show no significant drift, except at 2 sites for Metop A (Toronto and Bremen) and at 1 site for Metop B (Toronto). The drifts are usually smaller for Metop B. Except for these sites, the drifts are within 45%/decade and 1.2E10^{15} molec.cm^{-2}/decade. The fact that both sensors show a significant drift at Toronto might point to a drift of the ground-based data, which should be investigated by the FTIR network. The drift at Bremen, only present for Metop A might be due to the sparser coincidences at the beginning of the time-series (not shown), and should be taken with care.

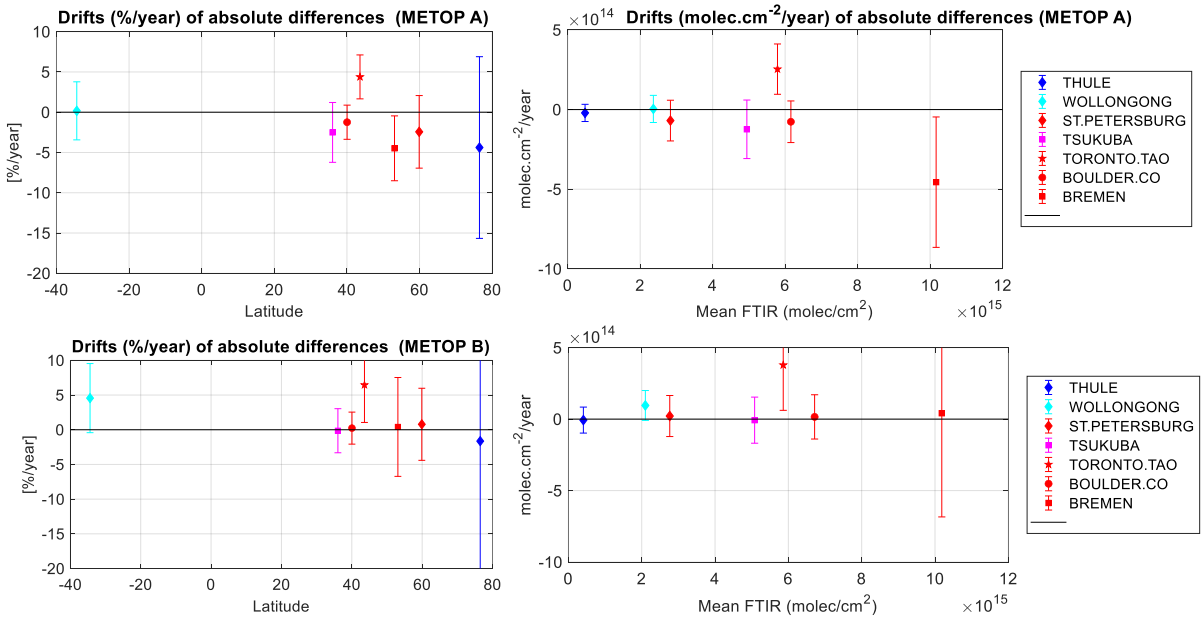


Figure 9-3: Drift (left: in %/year; right in molec/cm2/year) of the NH_3 L3 – FTIR absolute differences (top: Metop A; bottom: Metop B). The error bars are equivalent to 2-sigma uncertainty on the drifts.

9.3.2.4. Influence quantities

As no influence quantities are provided in the L3 files, we can not study their impact on the validation.

9.3.2.5. Impact of L2-to-L3 conversion on validation results

Not applicable yet for NH_3.

9.3.2.6. Compliance with user requirements

Table 9-1. Evaluation, compliance with GCOS user requirements (Table 4-5) and recommendations for L3 IASI NH_3 product.

Quantity	Requirements Threshold| Goal	Compliance/evaluation	Remark
Horizontal resolution	100 km | 10 km	L3: 0.125°x0.125° grid ~ 13km	The spatial resolution requirement is almost achieved. However, the validation results using a single pixel at this resolution are worse even in term of bias. We would advise for a **wider grid**.
Temporal resolution	1 week | 1 hour	L3: Monthly (L2: daily)	Requirement not reached for L3. Users referred to L2 for daily res.
Bias	Not specified	**< 2E15 molec/cm²**: from +77% to +126% (6 to 7E14 molec.cm^{-2}) **> 4E15 molec/cm²**: from -24% to 32% (2.1 to 2.5E15 molec.cm^{-2})	- Bias is positive for clean sites; negative for polluted sites - This can lead to some seasonal cycles on the relative differences at sites having a strong NH_3 seasonal cycle. - Metop C shows a higher precision than A&B, to be confirmed when more coincidences will be available. - Bias and dispersion will be used in the next PVIR to validate random and systematic uncertainites provided in L3 files.
Dispersion	Not specified	**< 2E15 molec/cm²**: 1.3-1.5E15 molec.cm^{-2} for A&B; 0.7E15 molec.cm^{-2} for C (270%,242%,193%). **> 4E15 molec/cm²**: 2.4E15 molec.cm^{-2} for A&B (30%); 2.1E15 molec.cm^{-2} for C (27%).	
Total uncertainty (**1-sigma**)	Absolute: 5 PMC | 1.25 PMC Relative: 50% | 20%	Sqrt(median^2+1/2IP68^2)= **< 2E15 molec/cm²**: 1.3, 1.5, 0.9E15 molec.cm^{-2} for A,B, C, respectively; (281%,258%,231%). Sqrt(median^2+1/2IP68^2)= **> 4E15 molec/cm²**: 3.4, 3.2, 3.1E15 molec.cm^{-2} for A, B, C respectively. (44%, 39%, 36%).	The "goal" is reached in absolute values for clean conditions, "threshold" in % for polluted conditions. However, these validation metrics are obtained using the 50km average. For nearest pixel only "threshold" is reached: **< 2E15 molec/cm²**: 3.1, 3.4E15 molec.cm^{-2} for A,B. **> 4E15 molec/cm²**: 4.4, 4.0E15 molec.cm^{-2} for A, B Advise to **include profiles** for more complete validation
Dependencies		Not assessed as no influence quantities in L3 files	Advise to **include influence quantities** in L3 files
Stability	Absolute: 2 | 1 PMC/decade Relative: 10 | 2 %/decade	Drifts are within 45%/decade and 1.2E10^{15} molec.cm^{-2}/decade	The "goal" is almost achieved in absolute value. However, % values exceed by far the requirement. . The drifts are non significant and with large uncertainties (length of time-series; gaps in the GB data, high variability in the differences,…)

Color code:

x>Threshold	Threshold≥x>Goal	Goal≥x

10. Compliance with Document Requirement Definitions

This section reviews the compliance of this PVIR with the Document Requirement Definitions provided on Page 6 of CCI+ Phase 2 Theme (I) – Precursors ECV SoW – Annex A - [RD1:DRD].

Table 10-1. Mapping between SoW *Document Requirements Definitions* for the Precursors_cci+ PVIR (Annex A, Page 6) and this Version 1 of the *Product Validation and Intercomparison Report*.

SOW ANNEX A REQUIREMENT: THE PVIR SHALL …	LINK TO SECTION	REMARKS
(1) give a full description of the validation and intercomparison strategy, methodologies and tools, citing any community standard validation protocols employed (e.g. CEOS-WGCV, WCRP). Information from the PVP may be reused and updated where appropriate, so that the PVIR provides a self-contained account of all validation and intercomparison work.	General: 4,4.1-4.5 Product-specific: sections 5.2,6.2,7.2,8.2,9.3	
(2) list the data sets (ground-based, radiosonde, in situ, airborne, ship based, etc.) used as independent reference measurements for validation	Section 4.6.	
(3) list the EO and model-based products that are used for intercomparison	CRDP itself: section 3. NO_2 STREAM: 3.1 HCHO variants: 3.2 SO_2 variants: 3.3	
(4) fully describe results of the validation of the CRDP, including the prognostic uncertainty estimates, taking into account the reliability and accuracy of the independent reference measurements.	Sections 5.3,6.3,7.3,8.3,9.3 Uncertainty budget SO_2: 7.3.1.5	Mostly without uncertainty budget. More is planned for PVIR v2.
(5) discuss what relevance the validation applied to one CRDP product level (e.g. Level-2) has for usage of another product level (e.g. Level-3).	1. sect 5.3.1.4 2. sect 6.3.2.4, 8.3.2.5	1. StratNO_2 val on tropNO_2 2. L2 to L3 conversion
(6) validate the CRDP's stability, if it is feasible to do so.	5.3.1.2,6.3.1.2,6.3.2.2,7.3.1.2,8.3.2.3	L2 HCHO+NO2: qualitative. SO_2 : lack of data
(7) compare the CRDP with other commonly used EO and model-based products for the same ECV	Strat NO_2: Table 5-1 HCHO: 6.3.1.4 SO_2: 7.3.1.6	Strat NO_2: STREAM HCHO: product variants SO2: product variants
(8) compare the stability of the CRDP with different EO and model-based products, paying particular attention to breakpoints where different satellites start or stop contributing to the time series.	Not done.	
(9) identify and discuss the impacts of any weaknesses in the validation and intercomparison work (e.g. retrieved quantities that cannot be confidently validated due to lack of reference measurements, or due to large uncertainties or poor geographic sampling of the reference measurements)	SO_2 limited scope: 7.1. No influence quantity in L3: 8.3.2.4,9.3.2.4. Uncertainty in drift estimation: 9.3.2.6.	
(10) quantitatively compare the quality of the CRDP with the GCOS accuracy requirements, and also with the user's accuracy requirements reported in the URD (if different from GCOS).	5.3.1.4,6.3.2.5,7.3.1.7,8.3.2.6,9.3.2.6	
(11) identify limitations in the CRDP, e.g. anomalies, or conditions under which the product quality is lower than expected.	1. 5.3.1.1 2. 6.3.1.2	1. SCIA & GOME-2A QA4ECV high-lat issue 2. GOME2A degradation
(12) provide recommendations to the EO developers	6.3.1.5,8.3.2.6,9.3.2.6	Skip last years GOME2A,

on aspects of CRDP quality that are priorities for future improvement		Optimal grid size, separate random/systematic unc, more ancillary data
(13) provide overall conclusions to users on the CRDP quality, and recommendations for use regarding any limitations identified.	5.3.1.4, 6.3.1.5,6.3.2.5,7.3.1.7,8.3.2.6,9.3.2.6	

11. Terms, abbreviations and definitions

11.1. Terms and definitions

In Table 11-1 terms and definitions as recommended by the CEOS WGCV and by standardization bodies have been reproduced.

Table 11-1. Recommended terms and definitions.

TERM	DEFINITION	SOURCE
accuracy	closeness of agreement between a quantity value obtained by measurement and the true value of the measurand; note that it is not a quantity and it is not given a numerical quantity value	VIM, GUM
area (volume) of representativeness	the area (volume) in which the concentration does not differ from the concentration at the station by more than a specific range	Larssen
bias	(1) systematic error of indication of a measuring system (2) estimate of a systematic measurement error (3) estimate of a systematic forecast error	VIM VIM GAS
calibration	(1) the process of quantitatively defining the system responses to known, controlled signal inputs (2) operation that, under specified conditions, in a first step, establishes a relation between the quantity values with measurement uncertainties provided by measurement standards and corresponding indications with associated measurement uncertainties and, in a second step, uses this information to establish a relation for obtaining a measurement result from an indication	CEOS VIM
dead band (or neutral zone)	maximum interval through which a value of a quantity being measured can be changed in both directions without producing a detectable change in the corresponding indication	VIM
detection limit	measured quantity value, obtained by a given measurement procedure, for which the probability of falsely claiming the absence of a component, given a probability α of falsely claiming its presence	VIM
error	(1) measured quantity value minus a reference quantity value (2) difference of quantity value obtained by measurement and true value of the measurand (3) difference of forecast value and a, estimate of the true value	VIM CEOS
fiducial reference measurement	the suite of independent ground measurements that provide the maximum return on investment for a satellite mission by delivering, to users, the required confidence in data products, in the form of independent validation results and satellite measurement uncertainty estimation, over the entire end-to-end duration of a satellite mission	Donlon and Zibordi (2014)
field-of-regard	an area of the object space scanned by the field-of-view of a scanning sensor	NIST
field-of-view	the solid angle from which the detector receives radiation	NIST

TERM	DEFINITION	SOURCE
footprint	the area of a target encircled by the field-of-view of a detector of radiation, or irradiated by an active system	NIST
influence quantity	quantity that, in a direct measurement, does not affect the quantity that is actually measured, but affects the relation between the indication and the measurement result	VIM
in situ measurement	(1) a direct measurement of the measurand in its original place (2) any sub-orbital measurement of the measurand	GEOSS
Level-0 data	reconstructed, unprocessed instrument and payload data at full resolution (atmospheric and calibration modes, housekeeping data), with any and all measurement and communications artefacts	
Level-1a data	reconstructed, unprocessed data at full resolution, time referenced, and annotated with ancillary information, including radiometric and geometric calibration coefficients and geo-referencing parameters (e.g., ephemeris) computed and appended but not applied to the Level 0 data	
Level-1b data	calibrated, geo-located Earth reflectance and radiance spectra in all spectral bands; solar irradiance data, annotation data and references to used calibration data	
Level-2 data	geophysical measurand at the same resolution and geolocation as the Level 1 source data from which it is derived	
Level-3 data	data or retrieved geophysical parameters (i.e. derived from Level 1 or 2 data products) mapped on uniform space-time grid scales, usually with some completeness and consistency. Such re-sampling may include averaging, compositing, kriging, use of Kalman filters…	
Level-4 data	model output or results from analyses of lower level data, i.e., parameters that are not directly measured by the instruments, but are derived from these measurements	
measurand	quantity intended to be measured	VIM
metadata	data about the data; parameters that describe, characterise, and/or index the data	WMO
monitoring	(1) systematic evaluation over time of some quantity (2) by extension, evaluation over time of the performance of a system, of the occurrence of an event etc.	NIST
point-to-area (point-to-volume) representativeness	the probability that a point measurement lies within a specific range of area-average (volume-average) concentration value	Nappo
precision	closeness of agreement between quantity values obtained by replicate measurements of a quantity on the same or similar object under specified conditions	VIM
process validation	establishing documented evidence of a high degree of assurance that a specific process will consistently produce a product meeting its pre-determined specifications and quality characteristics	CDRH
quality assessment (QA)	QA refers to the overall management of the processes involved in obtaining the data	CEOS
quality control (QC)	QC refers to the activities undertaken to check and optimise accuracy and precision of the data after its collection	CEOS

TERM	DEFINITION	SOURCE
quality indicator (QI)	a means of providing a user of data or derived product with sufficient information to assess its suitability for a particular application. This information should be based on a quantitative assessment of its traceability to an agreed reference or measurement standard (ideally SI), but can be presented as a numeric or a text descriptor, provided the quantitative linkage is defined.	QA4EO
radiometric calibration	a determination of radiometric instrument performance in the spatial, spectral, and temporal domains in a series of measurements, in which its output is related to the true value of the measured radiometric quantity	NIST
random error	component of measurement error that in replicate measurements varies in an unpredictable manner; note that random measurement error equals measurement error minus systematic measurement error	VIM
relative standard uncertainty	standard measurement uncertainty divided by the absolute value of the measured quantity value	VIM
repeatability	measurement precision under set of conditions including the same measurement procedure, same operator, same measuring system, same operating conditions and same location, and replicated measurements over a short period of time	VIM
representativeness	the extent to which a set of measurements taken in a given space-time domain reflect the actual conditions in the same or different space-time domain taken on a scale appropriate for a specific application	Nappo
reproducibility	measurement precision under a set of conditions including different locations, operators, and measuring systems	VIM
resolution	(1) the least angular/linear/temporal/spectral distance between two identical point sources of radiation that can be distinguished according to a given criterion (2) the least vertical/geographical/temporal distance between two identical atmospheric features that can be distinguished in a gridded numerical product or in time series of measurements; resolution is equal to or coarser than vertical/geographical/temporal sampling of the grid or the measurement time series	NIST
stability	ability of a measuring system to maintain its metrological characteristics constant with time	VIM
systematic error	component of measurement error that in replicate measurements remains constant or varies in a predictable manner	VIM
traceability	property of a measurement result relating the result to a stated metrological reference (free definition and not necessarily SI) through an unbroken chain of calibrations of a measuring system or comparisons, each contributing to the stated measurement uncertainty	VIM

TERM	DEFINITION	SOURCE
tropopause	the region of the atmosphere where the environmental temperature lapse rate changes from positive (in the troposphere) to negative (in the stratosphere) the lowest level at which the lapse rate decreases to 2 °C/km or less, provided that the average lapse rate between this level and all higher levels within 2 km does not exceed 2 °C/km occasionally, a second tropopause may be found if the lapse rate above the first tropopause exceeds 3 °C/km	WMO
uncertainty	non-negative parameter that characterizes the dispersion of the quantity values that are being attributed to a measurand, based on the information used	VIM
validation	(1) the process of assessing, by independent means, the quality of the data products derived from the system outputs (2) verification where the specified requirements are adequate for an intended use (3) the process of assessing, by independent means, the degree of correspondence between the value of the radiometric quantity derived from the output signal of a calibrated radiometric device and the actual value of this quantity. (4) confirmation by examination and provision of objective evidence that specifications conform to user needs and intended uses, and that the particular requirements implemented through software can be consistently fulfilled	CEOS VIM NIST CDRH
verification	(1) the provision of objective evidence that a given data product fulfils specified requirements; note that, when applicable, measurement uncertainty should be taken into consideration. (2) the provision of objective evidence that the design outputs of a particular phase of the software development life cycle meet all of the specified requirements for that phase	VIM CDRH

11.2. Abbreviations and acronyms

AD	Applicable Document
AK	Averaging Kernel
AMF	Air Mass Factor, or optical enhancement factor
ATBD	Algorithm Theoretical Basis Document
BIPM	Bureau International des Poids et Mesures
BIRA-IASB	Belgian Institute for Space Aeronomy
CCI	ESA's Climate Change Initiative programme
CDRH	Center for Devices and Radiological Health
CEOS	Committee on Earth Observation Satellites
CMUG	Climate Modelling User Group
CRG	Climate Research Group
DARD	Data Access Requirement Document
DFS	Degree of Freedom of the System

DHF	Data Host Facility
DOAS	Differential Absorption Optical Spectroscopy
DU	Dobson Unit – unit of vertical column density (2.69 10^{16} molec.cm^{-2})
EC	European Commission
Envisat	ESA's Environmental Satellite, launched March 1, 2002
EO	Earth Observation
EOST	Earth Observation Science Team
ERS-2	ESA's Earth Remote Sensing satellite 2, launched April 21, 1995
ESA	European Space Agency
ESRIN	European Space Research Institute
EUMETSAT	European Organisation for the Exploitation of Meteorological Satellites
FTIR	Fourier Transform Infra-Red
GAW	WMO's Global Atmosphere Watch
GCOS	Global Climate Observing System
GEOSS	Global Earth Observation System of Systems
GOME	Global Ozone Monitoring Experiment
GUM	Guide to the expression of uncertainty in a measurement
ISO	International Organization for Standardization
JCGM	Joint Committee for Guides in Metrology
KNMI	Royal Dutch Meteorological Institute
Lidar	Light detection and ranging
MAX-DOAS	Multi Axis Differential Optical Absorption Spectroscopy
MetOp	EUMETSAT's Meteorological Operational satellite
MPC	Mission Performance Centre
NDACC	Network for the Detection of Atmospheric Composition Change
OMI	Ozone Monitoring Instrument
Pandora	not an acronym; direct Sun UV-visible spectrometer
PVP	Product Validation Plan
QA4EO	Quality Assurance framework for Earth Observation
RD	Reference Document
S5P	Sentinel-5 Precursor
SCIAMACHY	SCanning Imaging Absorption spectroMeter for Atmospheric CHartographY
SZA	Solar Zenith Angle
TCCON	Total Carbon Column Observing Network
TROPOMI	TROPOspheric Monitoring Instrument
ULB	Université Libre de Bruxelles
URD	User Requirement Document
UT	Upper Troposphere
VALT	Validation team of the Ozone_cci project
VIM	International Vocabulary of Metrology
WGCV	CEOS Working Group on Calibration and Validation
WMO	World Meteorological Organization
ZSL-DOAS	Zenith-scattered-light Differential Optical Absorption Spectroscopy