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_02_00
Issue	02
Revision	00
Status	Final
Date of issue	23/09/2025
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	19/09/2025
CONTRIBUTING AUTHORS	Validation Team (VALT)	Tijl Verhoelst (BIRA-IASB) Bavo Langerock (BIRA-IASB) Gaia Pinardi (BIRA-IASB) Mahesh Kumar Sha (BIRA-IASB) Kai Krause (IUPB) Andreas Richter (IUPB) Corinne Vigouroux (BIRA-IASB)	07/07 to 15/09/2025
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)	21/08 to 15/09/2025
	ESA	Claire Macintosh	../../2025
REVIEWED AND APPROVED BY	Project Science Leaders	Michel Van Roozendael Folkert Boersma	23/09/2025
	ESA Project Officer	Simon Pinnock	../../2025

DISTRIBUTION	ESA Precursors_cci+ Consortium

rev0.1	2025/09/01-15	Versions circulated to validators and product providers for checking
rev0.2	2025/09/19	Finalized executive summary, small fixes (typos, broken links…)

Table of Contents

EXECUTIVE SUMMARY 15

1. INTRODUCTION 17

1.1. Purpose and scope 17

1.2. Document overview 17

2. APPLICABLE AND REFERENCE DOCUMENTS 18

2.1. Applicable documents 18

2.2. Reference documents 18
2.2.1. Climate data user requirement documents 18
2.2.2. Satellite data products 18
2.2.2.1. Multi target products 18
2.2.2.2. NO_2 data products 18
2.2.2.3. HCHO data products 19
2.2.2.4. SO_2 data products 19
2.2.2.5. CO data products 19
2.2.2.6. NH_3 data product 20
2.2.2.1. CHOCHO data products 20
2.2.3. Ground-based (reference) measurements 20
2.2.3.1. Zenith-scattered-light UV-Visible DOAS (ZSL-DOAS) 20
2.2.3.2. Multi-axis UV-Visible DOAS (MAX-DOAS) 21
2.2.3.3. Fourier Transform Infra-Red spectrometry (FTIR) 22
2.2.4. Geophysical validation 22
2.2.4.1. Validation protocol 22
2.2.4.2. Multi target validation reports 22
2.2.4.3. NO_2 column validation 23
2.2.4.4. HCHO column validation 24
2.2.4.5. SO_2 column validation 25
2.2.4.6. CO column validation 25
2.2.4.7. Methods, models, tools and miscellaneous 25

3. CLIMATE RESEARCH DATA PACKAGE (CRDP) 26

3.1. NO_2 column data sets 26

3.2. HCHO column data sets 28

3.3. SO_2 column data sets 29

3.4. CO column data sets 30

3.5. NH_3 column data sets 31

3.6. CHOCHO column data sets 31

4. ECV VALIDATION METHODOLOGY 32

4.1. General principles of the validation process 32

4.2. Compliance with user requirements 32
4.2.1. NO_2 data product requirements 32
4.2.2. HCHO data product requirements 33
4.2.3. SO_2 data product requirements 33
4.2.4. CO data product requirements 34
4.2.5. NH_3 data product requirements 34
4.2.6. CHOCHO data product requirements 34

4.3. Validation metrics 35

4.4. Drift estimation methodology 36

4.5. Validation of individual components of ECV processing chain 36

4.6. Comparison to independent reference measurements 38
4.6.1. NO_2 column validation data sources 38
4.6.1.1. Zenith scattered light DOAS 38
4.6.1.1. Multi-Axis DOAS 39
4.6.1.1. Direct sun DOAS from Pandora 40
4.6.2. HCHO column validation data sources 41
4.6.3. SO_2 column validation data sources 42
4.6.4. CO column validation data sources 43
4.6.5. NH_3 column validation data sources 44
4.6.6. CHOCHO column validation data sources 45
4.6.7. Error budget of the comparison of atmospheric data 45

5. VALIDATION OF NO_2 DATA PRODUCTS 47

5.1. Scope and generalities 47

5.2. Validation methodology 47

5.3. Validation results 50
5.3.1. Level-2 stratospheric NO_2 column products 50
5.3.1.1. Bias, dispersion and total uncertainty 52
5.3.1.2. Stability 55
5.3.1.3. Influence quantities 55
5.3.1.4. Compliance with user requirements 55
5.3.2. Level-3 NO_2 products 55
5.3.2.1. Product specification and file content checks 56
5.3.2.2. Level-3 stratospheric NO_2 columns 56
5.3.2.3. Level-3 tropospheric NO_2 columns 63
5.3.2.4. Level-3 total NO_2 columns 71
5.3.2.1. Compliance with user requirements 76

6. VALIDATION OF HCHO DATA PRODUCTS 77

6.1. Scope and generalities 77

6.2. Validation methodology 77

6.3. Validation results 79
6.3.1. Level-2 products 79
6.3.2. Level-3 products 79
6.3.2.1. Product specification and file content checks 79
6.3.2.2. Bias, dispersion, and uncertainties 79
6.3.2.3. Stability 87
6.3.2.4. Impact of L2-to-L3 conversion on validation results 88
6.3.2.5. Compliance with user requirements 88

7. VALIDATION OF SO_2 DATA PRODUCTS 90

7.1. Scope and generalities 90

7.2. Validation methodology 90

7.3. Validation results 91
7.3.1. Level-2 products 91
7.3.1.1. Bias and dispersion 91
7.3.2. Level-3 products 94
7.3.2.1. Product specification and file content checks 94
7.3.2.2. Bias and dispersion 94
7.3.2.3. Impact of L2-to-L3 conversion on validation results 96
7.3.2.4. Uncertainty budget of the comparison 96
7.3.2.5. Intercomparison with alternative EO data sets 96
7.3.2.6. Compliance with user requirements 96

8. VALIDATION OF CO DATA PRODUCTS 98

8.1. Scope and generalities 98

8.2. Validation methodology 98

8.3. Validation results 99
8.3.1. Level-2 products 99
8.3.2. Level-3 product (IASI merged) 99
8.3.2.1. Product specification and file content checks 99
8.3.2.2. Bias, dispersion, and uncertainty 99
8.3.2.3. Stability 101
8.3.2.4. Influence quantities 102
8.3.2.5. Impact of L2-to-L3 conversion on validation results 102
8.3.2.6. Compliance with user requirements 102
8.3.3. Level-3 IASI-MOPITT merged product 103
8.3.3.1. Product specification and file content checks 103
8.3.3.2. Bias, dispersion, and uncertainties 104
8.3.3.3. Stability 105
8.3.3.4 Compliance with user requirements 106

9. VALIDATION OF NH_3 DATA PRODUCTS 108

9.1. Scope and generalities 108

9.2. Validation methodology 108

9.3. Validation results 109
9.3.1. Level-2 products 109
9.3.2. Level-3 products 109
9.3.2.1. Product specification and file content checks 109
9.3.2.2. Bias, dispersion, and uncertainty 109
9.3.2.3. Stability 112
9.3.2.4. Influence quantities 113
9.3.2.5. Impact of L2-to-L3 conversion on validation results 113
9.3.2.6. Compliance with user requirements 114

10. VALIDATION OF CHOCHO DATA PRODUCTS 116

10.1. Scope and generalities 116

10.2. Validation methodology 116

10.3. Validation results 117
10.3.1. Level-2 products 117
10.3.2. Level-3 products 117
10.3.2.1. Verification with other TROPOMI products 117
10.3.2.2. Comparison to ground-based MAX-DOAS 118
10.3.2.3. Compliance with user requirements 121

11. COMPLIANCE WITH DOCUMENT REQUIREMENT DEFINITIONS 122

12. TERMS, ABBREVIATIONS AND DEFINITIONS 124

12.1. Terms and definitions 124

12.2. Abbreviations and acronyms 127

List of Figures

Figure 4-1: Geographical distribution of the NDACC ZSL-DOAS, the MAX-DOAS and the PGN@EVDC Pandora spectrometers as used for the validation of stratospheric, tropospheric and total NO_2 column data. Background is a multi-annual mean tropospheric vertical column density (VCD) based on S5P NO_2 observations. 39
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. 39
Figure 4-3: Data availability from the MAX-DOAS VCD-only as used for the validation of tropospheric NO_2 column data. 40
Figure 4-4: Data availability from the MAX-DOAS profiles as used for the validation of tropospheric NO_2 column data, including smoothing. 40
Figure 4-5. Temporal coverage of Pandora direct-sun DOAS NO_2 VCD measurements from PGN (86 stations), obtained via EVDC. 41
Figure 4-6: Geographical distribution of the NDACC FTIR and MAX-DOAS spectrometers as used for the validation of 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. 42
Figure 4-7: The MAX-DOAS spectrometer locations at Xianghe, Wuxi and Mohali, and the PGN Pandora instrument locations at Unam, Vallejo, Altzomoni and Wakkerstroom as considered 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. 43
Figure 4-8. Monthly mean SO_2 VCD per station for (left) PGN Pandora and (right) MAX-DOAS data..... 43
Figure 4-9: 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. 44
Figure 4-10: 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. 44
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. 50
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..... 51
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%. 52
Figure 5-4: Network-wide statistics for the QA4ECV DA (left-hand panel) and QA4ECV STREAM (right-hand panel) products of SCIAMACHY on ENVISAT. 53
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. 54
Figure 5-6: Time series of the Level-3 OMI stratospheric NO_2 columns (at a 1x1 deg^2 grid resolution) and co-located SAOZ sunset observations (adjusted to OMI overpass time) at the Observatoire de Haute Provence (France)..... 57
Figure 5-7: Regression analysis and histogram of the differences corresponding to the co-located Level-3 OMI & SAOZ stratospheric NO_2 time series at OHP visualized in Figure 5-6. 57
Figure 5-8: Time series of the Level-3 GOME-2B stratospheric NO_2 columns (at a 0.2x0.2 deg^2 grid resolution) and co-located SAOZ sunset observations (adjusted to MetOp-B overpass time) at the Observatoire de Haute Provence (France)..... 58
Figure 5-9: Regression analysis and histogram of the differences corresponding to the co-located Level-3 GOME-2B & SAOZ stratospheric NO_2 time series at OHP visualized in Figure 5-8..... 58
Figure 5-10: Box-whisker plots summarizing the comparison bias and dispersion for the GOME on ERS2 and SCIAMACHY on ENVISAT in blue and red respectively. Stations are ordered by latitude. Note that not all sites produced data overlapping the time coverage of both sensors. 59

Figure 5-11: Box-whisker plots summarizing the comparison bias and dispersion for the GOME-2 instruments on MetOp-A, B, and C (blue, red, and green respectively). Stations are ordered by latitude. Note that not all sites produced data overlapping the time coverage of all three sensors..... 59
Figure 5-12: Box-whisker plots summarizing the comparison bias and dispersion for the Aura-OMI and Sentinel-5 Precursor TROPOMI instruments (red and blue respectively). Stations are ordered by latitude. Note that (1) not all sites produced data overlapping the time coverage of both sensors, and (2) quality filtering of the OMI Level-3 data induced more data loss at high Southern latitudes than for TROPOMI. 60
Figure 5-13: Drift estimates, expressed in %/decade, derived from a robust linear regression to the relative differences between satellite (GOME and SCIAMACHY in this graph) and ground-based measurements. Error bars represent 1-sigma uncertainty on the drift estimate. 61
Figure 5-14: Similar to Figure 5-13 but for the three GOME-2 instruments (top panel) and for OMI and TROPOMI (bottom panel). Note that for GOME-2C and TROPOMI, the available time series are actually too short for a stable drift estimate..... 62
Figure 5-15: Illustration of the (a) time-series, (b) scatter plot and (c) histogram of the absolute differences for the OMI CCI+p L3 tropospheric NO_2 VCDs comparisons vs MAX-DOAS in Xianghe. The Theil-Sen regression, mean, median bias and dispersion statistics are given as inset. 64
Figure 5-16: Illustration of the (a) time-series, (b) scatter plot and (c) histogram of the absolute differences for the OMI QA4ECV L2 tropospheric NO_2 VCDs comparisons vs MAX-DOAS in Xianghe. The Theil-Sen regression, mean, median bias and dispersion statistics are given as inset. 65
Figure 5-17: Scatter density plots for the different MAX-DOAS to L3 satellite tropospheric NO_2 VCDs comparisons. The number of sites for each comparison is given in each subplot title, and the Theil-Sen regression, as well as the following statistics: mean and median difference, dispersion (1/2IP68), and Pearson and Spearman correlation, are given as inset. 66
Figure 5-18: Box-whisker plots summarizing the tropospheric NO_2 comparison absolute (left) and relative (right) bias and dispersion for all the satellites instruments against MAX-DOAS VCDs. Different pollution conditions are overlayed over the results for all the points (black): cyan for clean situations (MAX-DOAS data <5 Pmolec/cm^2), blue for moderate pollution (between 5 and 10 Pmolec/cm^2) and red for polluted cases (>10 Pmolec/cm^2). Median biases are summarized in the corresponding colors for the relative differences in the right panel. 67
Figure 5-19: Density plots for the different MAX-DOAS to satellite tropospheric NO_2 VCDs comparisons. The relative (SAT-GB)/GB values are shown as a function of the MAX-DAOS VCD columns. The median and dispersion for bins of 3 Pmolec/cm² are overlayed (in yellow), while in blue and orange the dispersion and total uncertainties are reported. The number of sites for each comparison is given in each subplot title, and the median difference and dispersion statistics are given as inset. The red zone represents the GCOS threshold requirement (Max(50%, 2.5E15 molec cm-2)). 68
Figure 5-20: Time series of Level-3 and Level-2 TROPOMI total NO_2 column measurements (red solid lines and “+” markers respectively) and the corresponding PGN data (in blue), over the highly polluted site of Vallejo in Mexico City. The various quality indicators confirm the

expected changes from Level-2 to Level-3: better correlation, lower dispersion, but also more pronounced underestimation due to increased spatial smearing of localized pollution. 71
Figure 5-21: Time series of Level-3 and Level-2 TROPOMI measurements (red solid lines and "+" markers respectively) and the corresponding PGN data (in blue), over Charles City in Virginia. While the total column time series are remarkably flat, the green and magenta lines, representing the satellite stratospheric and tropospheric columns respectively, reveal that this is a consequence of winter-time tropospheric highs compensating winter-time stratospheric NO_2 lows, and vice-versa in summer. 72
Figure 5-22: Box-whisker plots summarizing the total NO_2 comparison bias and dispersion for the three GOME-2 instruments against PGN. Stations are ordered by increasing long-term mean NO_2 levels (as measured by PGN). Note that not all sites produced data overlapping the time coverage of both sensors. In particular for GOME-2A, only a limited subset of PGN sites already produced measurements. 73
Figure 5-23: Box-whisker plots summarizing the total NO_2 comparison bias and dispersion for the Aura-OMI and Sentinel-5 Precursor TROPOMI instruments (red and blue respectively) against PGN. Stations are ordered by increasing long-term mean NO_2 levels (as measured by PGN). Note that not all sites produced data overlapping the time coverage of both sensors. 74
Figure 5-24: Bias (red), dispersion (magenta), and total uncertainty as a function of mean total NO_2 VCD using PGN data from all sites in a merged way, for all sensors yielding sufficient PGN co-locations to populate the matrix. Cyan and greenish zones represent GCOS threshold and breakthrough requirements respectively. 75
Figure 6-1: Biases for monthly L3 products "clear-sky" (i.e., no cloud correction applied) GOME, SCIAMACHY, GOME2A/B/C, OMI and TROPOMI versus FTIR, after smoothing the FTIR profile with the satellite averaging kernel. The biases for daily L3 OMI product is also shown. We do not show Daily TROPOMI biases, since they are similar to the monthly L3 ones. The grey bars are the dispersion of the differences, and the colored bars are the errors on the median (=2*dispersion/sqrt(nb coincidences)). 81
Figure 6-2. Correlation plots of GOME2A, GOME2C and OMI L3 monthly vs MAX-DOAS, without and with smoothing. Theil-Sen regression and Pearson-R are indicated as well. 81
Figure 6-3: a) and b) Validation Monthly OMI L3 random and total uncertainty, respectively; c) and d) Validation Monthly TROPOMI L3 random and total uncertainty, respectively. 87
Figure 6-4: Drift (%/year) of the monthly L3 products (trend regression including seasonal cycle of the SAT– FTIR absolute differences; divided by mean of FTIR to obtain drift in %) (SAT= GOME, SCIAMACHY, GOME2-A/B/C, OMI, TROPOMI). The error bars are equivalent to 2-sigma uncertainty on the drifts. The larger markers are for the comparisons taking AK into account. Similar drifts are obtained using L3 Daily for OMI and TROPOMI (not shown). 88
Figure 7-1. Ground-based SO_2 vertical column distribution, as collocated with OMI (left), and with S5P (right). Boxplots are ordered along mean SO_2 vertical column. Note that the Wuxi MAX-DOAS has no temporal overlap with S5P, and that Xianghe changes from a strongly polluted site within OMI temporal range, to a much less polluted site in S5P temporal range. 91
Figure 7-2. Boxplots of difference (left) and relative difference (right) for L2 CCI+p OMI (top) and L3 CCI+p (bottom), for cloud-corrected and clear-sky VCD. Green marker is the mean

difference. Boxplots are ordered by mean station SO_2 level (bottom most clean). In red dashed line the total uncertainty requirements (threshold, breakthrough, goal) are indicated. 92
Figure 7-3. Boxplots of difference (left) and relative difference (right) for L2 CCI+p S5P (top) and L3 CCI+p S5P (bottom), for cloud-corrected and clear-sky VCD. Green marker is the mean difference. Boxplots are ordered by mean station SO_2 level (bottom most clean). In red dashed line the total uncertainty requirements (threshold, breakthrough, goal) are indicated. 93
Figure 7-4. Left: L2 OMI SO_2 vs MAX-DOAS and PGN correlation plots. Theil-Sen robust regression is used, and normed Kendall-Tau correlation is indicated. Right: similar but for L2 S5P SO_2. 93
Figure 7-5. Difference and relative difference versus ground-based value for OMI (top) and S5P (bottom), for all PGN Pandora and MAX-DOAS stations combined. Per bin, bias (median difference), dispersion (1/2IP68) and the their combination as total uncertainty QI are indicated, and the GCOS threshold requirement on total uncertainty is indicated as a red zone. The bins are in steps of 0.73 DU; at the lowest bin, the absolute requirement of 0.37 DU holds and in the higher bins the relative requirement of 50% holds. 95
Figure 7-6. Comparison of diagnostic and prognostic total uncertainties. The derived reduced chi^2 value is also shown. 96
Figure 8-1: a) Left: Individual biases at all sites as a function of latitude; The grey bars are the dispersion of the differences (1-sigma), and the colored bars are the errors on the median (=2*dispersion/sqrt(nb coincidences)). b) Right: the scatter plot of the comparisons. 100
Figure 8-2: Left: CO time-series from IASI L3 multi-platform data set and FTIR at Toronto. Right: their relative differences. 101
Figure 8-3: Validation of Monthly IASI L3 total uncertainty. 101
Figure 8-4: a) Drift (%/year) of the IASI CO L3 – FTIR absolute differences. The error bars are equivalent to 2-sigma uncertainty on the drifts. b) Colors: Monthly means of absolute differences at individual stations; black squares and errors bars: Mean of the differences obtained at all stations for each month and standard deviation; Red line: regression model (trend and seasonal cycle, see Sect. 4.4) 102
Figure 8-5: a) Left: Individual biases at all sites as a function of latitude; The grey bars are the dispersion of the differences (1-sigma), and the colored bars are the errors on the median (=2*dispersion/sqrt(nb coincidences)). b) Right: the scatter plot of the comparisons. 104
Figure 8-6: Validation of Monthly IASI-MOPITT L3 total uncertainty. 105
Figure 8-7: a) Drift (%/year) of the IASI-MOPITT CO L3 – FTIR absolute differences. The error bars are equivalent to 2-sigma uncertainty on the drifts. b) Colors: Monthly means of absolute differences at individual stations; black squares and errors bars: Mean of the differences obtained at all stations for each month and standard deviation; Red line: regression model (trend and seasonal cycle, see Sect. 4.4) 106
Figure 9-1. For the three IASI sensors and the stitched product: a) The bias (median of the relative differences, in %; b) the dispersion (1/2IP68), in molec/cm^2; c) the Pearson correlation between IASI and FTIR. The number of coincidences is provided inside a): from top (Metop A) to bottom (Stitched A/B/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 clean sites can be up to 700% but the axis goes up to 130% for clarity. 110
Figure 9-2: Scatter plots between IASI A, B, C, stitched A/B/C and the FTIR stations. Correlation coefficients are provided as well as the regression slope and the intercept. 111

Figure 9-3. a) and b) Validation of Monthly stitched IASI A/B/C L3 random and total uncertainty, respectively. 112
Figure 9-4: Drift (in %/year) of the NH_3 L3 – FTIR absolute differences for IASI A, B, C, and stitched A/B/C; error bars are equivalent to 2-sigma uncertainty on the drifts. 113
Figure 10-1: Comparison of annual means for 2023 from the three glyoxal data products; the GLYRETRO PAL product (top left), the IUP-UB product (top right) and the new precursors-cci product (bottom) 117
Figure 10-2: Time series of TROPOMI glyoxal columns over six selected regions. 118
Figure 10-3: Timeseries of TROPOMI glyoxal columns over ground-based MAX-DOAS stations. 119
Figure 10-4: As Figure 10-3 but as scatter plots with regression lines. 120

List of Tables

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}. 26
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}. 28
Table 3-3. SO_2 column data sets produced within, the CCI+precursors project. In analogy to [RD5:PUG], product ID is given as {species}_{processing level}_{platform}_{sensor}. 30
Table 3-4. CO column data sets produced within the CCI+precursors project. In analogy to [RD5:PUG], product ID is given as {species}_{processing level}_{platform}_{sensor}. 30
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}. 31
Table 3-6. CHOCHO column data sets produced within the CCI+precursors project. In analogy to [RD5:PUG], product ID is given as {species}_{processing level}_{platform}_{sensor}. 31
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. 32
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. 33
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. 33
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. 34
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. 34
Table 4-6. Consortium specified user requirements for CHOCHO tropospheric column, adapted from the URD [RD4:URD]. Specifically, 2-sigma uncertainty is adapted to 1-sigma. 35
Table 4-7. Specific validation metrics to be derived from the comparison to reference data. We focus on robust metrics in this work. 35
Table 4-8. MAX-DOAS and Pandora data used for the SO_2 column validation 42

Table 4-9. Overview of MAX-DOAS stations providing glyoxal column observations used for validation. 45
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. 54
Table 5-2: Summary of the Quality Indicators derived from the comparison of the Level-3 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). Note that all these numbers include ground-based measurement and comparison uncertainties. 60
Table 5-3: Summary of the Quality Indicators derived from the comparison of the Level-3 tropospheric NO_2 VCD data sets to ground-based MAX-DOAS VCD measurements. The bias is computed as the mean or median difference of all the collocation pairs and the dispersion as their 1/2IP68 (i.e., half of the 68% central interpercentile). Rp and Rsp are Pearson and Spearman correlation respectively. Note that all these numbers include ground-based measurement and comparison uncertainties. The number of sites and number of measurement pairs are provided as well. 66
Table 5-4: Bias and dispersion estimates for the different products, differentiated into 3 distinct (MAXDOAS-measured) tropospheric NO_2 VCD bins. For GOME, an insufficient number of MAXDOAS measurements were available. Per NO_2 VCD bin, median difference and relative difference (first line) and 1/2IP68 of difference and relative difference (second line) is provided. 67
Table 5-5: Summary of the Quality Indicators derived from the comparison of the Level-3 tropospheric NO_2 VCD data sets to ground-based MAX-DOAS sites proving vertical profiles, and where smoothing is possible. The bias is computed as the mean or median (absolute or relative) difference over all the collocation pairs and the dispersion as the 1/2IP68 (i.e., half of the 68% interpercentile) of the differences. Note that all these numbers include ground-based measurement and comparison uncertainties. Note that since no valid AVK was present in L3 TROPOMI fv0122, smoothing could not be applied here. 69
Table 5-6: Bias estimates for the different products, differentiated by (PGN-measured) tropospheric NO_2 level into 3 distinct classes. For GOME and SCIAMACHY, an insufficient number of PGN measurements were available. 74
Table 5-7. Compliance with GCOS requirements for the tropospheric NO_2 L3 data. Bias, dispersion and total uncertainty are based on evaluation with MAX-DOAS. Regarding drift, stratospheric NO_2 VCD vs ZSL-DOAS is provided instead. Bias and dispersion are taken from Table 5-4, total uncertainty is then derived as $\sqrt{(bias^2+disp^2)}$. 76
Table 6-1: Overview statistics of the HCHO L3 files vs FTIR ground-based data. For bias and dispersion, the statistics are separated by pollution level: Clean<2.5E15 molec/cm^2; Polluted: >8E15molec/cm^2). PMC=E15 molec cm^{-2}. 81
Table 6-2. Overview statistics of the HCHO L3 files vs MAX-DOAS ground-based data. For bias and dispersion, the statistics are separated by pollution level: Clean<2.5E15 molec/cm2; Polluted: >8E15molec/cm2). PMC=E15 molec cm-2. The second entry per cell is for the

smoothed MAX-DOAS profile. 'Daily per month' means that monthly averages are taken for the daily product comparisons, to make it more intercomparable with the monthly product. 83
Table 6-3. Compliance with GCOS requirements for HCHO L3 data..... 88
Table 7-1. Compliance with GCOS user requirements (Table 4-3) for the L3 COBRA SO_2 products. Bias is calculated as median difference, dispersion as ½ IP68, and total uncertainty as √(bias^2+dispersion^2). 97
Table 8-1. Evaluation, compliance with GCOS user requirements (Table 4-4) for L3 IASI CO multi-platform product and recommendations..... 102
Table 8-2. Evaluation, compliance with GCOS user requirements (Table 4-4) for L3 IASI-MOPITT CO multi-platform product and recommendations. 106
Table 9-1. Evaluation, compliance with GCOS user requirements (Table 4-5) and recommendations for L3 IASI NH_3 product. 114
Table 10-1. Compliance with GCOS requirements for the S5P CHOCHO L3 data..... 121
Table 11-1. 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*. 122
Table 12-1. Recommended terms and definitions. 124

Executive Summary

This document is the *Product Validation and Intercomparison Report* (PVIR) Version 2 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 2 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 final products planned to be released are processed to Level 3. While version 1 of the PVIR had to focus on intermediate products, this version 2 of the report targets products in final release form or close to. Note that for some ECVs not all products were finalized at the time of validation.

Validation of the CCI+p NO_2 products relies on comparisons with ZSL-DOAS instruments of the SAOZ network, MAX-DOAS instruments (partly from NDACC) and direct sun DOAS Pandora instruments from PGN for the stratospheric, tropospheric, and total NO_2 column respectively. There is generally a good consistency in validation results across the different sensors. For the satellite **stratospheric NO_2** column, the comparison dispersion is 0.2 Pmolec/cm^2 for all sensors and dominated by the seasonal amplitude of the difference, while the bias is ~-0.3 Pmolec/cm^2 for the morning sensors and -0.1 Pmolec/cm^2 for the afternoon sensors. The drift is better than 5% for the recent long records by OMI and GOME-2B. The bias of the satellite **total NO_2** column is insignificant or slightly negative in clean conditions but can reach up to -50% in polluted conditions, while the dispersion is usually below 20% in polluted conditions. Finally, for the satellite **tropospheric NO_2** column, in clean conditions (≤5 Pmolec/cm^2) the bias is better than the goal requirement (≤0.5 Pmolec/cm^2) while the dispersion and the total uncertainty are at breakthrough level (≤1 Pmolec/cm^2). In moderately polluted conditions (≤10 Pmolec/cm^2), for most sensors, bias, dispersion, and total uncertainty meet the threshold requirement (≤50%), while in highly polluted conditions (≥10 Pmolec/cm^2) the total uncertainty QI exceeds the threshold requirement for most sensors except TROPOMI.

The CCI+p **HCHO** products were validated by comparing with FTIR and MAX-DOAS observations. The GCOS goal requirement on total uncertainty is met in clean conditions

(≤4E15 molec/cm^2), while it is at breakthrough level (≤8E15 molec/cm^2) in polluted conditions. Drifts, assessed with FTIR observations, were found to be non-significant and meeting the GCOS goal requirement (≤8E15 molec cm^{-2}/decade).

Two **SO_2** products, CCI+p OMI and CCI+p S5P based on the COBRA algorithm, were validated by comparing with MAX-DOAS and PGN Pandora. The bias and dispersion of the L3 S5P product, and the dispersion of the L3 OMI product, are within the GCOS threshold requirement (max(50%,0.37 DU)), while the OMI bias exceeds it at higher SO_2 levels. The total uncertainty QI, which combines bias and dispersion, meets the threshold requirement (≤0.37 DU) at low SO_2 concentrations for both OMI and TROPOMI, but exceeds it (>50%) at higher SO_2 concentration. The L2 validation results are consistent with those of L3. The prognostic total uncertainties from the data product is below the diagnostic ones, but there is an unknown contribution from comparison error.

The L3 **CO** products were validated by comparing with ground-based FTIR. For both the IASI-A/B/C multiplatform (intermediate) product and the merged IASI-MOPITT (final product) the following holds: the total uncertainty QI exceeds the threshold requirement (<5%), mainly due to the dispersion, while the bias is within the threshold level. The drift is usually within threshold level (<2%/decade). The prognostic total uncertainty from the data product agrees with the diagnostic one from validation.

Using ground-based FTIR, the following L3 **NH_3** products were validated: based on the individual sensors IASI-A, -B and -C, and the stitched IASI-A/B/C product. The bias is positive for clean sites and negative for polluted sites. The total uncertainty QI is around 2.6 Pmolec cm^{-2} at lower NH3 levels (<2 Pmolec cm^{-2}) and around 4.5 Pmolec cm^{-2} at higher NH3 levels (>4 Pmolec cm^{-2}) and within the threshold requirement of 5 Pmolec cm^{-2}. The prognostic uncertainties from the data product agree with the diagnostic ones from validation. The drift is assessed to be within threshold level (2 Pmolec cm^{-2}), but more FTIR stations would be needed for a better assessment.

Validation of the S5P **glyoxal** product relies on a small set of MAX-DOAS observations and comparison to existing S5P glyoxal products. The precursors-cci glyoxal product is close to the existing PAL product, but in general lower, bringing it closer to the IUP-UB data set. The scatter is low, and the overall behaviour is as expected. Validation with MAX-DOAS station data shows variable agreement with substantial scatter, but already fulfils GCOS goal requirements (Max(10%, 2E14 molec cm^{-2})) for bias, dispersion and total uncertainty.

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 2, and as opposed to version 1 of the PVIR [RD36:PVIRv1], the glyoxal product is now also in scope and the CDRs are generally in their final stage. Still, for a few ECVs not all products were available yet for validation.

1.2. Document overview

This Product Validation and Intercomparison Report Version 2 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 10 describe validation results and compliance assessment for the NO_2, HCHO, SO_2, CO, NH_3 and CHOCHO 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 11 maps the content of this document with the Document Requirement Definitions outlined in Annex A of the Precursors_cci+ Statement of Work.
- Section 12 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, issue 1, version 1, Precursors_cci+_D4.2_PUG_01_01, 29/05/2024.

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 NO_2 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

[RD10:Glissenaar2025] Glissenaar, I.; Boersma, K. F.; Anglou, I.; Rijsdijk, P.; Verhoelst, T.; Compernolle, S.; Pinardi, G.; Lambert, J.-C.; Van Roozendael, M. & Eskes, H. TROPOMI Level 3 tropospheric NO_2 Dataset with Advanced Uncertainty Analysis from the ESA CCI+ ECV Precursor Project Earth Syst. Sci. Data Discuss., 1-36, https://doi.org/10.5194/essd-2024-616, 2025.

2.2.2.3. HCHO data products

[RD11: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

[RD12: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.

[RD13: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.

[RD14: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

[RD15: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

[RD16: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.

[RD17: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

[RD18: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

2.2.2.1. CHOCHO data products

[RD19:Alvarado2020] Alvarado, L. M. A., Richter, A., Vrekoussis, M., Hilboll, A., Kalisz Hedegaard, A. B., Schneising, O. and Burrows, J. P.: Unexpected long-range transport of glyoxal and formaldehyde observed from the Copernicus Sentinel-5 Precursor satellite during the 2018 Canadian wildfires, Atmos. Chem. Phys., 20(4), 2057–2072, https://doi:10.5194/acp-20-2057-2020, 2020.

2.2.3. Ground-based (reference) measurements

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

[RD20:Solomon1987] Solomon, S.; Schmeltekopf, A. L. & Sanders, R. W. On the interpretation of zenith sky absorption measurements, *J. Geophys. Res.,* 1987*, 92*, 8311-8319

[RD21: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

[RD22: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 NO_2 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.

[RD23: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 O_3 and NO_2 during the NDSC intercomparison of zenith-sky UV-visible spectrometers in June 1996, J. Atmos. Chem., 32, 281–314, 1999.

[RD24: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.

[RD25: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

[RD26: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 NO_2, J. Geophys. Res., 110, D08305, https://doi:10.1029/2004JD005423.

[RD27: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 NO_2 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)

[RD28:Wang2017] Wang, Y., Lampel, J., Xie, P., Beirle, S., Li, A., Wu, D., and Wagner, T.: Ground-based MAX-DOAS observations of tropospheric aerosols, NO2, SO2 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.

[RD29: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.

[RD30: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 SO2 retrieved from MAX-DOAS measurements in Xianghe, China, Atmos. Chem. Phys., 14, 11149–11164, https://doi.org/10.5194/acp-14-11149-2014, 2014.

[RD31: Dimitropoulou2022] Dimitropoulou, E.; Hendrick, F.; Friedrich, M. M.; Tack, F.; Pinardi, G.; Merlaud, A.; Fayt, C.; Hermans, C.; Fierens, F. & Van Roozendael, M., Horizontal distribution of tropospheric NO_2 and aerosols derived by dual-scan multi-wavelength multi-axis differential optical absorption spectroscopy (MAX-DOAS) measurements in Uccle, Belgium, *Atmos. Meas. Tech., 15*, 4503-4529, https://doi.org/10.5194/amt-15-4503-2022, 2022.

2.2.3.3. Fourier Transform Infra-Red spectrometry (FTIR)

[RD32: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.

[RD33: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

[RD34:PVPv2] Precursors CCI+ Product Validation Plan (PVP), deliverable 1.3, issue 2, Precursors_cci+_D1.3_PVP_02_01, 05/08/2024.

2.2.4.2. Multi target validation reports

[RD35:PVASR] Precursors CCI+ Product Validation and Algorithm Selection Report (PVASR), deliverable 2.3, issue 1, Precursors_cci+_D2.3_PVASR_01_01, 29 April 2024.

[RD36:PVIRv1] Precursors CCI+ Product Validation and Intercomparison Report (PVIR), deliverable 4.1, issue 1, Precursors_cci+_D4.1_PVIR_01_01, 26 April 2024.

[RD37: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 NO_2 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.

[RD38:ROCVR] Quarterly Validation Report of the Copernicus Sentinel-5 Precursor Operational Data Products #28: April 2018 – August 2025. Lambert, J.-C., A. Keppens, S. Compernolle, K.-U. Eichmann, M. de Graaf, D. Hubert, B. Langerock, M.K. Sha, E. van der Plas, 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, J. Granville, P. Hedelt, K.-P. Heue, G. Jaross, M.E. Koukouli, E. Loots, R. Lutz, M.C Martinez Velarte, K. Michailidis, A. Pseftogkas, S. Nanda, S. Niemeijer, A. Pazmiño, G. Pinardi, A. Richter, N. Rozemeijer, M. Sneep, D. Stein Zweers, N. Theys, G. Tilstra, C. Topaloglou, O. Torres, P. Valks, J. van Geffen, C. Vigouroux, P. Wang, and M. Weber. S5P MPC Routine Operations Consolidated Validation Report series, Issue #28, Version 28.00.00, 227 pp., 15 September 2025.

2.2.4.3. NO_2 column validation

[RD39: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.

[RD40: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.

[RD41:Pinardi2020] Pinardi, G.; Van Roozendael, M.; Hendrick, F.; Theys, N.; Abuhassan, N.; Bais, A.; Boersma, F.; Cede, A.; Chong, J.; Donner, S.; Drosoglou, T.; Dzhola, A.; Eskes, H.; Frieß, U.; Granville, J.; Herman, J. R.; Holla, R.; Hovila, J.; Irie, H.; Kanaya, Y.; Karagkiozidis, D.; Kouremeti, N.; Lambert, J.-C.; Ma, J.; Peters, E.; Piters, A.; Postylyakov, O.; Richter, A.; Remmers, J.; Takashima, H.; Tiefengraber, M.; Valks, P.; Vlemmix, T.; Wagner, T. & Wittrock, F. Validation of tropospheric NO_2 column measurements of GOME-2A and OMI using MAX-DOAS and direct sun network observations Atmos. Meas. Tech., 13, 6141-6174, https://doi.org/10.5194/amt-13-6141-2020, 2020.

[RD42:ACSAF-VALNO2]. Pinardi, G., T. Verhoelst, J.-C. Lambert, J. Granville, S. Seo and P. Valks, AC SAF Validation Report of Reprocessed L2 NO_2, GOME-2/MetopA&MetopB, doc ref: SAF/AC/IASB/VR/NO2, October 2023. Draft.

[RD43:Chen2009] Chen, D.; Zhou, B.; Beirle, S.; Chen, L. M. & Wagner, T. Tropospheric NO_2 column densities deduced from zenith-sky DOAS measurements in Shanghai, China, and their application to satellite validation Atmos. Chem. Phys., 9, 3641-3662, https://doi.org/10.5194/acp-9-3641-2009, 2009.

[RD44:Ma2013] Ma, J. Z.; Beirle, S.; Jin, J. L.; Shaiganfar, R.; Yan, P. & Wagner, T. Tropospheric NO_2 vertical column densities over Beijing: results of the first three years of ground-based MAX-DOAS measurements (2008-2011) and satellite validation Atmos. Chem. Phys., 13, 1547-1567, https://doi.org/10.5194/acp-13-1547-2013, 2013.

[RD45:Chan2020] Chan, K. L.; Wiegner, M.; van Geffen, J.; De Smedt, I.; Alberti, C.; Cheng, Z.; Ye, S. & Wenig, M. MAX-DOAS measurements of tropospheric NO_2 and HCHO in Munich and the comparison to OMI and TROPOMI satellite observations Atmos. Meas. Tech., 13, https://doi.org/10.5194/amt-13-4499-2020, 2020.

2.2.4.4. HCHO column validation

[RD46:ACSAF-VALHCHO]. 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.

[RD47: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.

[RD48: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.

[RD49: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

[RD50: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

[RD51:Pinardi2025] Pinardi, G.; Friedrich, M. M.; Vigouroux, C.; Langerock, B.; De Smedt, I.; Fayt, C.; Hermans, C.; Beirle, S.; Wagner, T.; Zhou, M.; Wang, T.; Wang, P.; De Mazière,

M. & Van Roozendael, M. Intercomparison of MAX-DOAS, FTIR and direct sun HCHO vertical columns at Xianghe, China, EGUsphere, 2025, 2025, 1-56, https://doi.org/10.5194/egusphere-2025-3320

2.2.4.5. SO_2 column validation

[RD52: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

[RD53:VAL_FORLICO] AC SAF FORLI CO validation report, by B. Langerock et al. Submitted.

2.2.4.7. Methods, models, tools and miscellaneous

[RD54: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.

[RD55: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.

[RD56: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

[RD57:Rodgers2003] Rodgers, C. D. & Connor, B. J. Intercomparison of remote sounding instruments, *J. Geophys. Res.*, **2003**, *108*, 4116

[RD58:Errera2001] Errera, Q., and Fonteyn, D. (2001), Four-dimensional variational chemical assimilation of CRISTA stratospheric measurements, *J. Geophys. Res.*, 106(D11), 12253– 12265, https://doi.org/10.1029/2001JD900010.

[RD59:Eskes2003] Eskes, H. J. & Boersma, K. F. Averaging kernels for DOAS total-column satellite retrievals Atmos. Chem. Phys., 2003, 3, 1285-1291, https://doi.org/10.5194/acp-3-1285-2003

[RD60: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 in [RD36:PVIRv1] is an intermediate prototype product.

The L3 data files considered here for validation obey the following patterns (%Y: year, %m: month, %d: day):

ESACCI-PREC-L3C-NO2_TC_GOME1_ERSD2-DLR-1M-%Y%m%d_%Y%m%d-0360x720-fv0200.nc
ESACCI-PREC-L3C-NO2_TC_SCIAMACHY_ENVISAT-DLR-1M-%Y%m%d_%Y%m%d-0360x720-fv0200.nc
ESACCI-PREC-L3C-NO2-METOPA_GOME2_DLR-0900x01800_1M-%Y%m-fv1.0.nc
ESACCI-PREC-L3C-NO2-METOPB_GOME2_DLR-0900x01800_1M-%Y%m-fv1.0.nc
ESACCI-PREC-L3C-NO2-METOPC_GOME2_DLR-0900x01800_1M-%Y%m-fv1.0.nc
ESACCI-PREC-L3-NO2_TC-OMI_AURA-KNMI-1M-%Y%m%d_%Y%m%d-fv0103.nc
ESACCI-PREC-L3-NO2_TC-TROPOMI_S5P-KNMI-1M-%Y%m%d_%Y%m%d-fv0122.nc

Note that since the validation of OMI fv0103 and TROPOMI fv0122 for this report, more recent versions became available, with for TROPOMI fv0125 a corrected averaging kernel variable (it was NaN-filled in fv0122), and for OMI fv0108 the exclusion of descending node of the orbit from the L3 data, and a different weighting approach for the L3 gridding (both changes make it more consistent with the TROPOMI L3 processing).

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	Section 5.3.1[5]
NO2_L2_ENVISAT_SCIAMACHY	QA4ECV v1.1[1]	QA4ECV/KNMI	2002-2012	Section 5.3.1[5]
NO2_L2_METOP_GOME2_A	AC SAF proto[2]	AC SAF/DLR	2007-2020	Section 5.3.1
NO2_L2_METOP_GOME2_B	AC SAF proto[2]	AC SAF/DLR	2013-2024	N/A
NO2_L2_METOP_GOME2_C	AC SAF proto[2]	AC SAF/DLR	2019-2024	N/A
NO2_L2_AURA_OMI	QA4ECV v1.1/ L1v4[3]	QA4ECV/KNMI	2004-2021	N/A

NO2_L2_S5P_TROPOMI	PAL RPRO v2.3.1[1]	EU/ESA/KNMI	2018-2021	Section 5.3.1
L3 data sets				
NO2_L3_ERS2_GOME1	QA4ECV v1.1, L3: fv0200	QA4ECV/KNMI/ DLR	1995-2003	Section 5.3.2
NO2_L3_ENVISAT_SCIAMACHY	QA4ECV v1.1, L3: fv0200	QA4ECV/KNMI/ DLR	2002-2012	Section 5.3.2
NO2_L3_METOP_GOME2_A	AC SAF proto, L3: fv1.0	AC SAF/DLR	2007-2020	Section 5.3.2
NO2_L3_METOP _GOME2_B	AC SAF proto, L3: fv1.0	AC SAF/DLR	2013-2024	Section 5.3.2
NO2_L3_METOP _GOME2_C	AC SAF proto, L3: fv1.0	AC SAF/DLR	2019-2024	Section 5.3.2
NO2_L3_AURA_OMI	QA4ECV v1.1, L3: fv0103	QA4ECV/KNMI	2004-2021	Section 5.3.2
NO2_L3_S5P_TROPOMI	PAL RPRO v2.3.1, L3: fv0122	EU/ESA/KNMI	2018-2021	Section 5.3.2 & [RD10:Glissenaar2025]

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 [RD36:PVIRv1]. 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 version 1 products consist of the S5P and QA4ECV products [RD11:DeSmedt2018][RD13:PSD_QA4ECVHCHO][RD12:PRF_S5PHCHO], and AC SAF GOME2 product [RD14:PUM_ACSAFHCHO] which are publicly available products and were input to the L3 v1 products, which were prototype products. The validation results of L2 and L3 v1 products were provided in [RD36:PVIRv1]. The version 2 L2 and L3 products are significant upgrades. Validation results are provided in the current document only for L3 v2, not for L2 v2.

The filenames of the L3 version 2 ('fv0200') products targeted here obey the following patterns (%Y: year, %m: month, %d: day, ?: single character)

ESACCI-PREC-L3-HCHO_TC-GOME_ERS2_vfitB_v?_?-BIRA-%Y%m??_%Y%m??-fv0200.nc
ESACCI-PREC-L3-HCHO_TC-SCIAMACHY_ENVISAT_vfitB_v?_?-BIRA-%Y%m??_%Y%m??-fv0200.nc
ESACCI-PREC-L3-HCHO_TC-GOME2_METOP-A_vacsaf_proto_??.?-BIRA-%Y%m??_%Y%m??-fv0200.nc
ESACCI-PREC-L3-HCHO_TC-GOME2_METOP-B_vacsaf_proto_??.?-BIRA-%Y%m??_%Y%m??-fv0200.nc
ESACCI-PREC-L3-HCHO_TC-GOME2_METOP-C_vacsaf_proto_??.?-BIRA-%Y%m??_%Y%m??-fv0200.nc
ESACCI-PREC-L3-HCHO_TC-OMI_AURA_v01.00-BIRA-%Y%m??_%Y%m??-fv0200.nc
ESACCI-PREC-L3-HCHO_TC-OMI_AURA_v01.00-BIRA_1D-%Y%m%d-fv0200.nc
ESACCI-PREC-L3-HCHO_TC-TROPOMI_S5P_v0301-BIRA-%Y%m??_%Y%m??-fv0200.nc
ESACCI-PREC-L3-HCHO_TC-TROPOMI_S5P_v0301-BIRA_1D-%Y%m%d-fv0200.nc

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	v1: QA4ECV v1.2[1] v2[5]	QA4ECV/ BIRA-IASB	1995-2003	N/A
HCHO_L2_ENVISAT_SCIAMACHY	v1: QA4ECV v1.2[1] v2[5]	QA4ECV/ BIRA-IASB	2003-2012[4]	N/A
HCHO_L2_GOME2_A	v1: GDP 4.8/4.9 RPRO[1] v2[5]	AC SAF/DLR	2007-2018	v1: [RD36:PVIRv1], [RD46:ACSAF-VALHCHO] v2: not validated
HCHO_L2_GOME2_B	v1: GDP 4.8/4.9 RPRO[1] v2[5]	AC SAF/DLR	2013-now	v1: [RD36:PVIRv1], [RD46:ACSAF-VALHCHO] v2: not validated
HCHO_L2_GOME2_C	v1: GDP 4.8/4.9 RPRO[1]	AC SAF/DLR	2018-now	v1: [RD36:PVIRv1],

	v2[5]			[RD46:ACSAF-VALHCHO] v2: not validated
HCHO_L2_AURA_OMI	v1: QA4ECV v1.2[1] v2[5]	QA4ECV/ BIRA-IASB	2004-2021	v1: [RD36:PVIRv1], [RD49:Mueller2024] v2: not validated
HCHO_L2_S5P_TROPOMI	v1: Operational[1], collection 3, ≥v2.4 v2[5]	EU/ESA/BIRA-IASB/DLR	2018-now	v1: [RD36:PVIRv1], [RD38:ROCVR] v2: not validated
L3 data sets[2]				
HCHO_L3_ERS2_GOME1 Monthly	v2: cci+p v1.0	BIRA-IASB	1996-2002[3]	6.3.2
HCHO_L3_ENVISAT_SCIAMACHY Monthly	v2: cci+p v1.0	BIRA-IASB	2003-2012[4]	6.3.2
HCHO_L3_METOP_GOME2_A Monthly	v2: AC SAF prototype	BIRA-IASB/DLR/EUMETSAT	2007-2018	6.3.2
HCHO_L3_METOP_GOME2_B Monthly	v2: AC SAF prototype	BIRA-IASB/DLR/EUMETSAT	2013-2020	6.3.2
HCHO_L3_METOP_GOME2_C Monthly	v2: AC SAF prototype	BIRA-IASB/DLR/EUMETSAT	2019-2023	6.3.2
HCHO_L3_AURA_OMI Daily, monthly	v2: cci+p v1.0	BIRA-IASB	2005-2022	6.3.2
HCHO_L3_S5P_TROPOMI Daily, monthly	v2: cci+p v1.0	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 which 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. 5: L2 v1 products were validated in [RD36:PVIRv1]; no separate validation is done for L2 v2 products.

We furthermore note that in [RD35: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 and L3 data products were produced during the CCI+precursors project. The retrieval is based on the COBRA algorithm [RD15:Theys2021]. In the current document, validation results for L2 and L3 daily OMI & TROPOMI are presented. The L3 daily products have a spatial grid of 0.25°x0.25°. The GOME product is not validated because of lack of FRM data within the temporal range. The SCIAMACHY product was not yet available at the time of creating the current document.

The L2 products are currently not scheduled for release.

The filenames of the L3 version 1 ('fv0100') products (planned to be publicly released) targeted here obey the following patterns (%Y: year, %m: month, %d: day).

ESACCI-PREC-L3-SO2_TC-OMI_AURA-BIRA-1D-%Y%m%d-fv0100.nc
ESACCI-PREC-L3-SO2_TC-TROPOMI_S5P-BIRA-1D-%Y%m%d-fv0100.nc

Table 3-3. SO_2 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
L2 data sets				
SO2_L2_ERS2_GOME1	CCI+p v1.0	BIRA-IASB	1996-2003	N/A[1]
SO2_L2_ENVISAT_SCIAMACHY	CCI+p v1.0	BIRA-IASB	2003-2012 (planned)	N/A[2]
SO2_L2_AURA_OMI	CCI+p v1.0	BIRA-IASB	2004-2022	7.3.1
SO2_L2_S5P_TROPOMI	OFFL v0.3_02/ CCI+p v1.0	BIRA-IASB	2018-2023	7.3.1
L3 data sets				
SO2_L3_ERS2_GOME1	CCI+p v1.0	BIRA-IASB	1996-2003	N/A[1]
SO2_L3_ENVISAT_SCIAMACHY	CCI+p v1.0	BIRA-IASB	2003-2012 (planned)	N/A[2]
SO2_L3_AURA_OMI	CCI+p v1.0	BIRA-IASB	2004-2022	7.3.2
SO2_L3_S5P_TROPOMI	OFFL v0.3_02/ CCI+p v1.0	BIRA-IASB	2018-2023	7.3.2

1. No matching reference data. 2. Product not yet available.

We furthermore note that in [RD35:PVASR] comparisons are made for OMI with alternative settings, such as alternative prior profiles.

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 [RD16:ATBD_ACSAFCO]). The L2 CO total columns from IASI A, B and C are filtered with an a posteriori 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 is a combination of IASI and MOPITT v9T [RD17:Deeter2022] data.

Validation results of IASI L2 data (reprocessed by EUMETSAT in 2021) took place within the AC SAF project and are presented in [RD53:VAL_FORLICO]. Only a very brief summary is given in Section 8.3.1.

Table 3-4. CO 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				
CO_L3_IASI_A/B/C[1]	V2	LATMOS-ULB	2008-2022	8.3.2

CO_L3_MRG (IASI-MOPITT)	V2	LATMOS-ULB	2008-2022	8.3.3

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 [RD18: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	2007-2022	9.3.2
NH3_L3_MRG	ANNI v4.0.0	ULB-LATMOS	2007-2022	9.3.2

3.6. CHOCHO column data sets

The CHOCHO L3 data products are produced during the CCI+precursors project. The retrieval is based on the GLYRETRO glyoxal product available through the PAL system with a number of improvements. Only the product from S5P TROPOMI was available for validation. Two variants are available: a daily product at 0.25°x0.25° spatial resolution, and a monthly product at 0.125°x0.125° resolution.

The L3 daily and monthly files considered here obey the following pattern (%Y=year, %m=month, %d=day).

ESACCI-PREC-L3C-CHOCHO-TROPOMI_S5P_BIRA-0720x1440_1D-%Y%m%d-fv01.nc
ESACCI-PREC-L3C-CHOCHO-TROPOMI_S5P_BIRA-1440x2880_1M-%Y%m-fv01.nc

Table 3-6. CHOCHO 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				
CHOCHO_L3_S5P_TROPOMI Daily and monthly	S5P_PAL v1.0.2 / CCI+p v1.0	BIRA-IASB	2018 - 2023	10

4. ECV Validation Methodology

4.1. General principles of the validation process

The Precursors_cci+ Product Validation Plan [RD34:PVPv2] 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 [RD34:PVPv2]. 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 [RD34:PVPv2] 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(25%, 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 5E15 molec cm^{-2}.

4.2.6. **CHOCHO data product requirements**

At the time of writing, no GCOS requirements were formulated for CHOCHO column products, therefore requirements were formulated in [RD4:URD]. CHOCHO 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 4-6. Consortium specified user requirements for CHOCHO 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	Once per 30 days	Hourly
Measurement uncertainty (1-sigma)	Max(50%, 4E14 molec cm^{-2})	Max(10%, 2E14 molec cm^{-2})
Stability	Max(25%, 2e+14 molec. cm^{-2} /decade)	Max(10%, 1e+14 molec. cm^{-2} /decade)

The absolute (relative) threshold requirement on measurement uncertainty is applicable below (above) measurand values of 8E14 molec cm^{-2}. The absolute (relative) goal requirement on measurement uncertainty holds below (above) measurand values of 2E15 molec cm^{-2}.

4.3. Validation metrics

In [RD34:PVPv2] validation metrics were specified for bias, dispersion and stability. See Table 4-7. To keep the number of reporting metrics manageable, we will focus on robust metrics in the compliance tables.

Table 4-7. 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**

When the satellite product includes both random and systematic uncertainty budget, we evaluate, in addition to the total uncertainty as described in Table 4-7, the random uncertainty σ by comparing with the 1/2IP68 dispersion. For this, we calculate the reduced chi-square χ^2_{red} for both random and total uncertainties as, with *i* each individual station and *N* the number of stations:

$$\chi^2_{\mathrm{red,rand}} = 1/N \sum_i^{\mathrm{station}} \frac{[\frac{1}{2}\mathrm{IP68}(\Delta)_i]^2}{\sigma^2_{\mathrm{rand},i}} \qquad \text{Eq(1)}$$

$$\chi^2_{\text{red,tot}} = 1/N \sum_i^{\text{station}} \frac{[\text{med}(\Delta)_i]^2 + [\frac{1}{2}\text{IP68}(\Delta)_i]^2}{\sigma^2_{\text{rand},i}} \quad \text{Eq(2)}$$

where $\sigma_{rand,I}$, $\sigma_{tot,I}$ per station is obtained by averaging the individual prognostic uncertainties.

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-ground-based 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**:

$$Y(t) = a + B * t + C_1 * \cos(2p * t) + C_2 * \sin(2p * t) + C_3 * \cos(4p * t) + C_4 * \sin(4p * t) \quad \text{Eq(3)}$$

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 [RD35:PVASR], from which conclusions are taken over ([RD36:PVIRv1], section 6.3.1.4 for HCHO and this document, section 7.3.2.5 for 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, and by the validation of the total NO_2 column, which is obviously a check on the combined quality of the stratospheric and tropospheric column. See section 5.3.

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 5.3.1 (for NO_2), 6.3.2.4 (for HCHO), 7.3.1 (for SO_2) 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 [RD21:Pommereau1988][RD25: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 [RD40:Compernolle2020]). A key property of these zenith-sky measurements at twilight is the geometrical enhancement of the optical path in the stratosphere [RD20:Solomon1987], which offers high sensitivity to stratospheric absorbers of UV-visible radiation and lower sensitivity to clouds and tropospheric species. For the current validation, we limited ourselves to the SAOZ network (see [RD34:PVPv2] for a more complete overview of ZSL-DOAS instruments).

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).

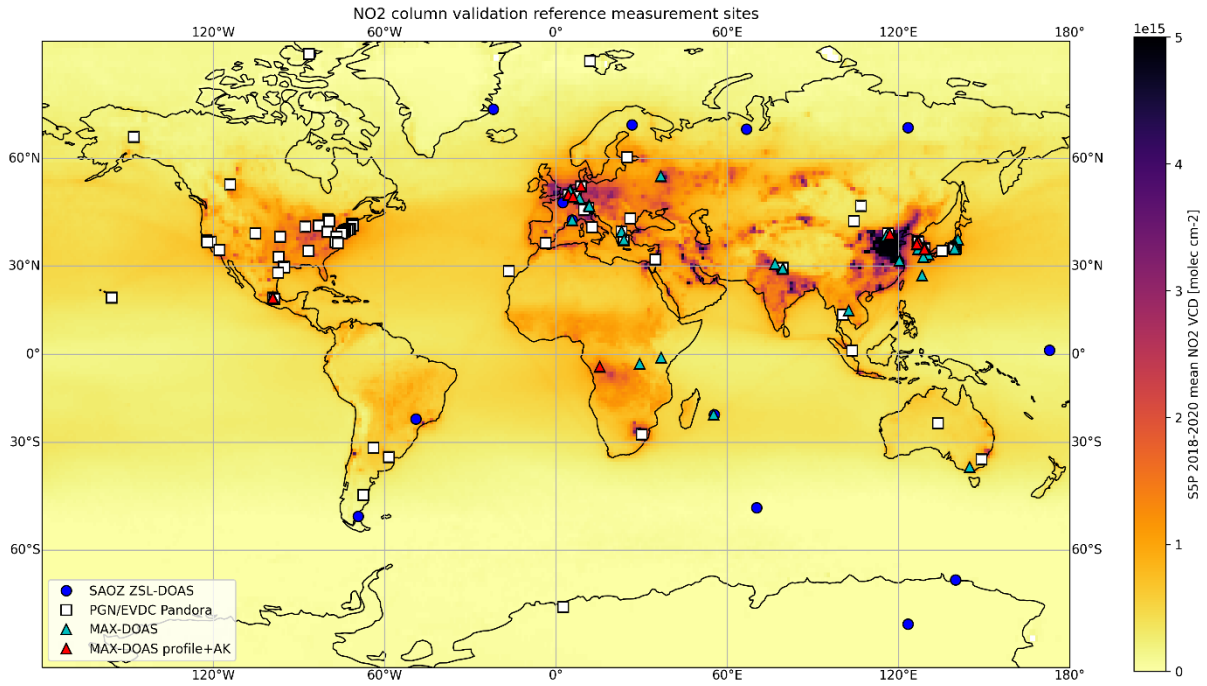


Figure 4-1: Geographical distribution of the NDACC ZSL-DOAS, the MAX-DOAS and the PGN@EVDC Pandora spectrometers as used for the validation of stratospheric, tropospheric and total NO_2 column data. Background is a multi-annual mean tropospheric vertical column density (VCD) based on S5P NO_2 observations.

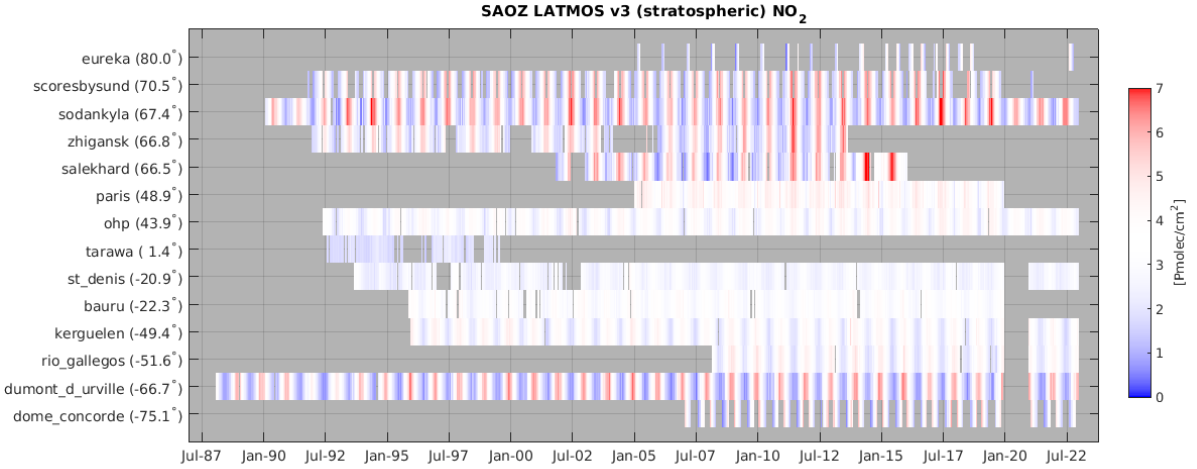


Figure 4-2: Data availability from the SAOZ network of ZSL-DOAS instruments (in consolidated LATMOSv3 processing), as used for the validation of stratospheric NO_2 column data.

NDACC field intercomparison campaigns [RD23:Roscoe1999][RD26: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 [RD27:Yela2017][RD22: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 [RD34:PVPv2] for a discussion on the ZSL-DOAS data.

4.6.1.1. Multi-Axis DOAS

MAX-DOAS are more and more used for the validation of tropospheric NO2 and HCHO [Pinardi2020][Compernolle2020][Verhoelst2021]. Within the project a large effort has been put in collecting MAX-DOAS data from past publications and harmonizing them into a common harp-compatible format. This effort led to the availability of about 70 datasets for the validation of tropospheric NO_2 VCD (only providing columns) and about 23 sites also providing profiles and averaging kernels, that allow a more in-depth study of the a priori impact in the satellite VCD, by smoothing the MAX-DOAS profiles with the satellites averaging kernels. The 2 ensembles of data are shown in Figure 4-3 and Figure 4-4, while the geographical distribution is shown in Figure 4-1. Some of the data only cover a few months (campaign data), while some others are more permanent sites measuring for over a decade (eg. Xianghe or Yokosuka). For some sites, several versions for different retrievals or separation in time due to different formats of the original data exist and are all considered here.

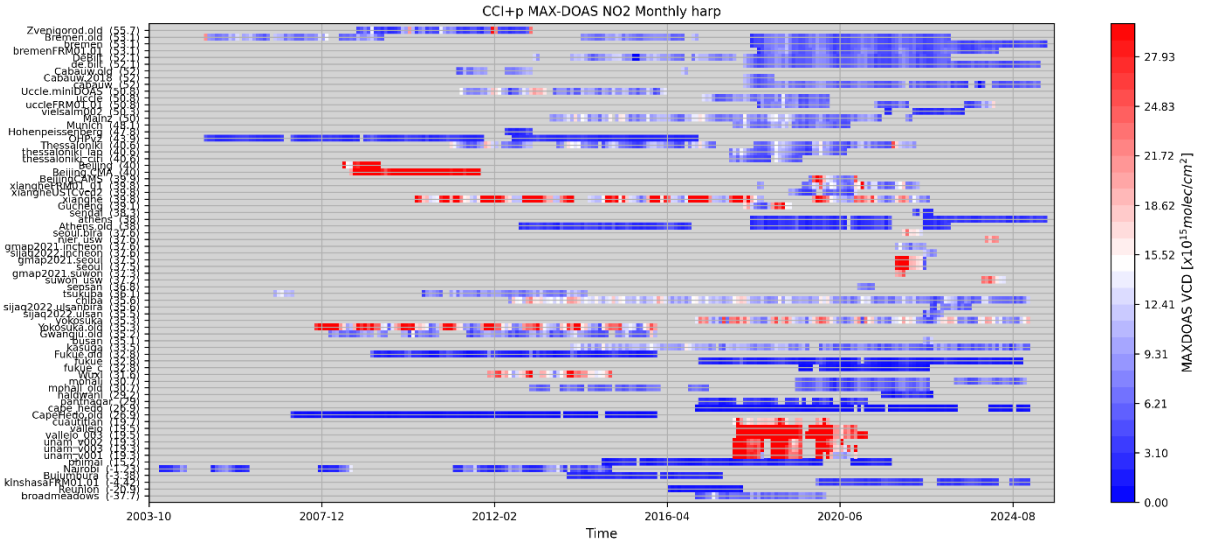


Figure 4-3: Data availability from the MAX-DOAS VCD-only as used for the validation of tropospheric NO_2 column data.

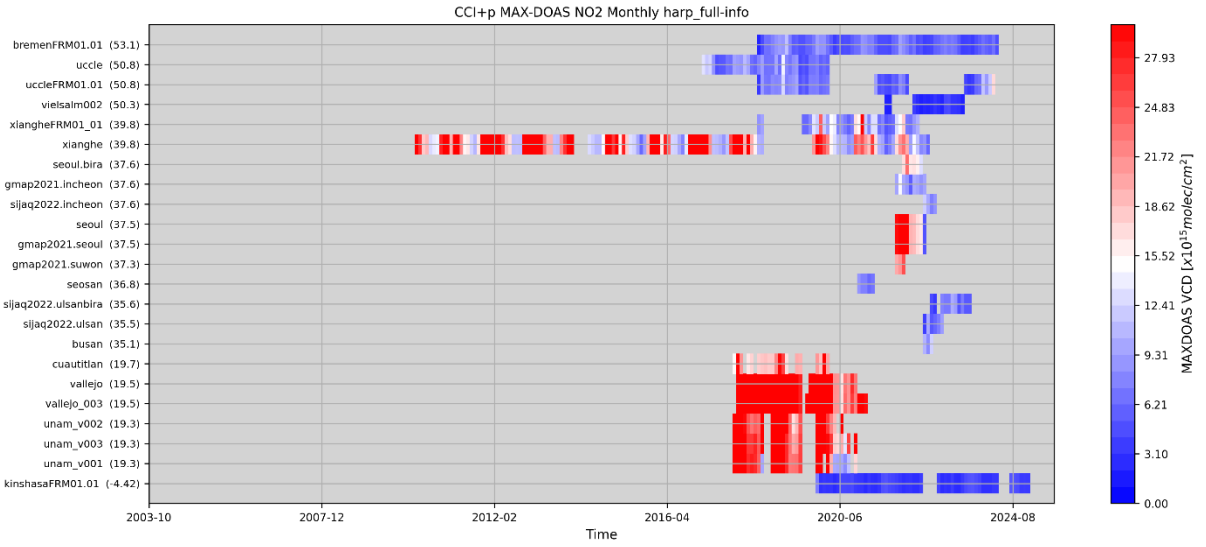


Figure 4-4: Data availability from the MAX-DOAS profiles as used for the validation of tropospheric NO_2 column data, including smoothing.

4.6.1.1. Direct sun DOAS from Pandora

The Pandora instrument infers NO_2, HCHO, SO_2 and ozone total column data from direct Sun UV-visible observations. The Pandonia Global Network (PGN) of standardized Pandora instruments acquires data in an automatic way across the globe, at regular frequency throughout the day. We only consider data uploaded to EVDC, as these are more thoroughly quality checked. 86 PGN/Pandora instruments contribute to the current NO_2 validation (see Figure 4-1). Their temporal coverage does not extend before 2016, with the exception of Innsbruck and Boulder which started in 2015 and 2014 respectively. They have been used in

the validation of e.g., TROPOMI operational NO_2 product [RD39:Verhoelst2021]. Temporal coverage is provided in Figure 4-5.

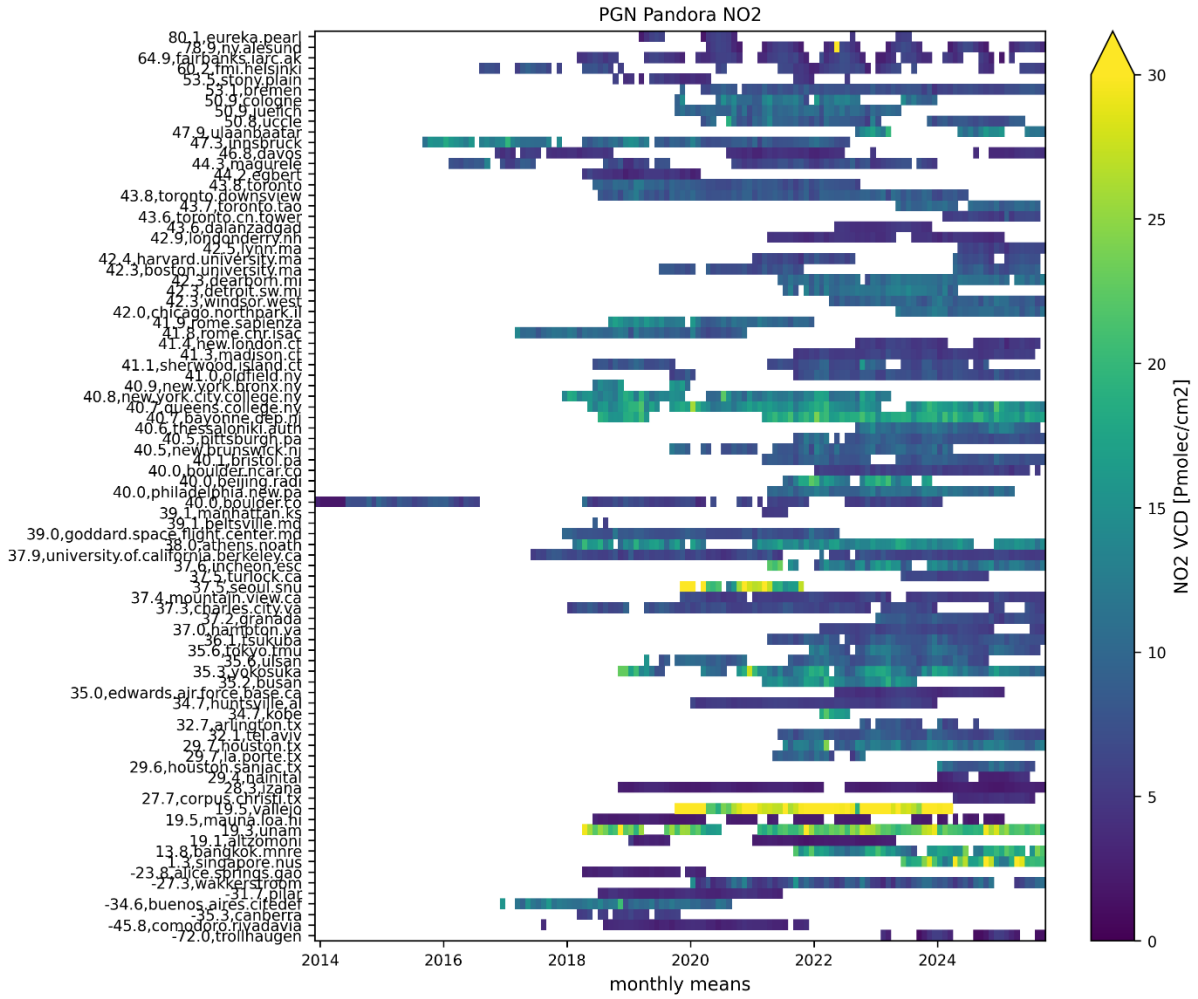


Figure 4-5. Temporal coverage of Pandora direct-sun DOAS NO_2 VCD measurements from PGN (86 stations), obtained via EVDC.

4.6.2. **HCHO column validation data sources**

29 FTIR stations, and 20 MAX-DOAS stations were used for the L3 validation. Only MAX-DOAS processings that include profile retrieval and an averaging kernel were considered, this is therefore a subset of those presented in [RD34:PVPv2]. The MAX-DOAS data are generated following different processings. Regarding Xianghe and Uccle: the bePRO processing has been succeeded by the FRM4DOAS processing, but the latter was not applied to the oldest measurements. Therefore, the bePRO processing was used for the satellite sensors with older historical data (OMI, SCIAMACHY, GOME2A, GOME2B) and the FRM4DOAS processing for the satellite sensors with only recent data (TROPOMI, GOME2C). Their geographical distribution is visualized in Figure 4-6.

More detail about the FTIR network can be found in [RD32:Vigouroux2018]. See also [RD34:PVPv2] for a discussion on the MAX-DOAS and FTIR data. In [RD51:Pinardi2025] a detailed comparison of MAX-DOAS, FTIR and direct sun measurements at Xianghe is carried

out, establishing an excellent comparison between FTIR and direct sun, but differences between FTIR and MAX-DOAS related to prior choice and the different vertical sensitivity.
We note finally that the FTIR Altzomoni station, and the UNAM MAX-DOAS Vallejo, Unam/Mexico City and Cuautitlan stations were not used, as its data strongly deviates from the rest.

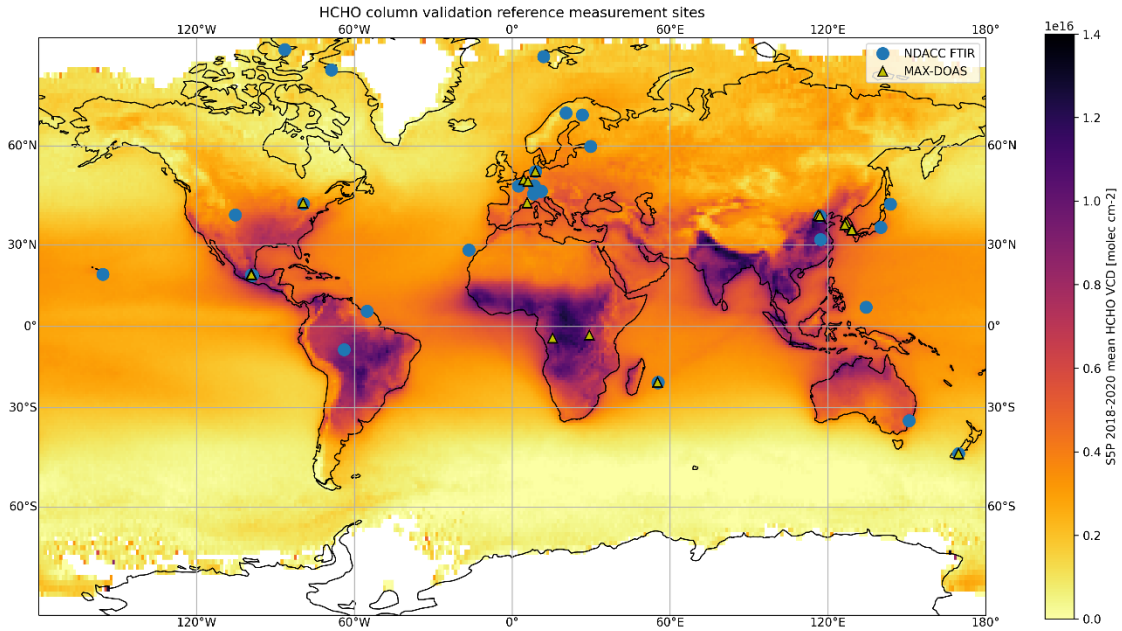


Figure 4-6: Geographical distribution of the NDACC FTIR and MAX-DOAS spectrometers as used for the validation of L3 tropospheric HCHO satellite column data in this report. Background is a multi-annual mean tropospheric vertical column density (VCD) based on S5P HCHO observations.

4.6.3. SO_2 column validation data sources

The FRM data for SO_2 column validation is quite limited. MAX-DOAS data is collected for the sites Xianghe, Wuxi and Mohali from different providers. PGN Pandora data is available for many sites (see overview in [RD34:PVPv2]), but only at a small subset enhanced SO_2 is present. The overview of the considered subset is shown in Table 4-8, Figure 4-7 and Figure 4-8. The PGN station Altzomoni was finally not kept, as the SO_2 signal, presumably from volcanic emission, was often perceived by the satellite but missed by the Pandora instrument.

Table 4-8. MAX-DOAS and Pandora data used for the SO_2 column validation.

	Instrument	lat, lon	Temporal range	Institute or PI
Xianghe	MAX-DOAS	39.75°, 116.96°	2010/03-2021/10	BIRA-IASB
Wuxi	MAX-DOAS	31.57°, 120.31°	2012/01-2014/11	Yang Wang
Mohali	MAX-DOAS	30.667°,75.729°	2019/05-2024/06	MPIC/IISER
Vallejo	PGN Pandora	19.483°, -99.147°	2019/11-2024/04	PGN, UNAM
Unam, Mexico city	PGN Pandora	19.326°, -99.176°	2018/04-2022/12	PGN, UNAM
Wakkerstroom	PGN Pandora	-27.349°, 30.144°	2020/01-now	PGN, WITS

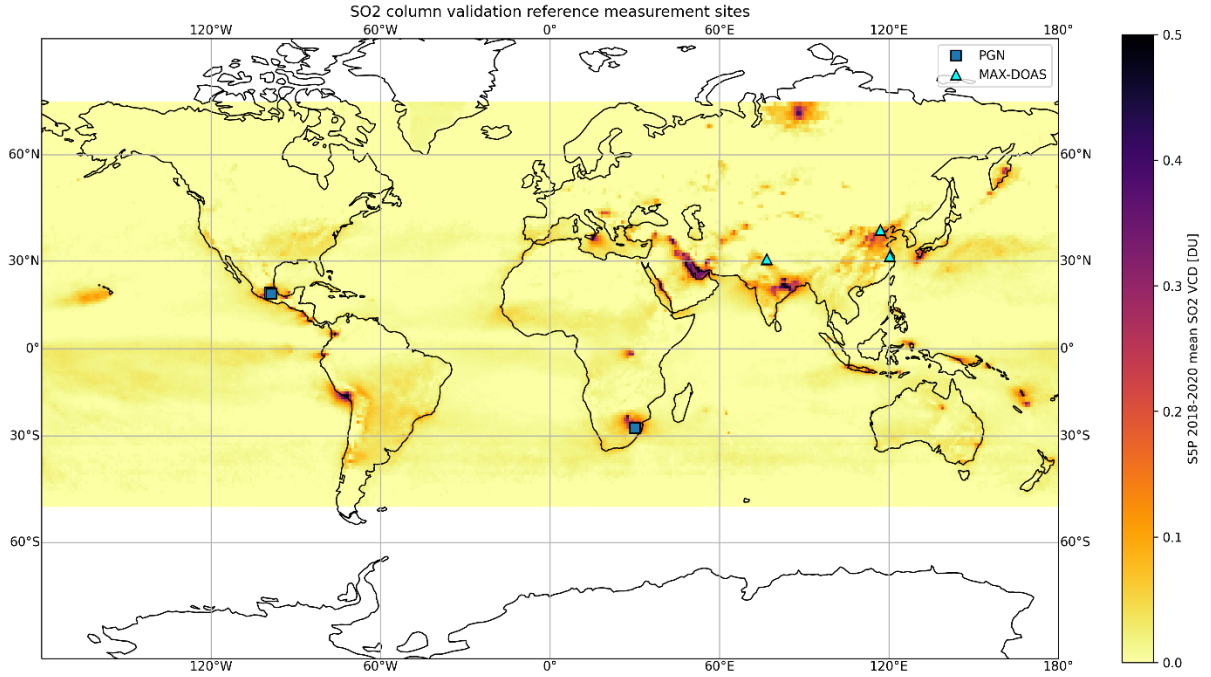


Figure 4-7: The MAX-DOAS spectrometer locations at Xianghe, Wuxi and Mohali, and the PGN Pandora instrument locations at Unam, Vallejo, Altzomoni and Wakkerstroom as considered for the validation of L2 SO_2 satellite column data in this report. Background is a multi-annual mean tropospheric vertical column density (VCD) based on S5P SO_2 observations.

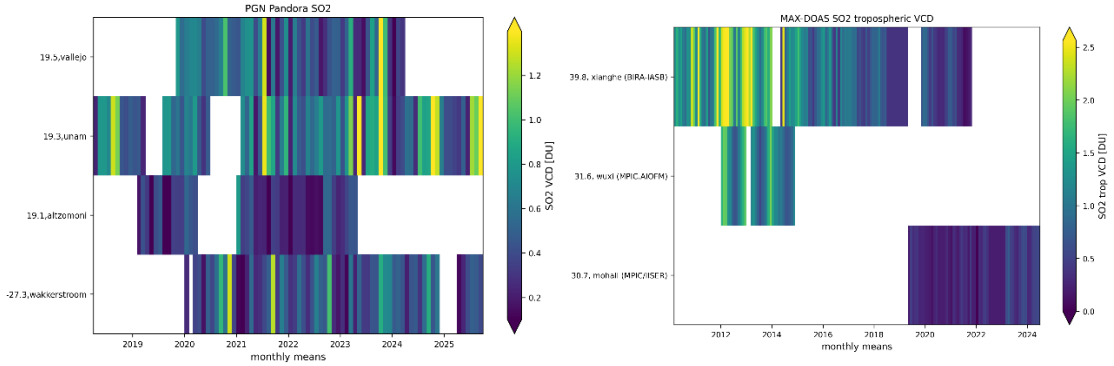


Figure 4-8. Monthly mean SO_2 VCD per station for (left) PGN Pandora and (right) MAX-DOAS data.

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 [RD34:PVPv2] for a discussion on the NDACC FTIR CO data. The geographical distribution is displayed in Figure 4-9.

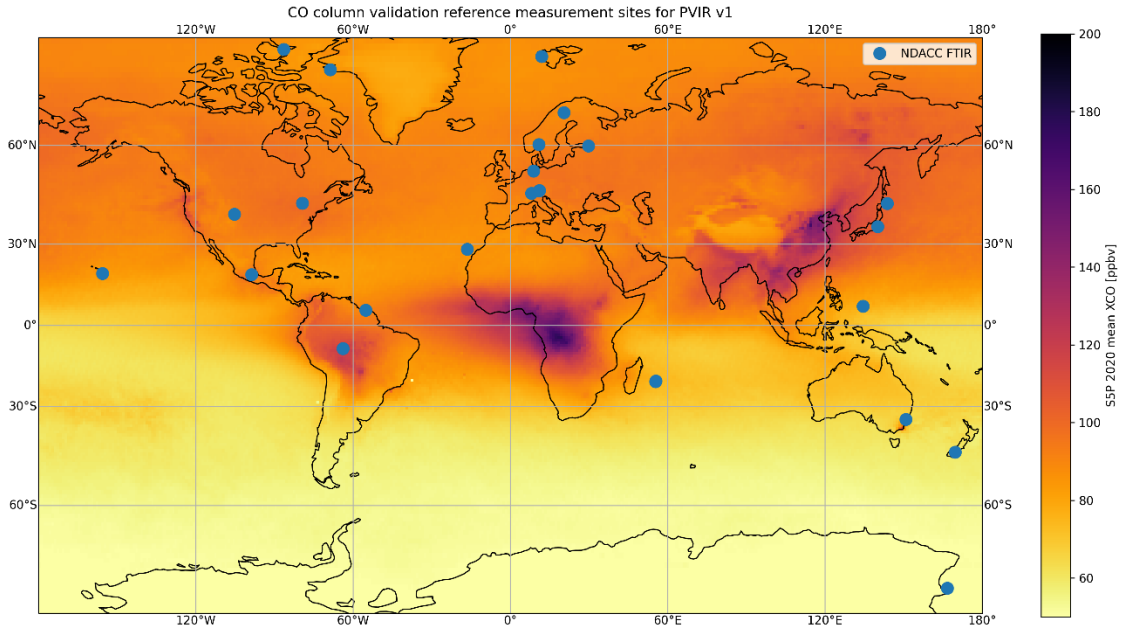


Figure 4-9: Geographical distribution of the NDACC FTIR spectrometers as used for the validation of L3 CO satellite column data in this report. Background is the 2020 annual mean dry-air column VMR based on S5P CO observations.

4.6.5. NH_3 column validation data sources

Regarding NH_3 measurements by FTIR [RD33:Dammers2015], although there is in principle data from over 16 FTIR stations [RD34:PVPv2], 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 4-10.

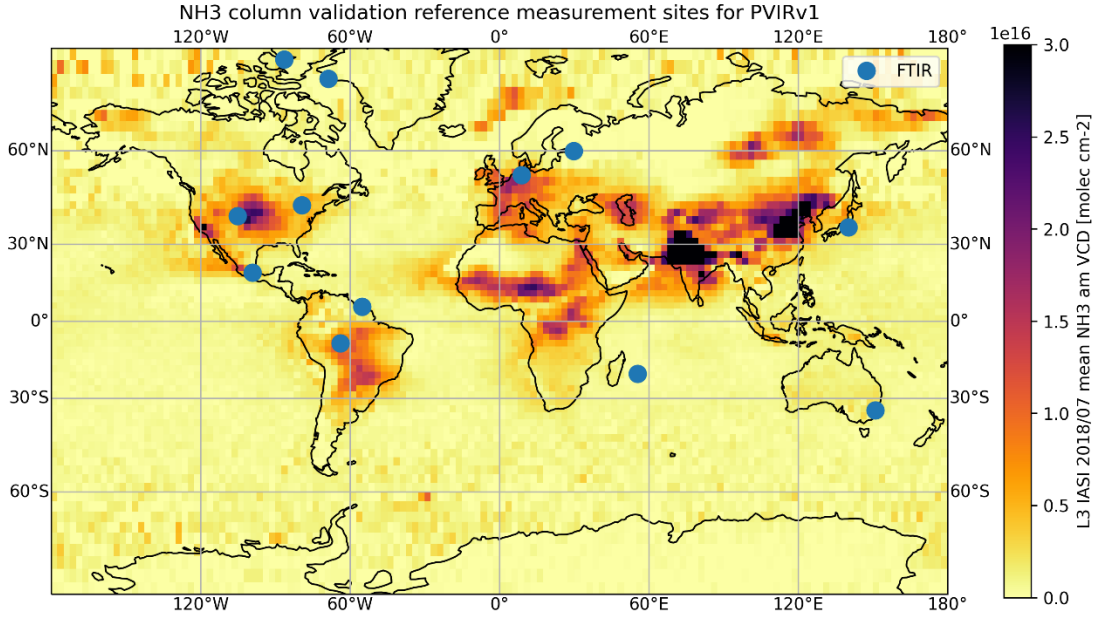


Figure 4-10: Geographical distribution of the NDACC FTIR spectrometers as used for the validation of L3 NH_3 satellite column data in this report. Background is the 2018/07 mean VCD based on the IASI L3 product.

4.6.6. **CHOCHO column validation data sources**

Regarding glyoxal, observations from MAX-DOAS stations were collected. The number of data sets that could be used is limited and the overview is provided in Table 4-9.

Table 4-9. Overview of MAX-DOAS stations providing glyoxal column observations used for validation.

Station	Institute	Lat	Lon	Time
Orleans	IUP-UB	47.96°N	2.11°E	03.07.2023 – 31.12.2024
Athens	IUP-UB	38.05°N	23.86°E	01.01.2018 – 31.12.2024
Incheon	IUP-UB	37.57°N	126.64°E	06.10.2021 – 15.11.2022
Chiba	Chiba University	35.63°N	140.10°E	01.01.2019 – 31.12.2021
Uccle	IASB/BIRA	50.8°N	4.36°E	01.01.2017 – 29.01.2020
Xianghe	IASB/BIRA	39.75°N	116.96°E	08.03.2010 – 30.04.2021

4.6.7. **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.) [RD54: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* [RD57:Rodgers2003][RD59: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 [RD55: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 covers the following data sets:

1. L2 stratospheric vertical column densities (VCD) obtained with GOME, SCIAMACHY, GOME-2A, and 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.
2. L3 total, tropospheric, and stratospheric columns for all L3 products listed in Table 3-1: GOME, SCIAMACHY, GOME-2A/B/C, OMI and TROPOMI. While several horizontal gridding resolutions are available for OMI and TROPOMI, results here cover in general the 0.2x0.2 degree^2 grids that are available for all L3 products, except for the stratospheric comparisons for TROPOMI and OMI, where the coarser resolution 1.0x1.0 degree^2 grid was used to better match satellite and ground-based measurement resolution (horizontal smoothing of ZSL-DOAS measurements is large).

5.2. Validation methodology

The validation of the NO_2 products generaly follows the L2-validation methodology already published in [RD40:Compernolle2020], [RD41:Pinardi2020] and [RD39:Verhoelst2021], which builds upon previous developments in the AC SAF, H2020 QA4ECV, and the S5P MPC VDAF. Specific settings and adaptations needed to meaningfully validate L3 data are detailed below.

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 (OMI, TROPOMI) data were additionally filtered on the snow-ice flag for some of the analyses, as recommended for tropospheric VCD[RD8:PSD_QA4ECVNO2]. PGN reference measurements were filtered by excluding flags 1,2,11,12,20,21, and 22. No filtering was applied on the consolidated LATMOS ZSL-DOAS data. No specific filters were applied on the MAX-DOAS VCD datasets either, although some of the these datasets provide cloud flags, that were sometimes used in past L2 validation exercises. The UNAM MAX-DOAS sites were excluded from the global VCD analysis because of some inconsistent behaviour w.r.t. other high-VCD sites.

Co-location. The co-location procedure depends on the type of (sub-)column.

Stratospheric columns

For stratospheric columns, satellite data are co-located with sunrise and sunset ZSL-DOAS measurements separately, allowing at most 12 hour time difference. Analysis of the sounders with a morning equatorial overpass time 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 L2 GOME and SCIAMACHY data, only an intersection between satellite pixel and this observation operator is required. For higher-resolution L2 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).
For the L3 data sets, the satellite grid cell covering the monthly mean spatial barycenter of the observation operator was selected.

Total and tropospheric columns

Satellite total and tropospheric columns are compared to direct-sun and MAX-DOAS ground-based measurements respectively. Co-location consists here of taking only satellite pixels (or L3 grid cells) covering the stations, and using only the ground-based measurements obtained within a window of -30 to + 30 minutes w.r.t. the satellite overpass time (or monthly mean overpass time-of-day in the case of L3).

In all L3 cases, all available reference measurements within a month (and within the mean overpass time-of-day window) were used to construct the monthly means, i.e., the ground-based data was not subsampled to the underlying availability of satellite L2 for the satellite L3 creation, as this filter information is not available in the L3 files. As a result, and especially for cloudy months, the ground-based sampling can be very different from the (unknown as based on L2) satellite sampling. As discussed in Sect. 5.3.2.3 this can cause large differences with respect to L2 validation results, where the ground-based data is matched to the subsampled (especially due to cloud filtering) L2 data before constructing monthly averages. For high pollution cases (especially in the morning and in the early 2000-2010), this can have a strong impact on the ground-based monthly average. The impact is expected to be higher for comparisons with MAX-DOAS than for comparisons with direct sun DOAS, since the latter observations are already tilted towards clear-sky conditions.

Harmonization.

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 [RD58:Errera2001][RD24: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%. As pointed out above, only sunset data are used, also for the satellites with an AM nominal overpass time.

For the subset of MAX-DOAS sites providing retrieved profiles (around 23 out of 70), apart from the column-to-column comparison, also vertical harmonization was done as follows:

- The MAX-DOAS profile is vertically regridded with conservation of mass [RD56: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** [RD59:Eskes2003].

*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.
**Specifically regarding the GOME-2 A/B/C NO_2 data, which include as only NO_2 products a prior profile, a substitution of the MAX-DOAS prior before application of the satellite averaging kernel by that of the satellite was attempted ([RD57: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.

5.3. Validation results

While the primary Precursors_cci+ NO_2 products are the L3 gridded data, the stratospheric column data in the underlying L2 products have been validated first so as to provide context for the L3 product validation. The focus is on the stratospheric columns as (1) these can benefit from long-term reference measurements available from before the start of the first considered satellite mission (GOME since 1996), and (2) the relatively smooth stratospheric NO_2 field ensures that effects due to resolution differences (between L2 and L3 products) remain minimal. The latter ensures that validation results from L2 should translate smoothly to validation results for L3, at least for long-term biases (the monthly averaging in the L3 products is expected to reduce comparison dispersion, which can also be verified).

5.3.1. Level-2 stratospheric NO_2 column products

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

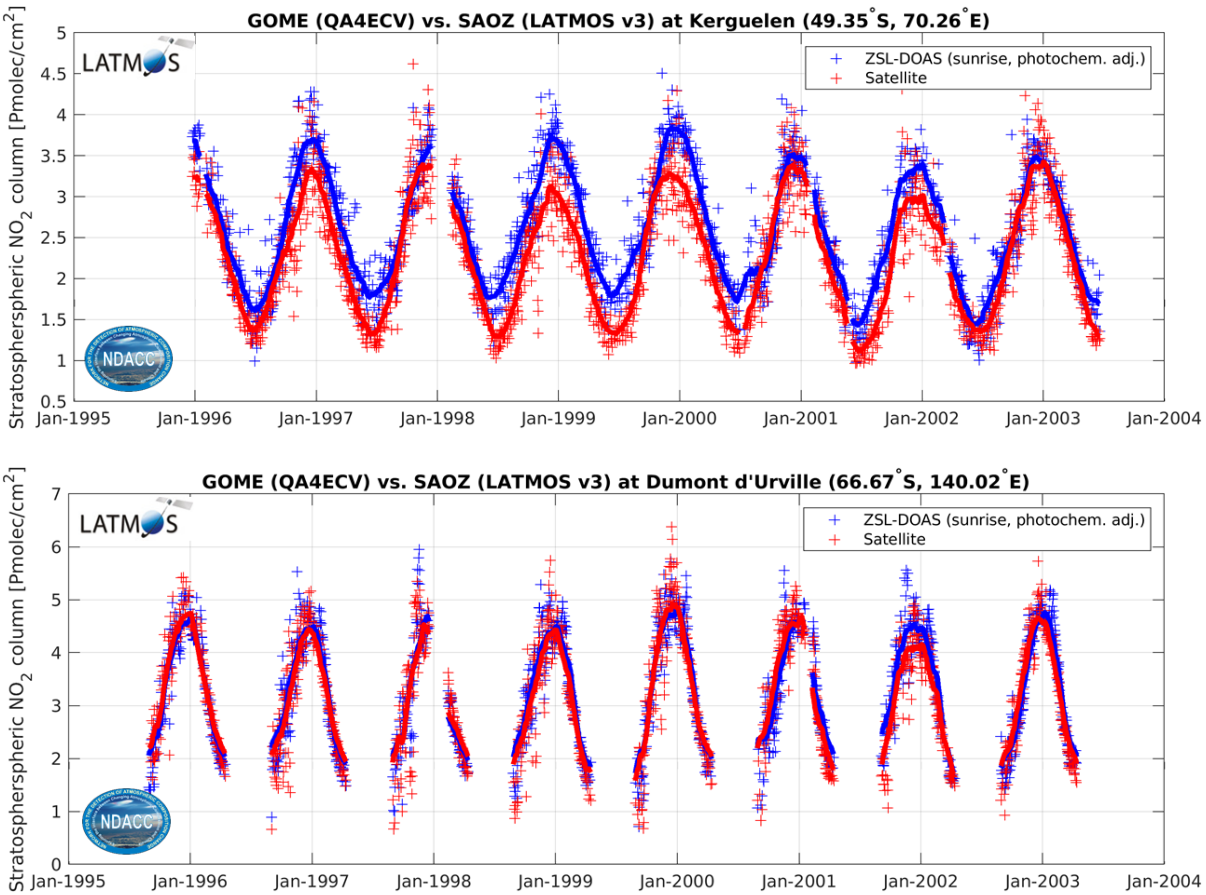


Figure 5-1: Illustration of typical comparison time series, here for GOME on ERS-2 versus the SAOZ on the Kerguelen Archipelago in the Southern Indian Ocean (top panel) and versus the SAOZ at Dumont d'Urville on the East coast of Antactica (bottom panel). Data gaps in the latter correspond to polar night.

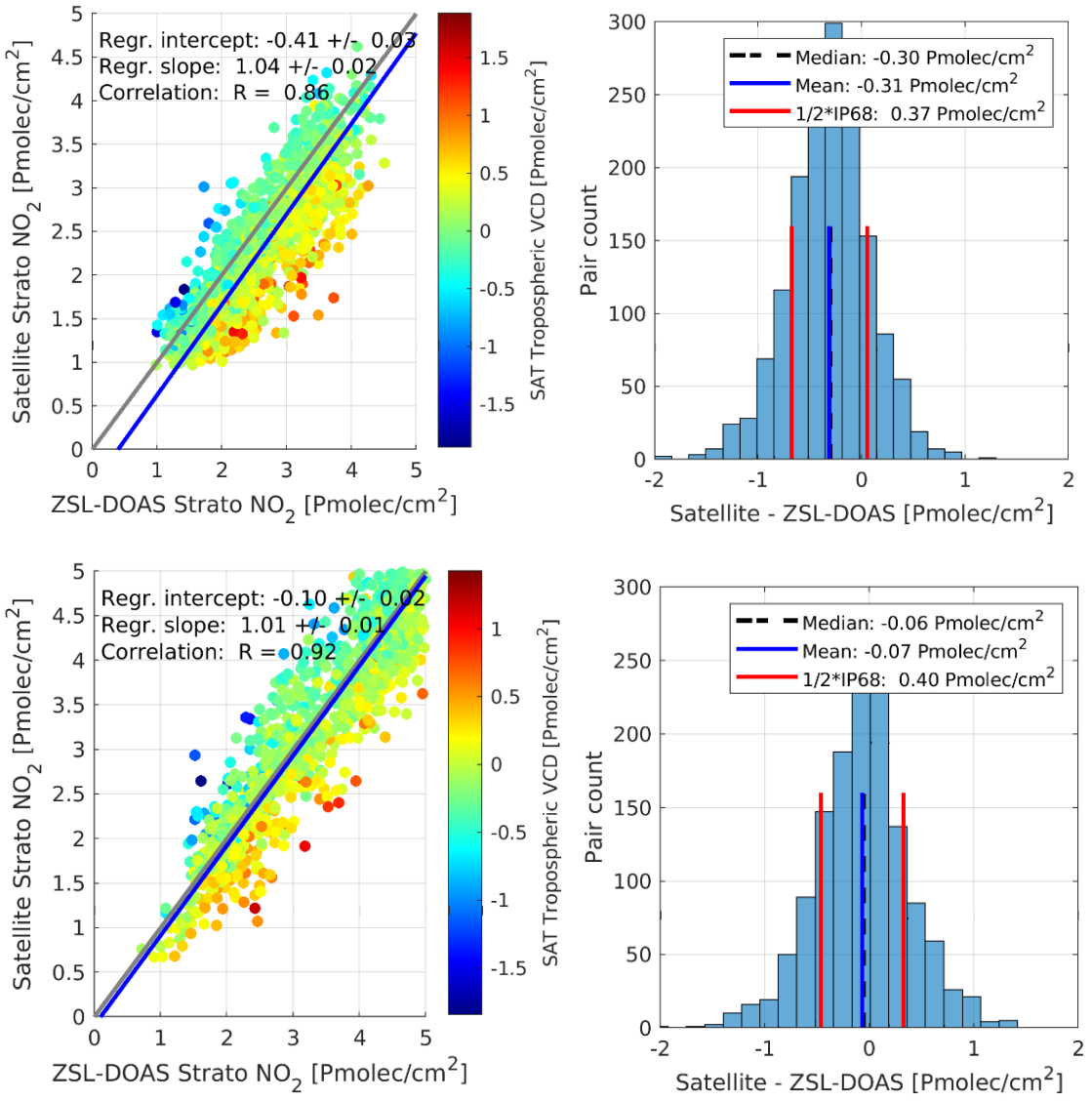


Figure 5-2: Scatter plots with regression analysis (left-hand panels) and histograms of differences (right-hand panels) for the comparisons at Kerguelen (top row) and Dumont d'Urville (bottom row) already visualized in Figure 5-1.

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 5-3 to Figure 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/cm^2, 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.

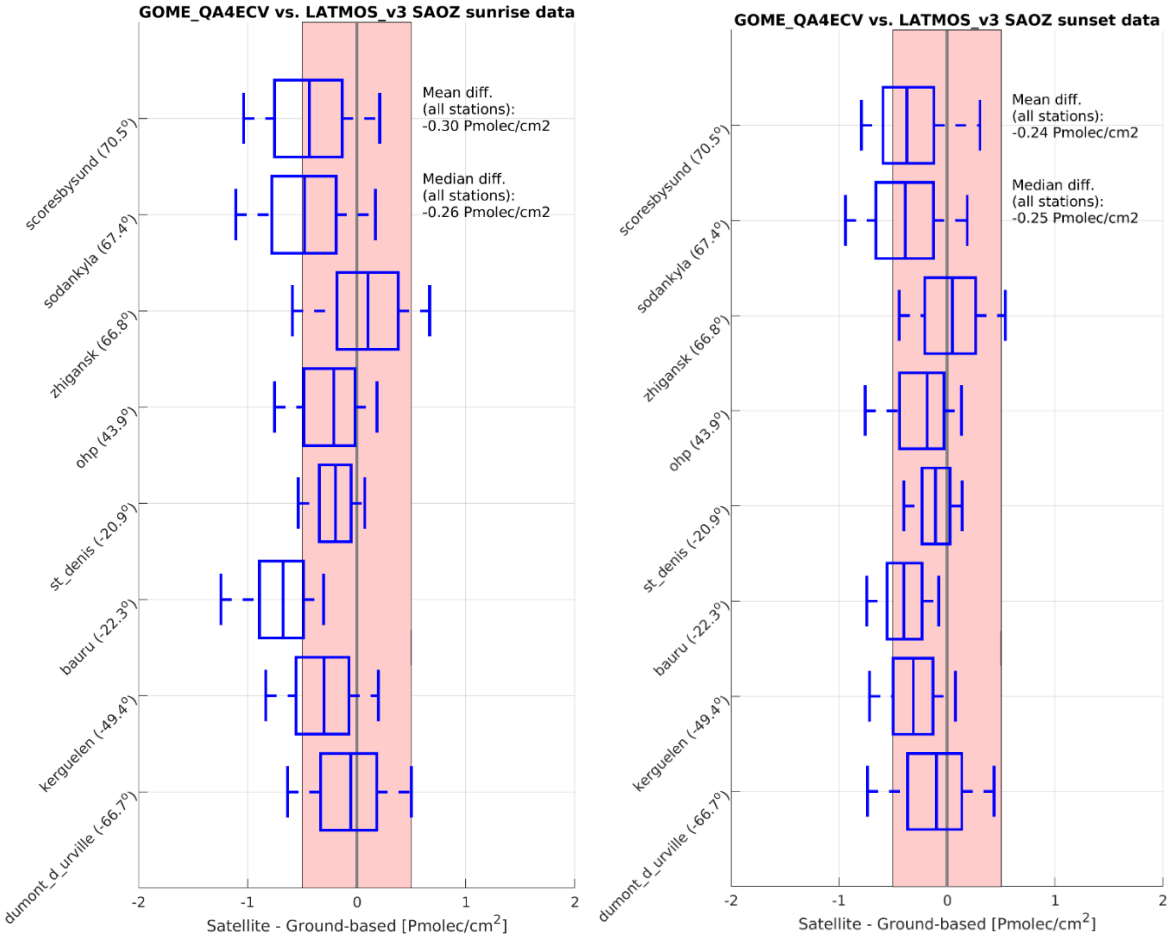


Figure 5-3: Network-wide statistics for the QA4ECV DA product of GOME on ERS-2, either versus sunrise SAOZ data (left-hand panel) or versus sunset SAOZ data (right-hand panel). Stations are ordered from South (bottom) to North (top). The boxes cover 25 – 75% of the differences, the whiskers extend from 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 [RD35:PVASR]).

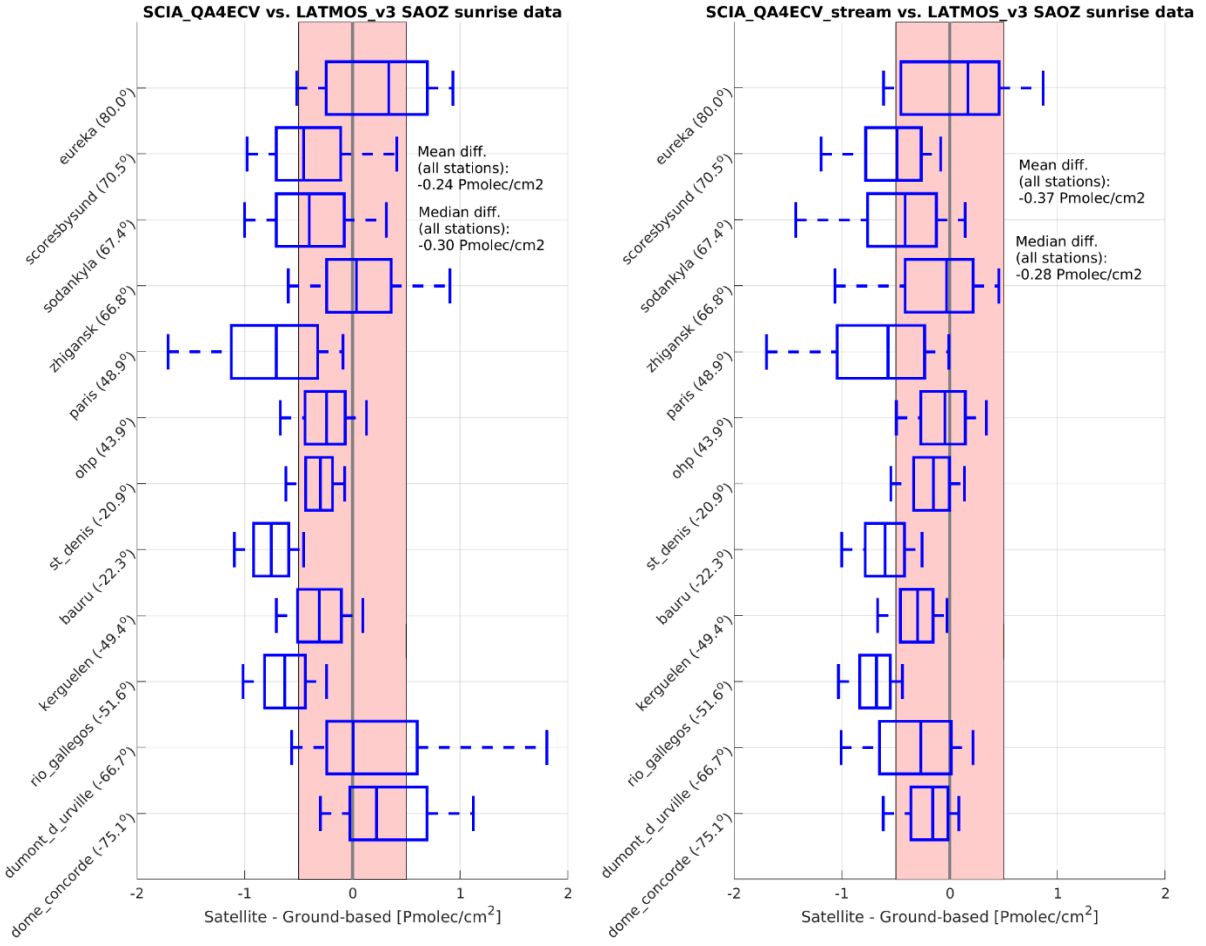


Figure 5-4: Network-wide statistics for the QA4ECV DA (left-hand panel) and QA4ECV STREAM (right-hand panel) products of SCIAMACHY on ENVISAT.

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 therefore 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 5-3 and Figure 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).

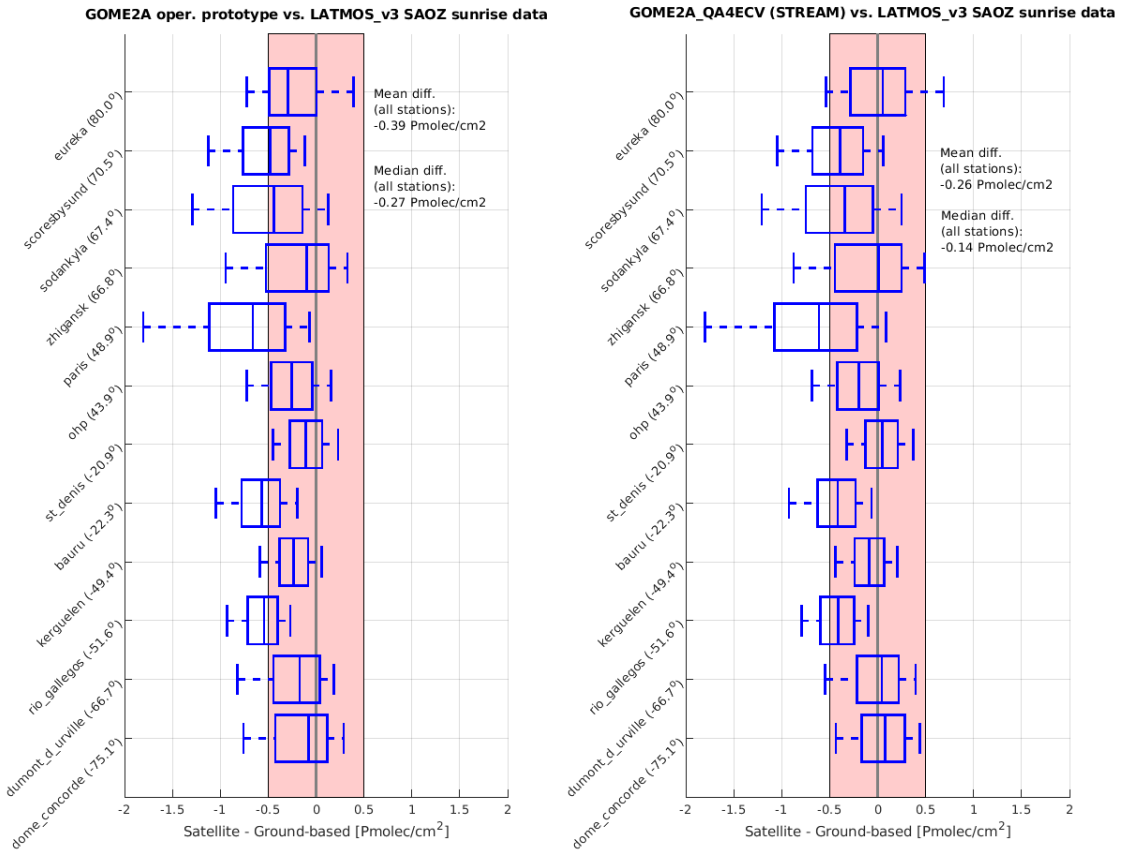


Figure 5-5: Network-wide statistics for the AC SAF prototype (left-hand panel) and QA4ECV STREAM (right-hand panel) products of GOME-2 on Metop-A.

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-2 and SCIAMACHY QA4ECV DA data, it is recommended to exclude data using the snow-ice flag.

5.3.2. Level-3 NO_2 products

The Level-3 products validated below are built on the Level-2 products validated above. The following steps were implemented along the course of the project to create the Level-3 data:

- Removal of the measurements in the ascending part of the orbit for GOME-2 A/B/C (impacting high-latitude measurements).

- Specifically for TROPOMI, an averaging scheme that accounts for representativeness uncertainty [RD10:Glissenaar2025].

5.3.2.1. Product specification and file content checks

We note the following:

- The GOME2-A/B/C products have the prior profile included which is useful for validation, but the other NO_2 products have not. We recommend to have the prior profiles included for all products.
- The GOME2-A/B/C products don't have the time variable within the data file. This information can be extracted from the filename or the netcdf attributes, but it would be more convenient, and consistent with the other NO_2 products, if the info is available as variable. Therefore we recommend to include it.
- The TROPOMI averaging kernel is NaN-filled in the considered version fv0122. We were informed by the product providers that this is fixed in the later version fv0125.
- NO_2 and HCHO products can be relevant for the same users (e.g., study of NO_2/HCHO concentration ratio). Therefore, it would be good if the resolution of the NO_2 and HCHO products were aligned, which is currently not the case. E.g., the S5P NO2 monthly product has a spatial resolution 0.2°x0.2°, while the S5P HCHO monthly product has resolution 0.125°x0.125°.
- Some variable names are inconsistent over the different products which makes it difficult for the user. For example,
    - The averaging kernel, (OMI: NO2_column_number_density_avk; S5P: NO2_averaging_kernel, others: averaging_kernel).
    - The uncertainty of the tropospheric NO_2 VCD has different names for OMI, for S5P and for the other products.
    - Names for the tropospheric AMF and for the L3 qa flag also vary.
    - The stratospheric NO_2 uncertainty is not provided in all products.

5.3.2.2. Level-3 stratospheric NO_2 columns

As a representative example of Level-3 stratospheric NO_2 column comparisons, Figure 5-6 shows the comparison between the long Aura-OMI record and the SAOZ at Observatoire de Haute Provence (OHP), which is a mid-latitude station representing relatively clean conditions in summer while being affected by some tropospheric pollution in winter.

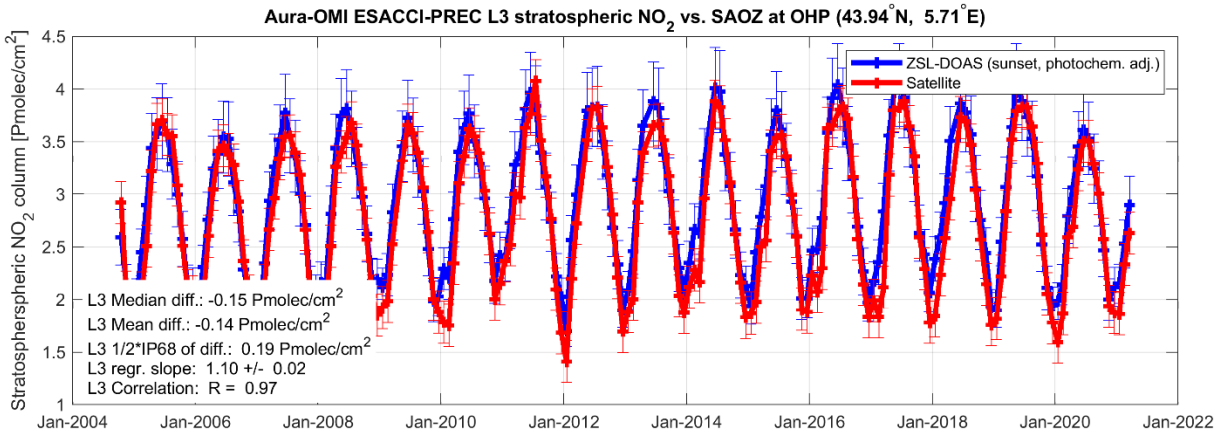


Figure 5-6: Time series of the Level-3 OMI stratospheric NO_2 columns (at a 1x1 deg^2 grid resolution) and co-located SAOZ sunset observations (adjusted to OMI overpass time) at the Observatoire de Haute Provence (France).

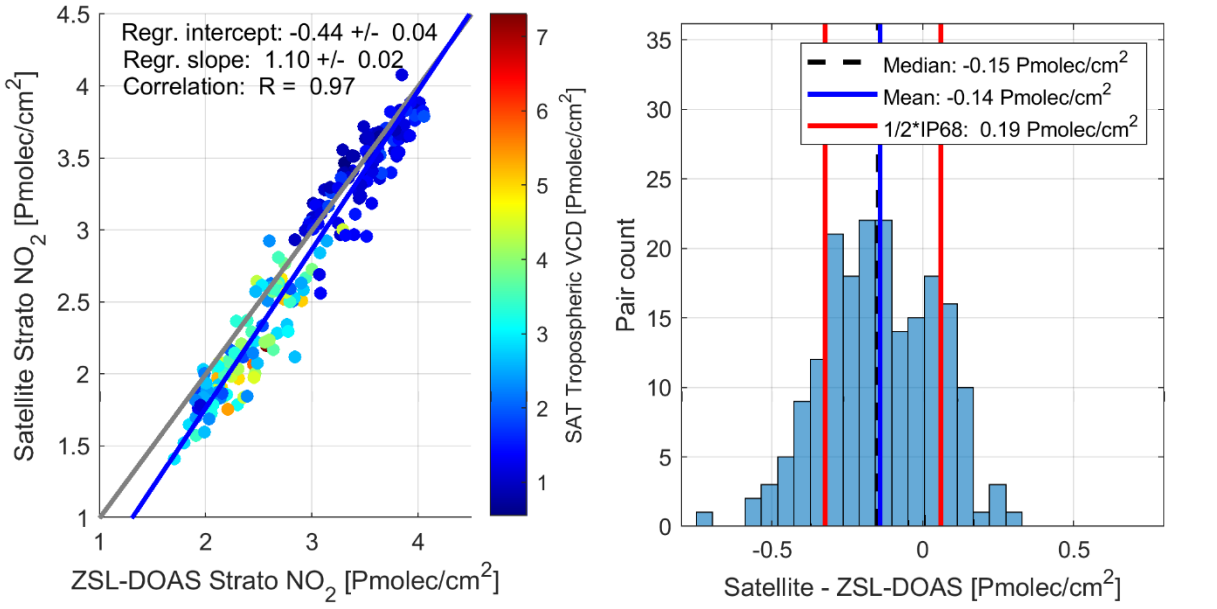


Figure 5-7: Regression analysis and histogram of the differences corresponding to the co-located Level-3 OMI & SAOZ stratospheric NO_2 time series at OHP visualized in Figure 5-6.

As further demonstrated in Figure 5-7, the correlation between satellite and ground-based records is excellent (R = 0.97), the histogram of the differences quite symmetrical, and the bias and dispersion are similar to what is seen for OMI Level-2 validation. The regression analysis shows that the small negative bias is mostly accrued during winter months. A similar analysis, also at OHP but for the morning sensor GOME-2 on MetOp-B is presented in Figure 5-8 and Figure 5-9, revealing overall similar performance albeit with a slightly stronger negative bias, in particular during winter-time low stratospheric NO_2 conditions.

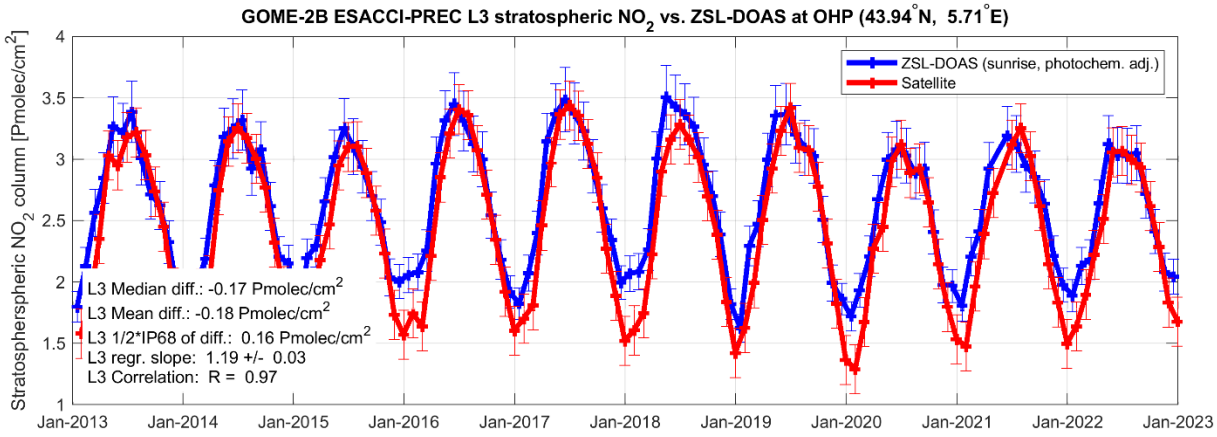


Figure 5-8: Time series of the Level-3 GOME-2B stratospheric NO_2 columns (at a 0.2x0.2 deg^2 grid resolution) and co-located SAOZ sunset observations (adjusted to MetOp-B overpass time) at the Observatoire de Haute Provence (France).

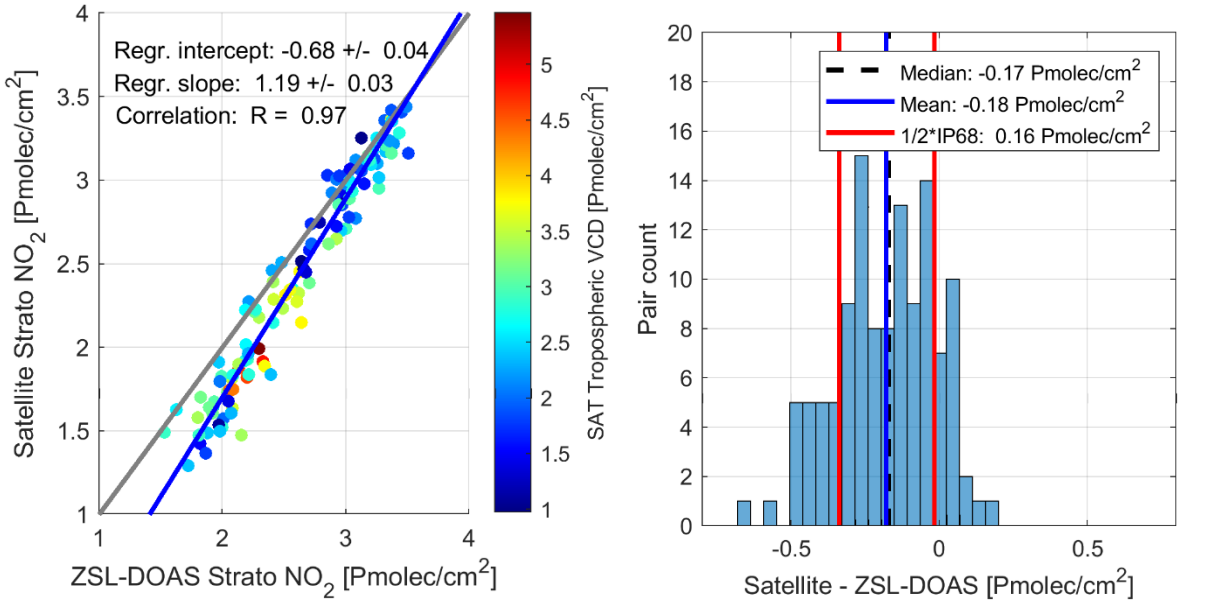


Figure 5-9: Regression analysis and histogram of the differences corresponding to the co-located Level-3 GOME-2B & SAOZ stratospheric NO_2 time series at OHP visualized in Figure 5-8.

Network-wide results are summarized as box-whisker plots (ordered by station latitude) in Figure 5-10 through Figure 5-12. Overall, we see highly consistent behavior between comparable missions (afternoon or morning overpass, type of stratospheric estimate, data producer), and across all missions.

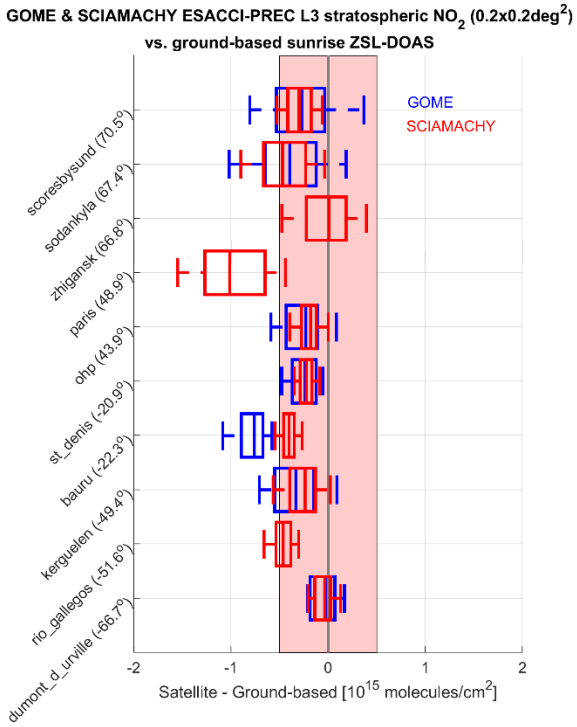


Figure 5-10: Box-whisker plots summarizing the comparison bias and dispersion for the GOME on ERS2 and SCIAMACHY on ENVISAT in blue and red respectively. Stations are ordered by latitude. Note that not all sites produced data overlapping the time coverage of both sensors.

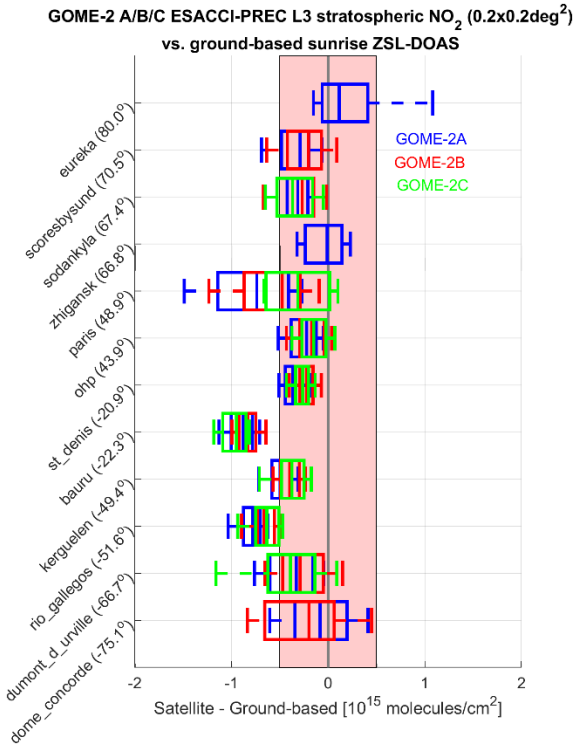


Figure 5-11: Box-whisker plots summarizing the comparison bias and dispersion for the GOME-2 instruments on MetOp-A, B, and C (blue, red, and green respectively). Stations are ordered by latitude. Note that not all sites produced data overlapping the time coverage of all three sensors.

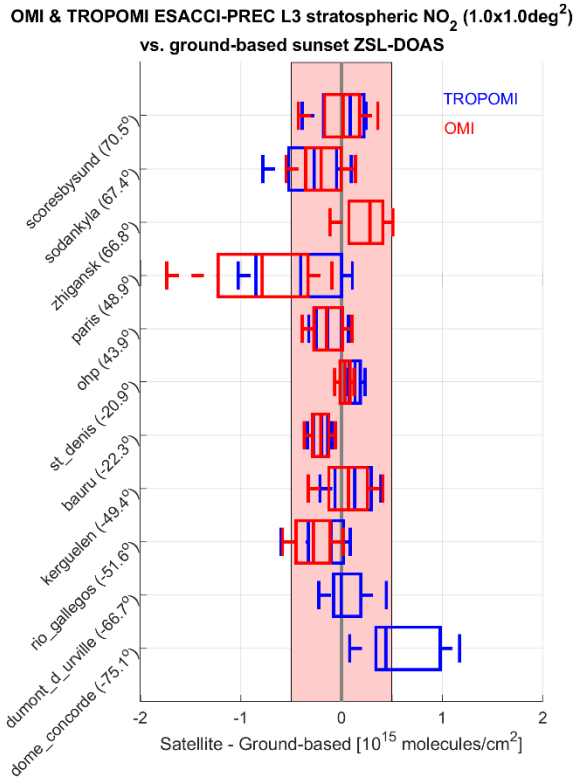


Figure 5-12: Box-whisker plots summarizing the comparison bias and dispersion for the Aura-OMI and Sentinel-5 Precursor TROPOMI instruments (red and blue respectively). Stations are ordered by latitude. Note that (1) not all sites produced data overlapping the time coverage of both sensors, and (2) quality filtering of the OMI Level-3 data induced more data loss at high Southern latitudes than for TROPOMI.

Poor results at Paris and Bauru (sunrise reference data in particular) can be related to large tropospheric pollution potentially affecting the ZSL-DOAS measurements. A more quantitative summary is presented in Table 5-2.

Table 5-2: Summary of the Quality Indicators derived from the comparison of the Level-3 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). Note that all these numbers include ground-based measurement and comparison uncertainties.

Product	Bias (Pmolec/cm2)	Dispersion (Pmolec/cm2)	Comments
GOME/ERS-2 L3 (0.2x0.2deg^2)	-0.3	0.2	Only 7 sites
SCIAMACHY/ENVISAT L3 (0.2x0.2deg^2)	-0.3	0.2	10 sites
GOME-2/MetOp-A L3 (0.2x0.2deg^2)	-0.3	0.2	12 sites
GOME-2/MetOp-B L3 (0.2x0.2deg^2)	-0.4	0.2	10 sites
GOME-2/MetOp-C L3 (0.2x0.2deg^2)	-0.4	0.2	8 sites
OMI/Aura L3 (1.0x1.0deg^2)	-0.1	0.2	Strong filtering at high latitudes
TROPOMI/Sentinel-5P L3 (1.0x1.0deg^2)	-0.1	0.2	10 sites

From this table, it is clear that the dispersion is virtually the same for all records, and significantly below that in the Level-2 comparisons (see Table 5-1), as can be expected from the monthly averaging. In fact, the largest contributor to this dispersion is the difference in seasonal amplitude observed between satellite and ground-based data records. As discussed

earlier, this may (in part) be attributable to the assumption of constant *Teff* in the ground-based data retrieval. Also a clear result is the more pronounced negative bias for the morning sensors w.r.t. the afternoon sensors. It must be kept in mind though that morning sensors were compared to sunrise data and afternoon sensors to sunset data, and that the photochemical adjustment may not be entirely bias free between morning and evening adjustments. Still, a similar differential bias is seen when comparing both morning and afternoon sensors to the sunset data only, suggesting a real (minor) discrepancy between the records. Overall, at low-to-mid latitude sites not impacted by strong tropospheric pollution, the L3 satellite stratospheric columns agree with ground-based reference data within a typical user requirement of at most 0.5 Pmolec/cm^2 difference.

Figure 5-13 and Figure 5-14 summarize drift estimates for each of the stratospheric NO_2 column records. These drifts are computed as the slope of a robust linear regression on the relative (percentage) differences between satellite and reference measurements. Results are only shown if comparisons for at least three years, sampled by at least 20 months, are available. Three years is still short to derive meaningful drift estimates, but for e.g, GOME-2C and TROPOMI, not more is available, and the large uncertainty on the drift estimate can be gauged from the depicted 1-sigma error bars. For the longest recent records (OMI, GOME-2B) and at clean, high-quality ground sites, the drift seems to be less than 5% decade. For older and/or shorter records, and at sites with (changing) tropospheric NO_2 burden (e.g, Paris), a large variety in drift estimates is observed, albeit without a clearly negative or positive pattern.

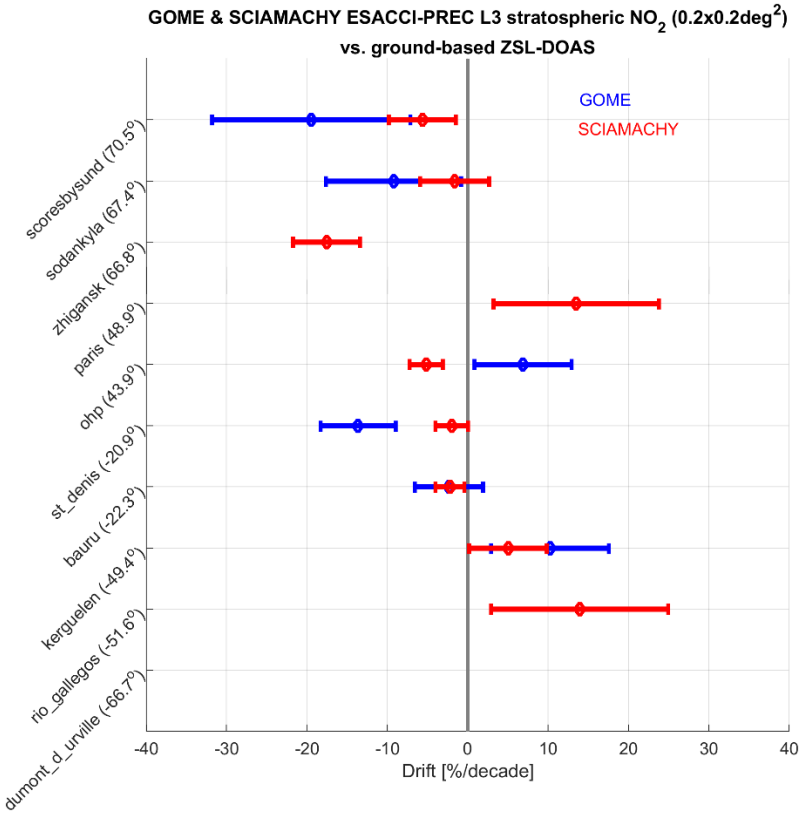


Figure 5-13: Drift estimates, expressed in %/decade, derived from a robust linear regression to the relative differences between satellite (GOME and SCIAMACHY in this graph) and ground-based measurements. Error bars represent 1-sigma uncertainty on the drift estimate.

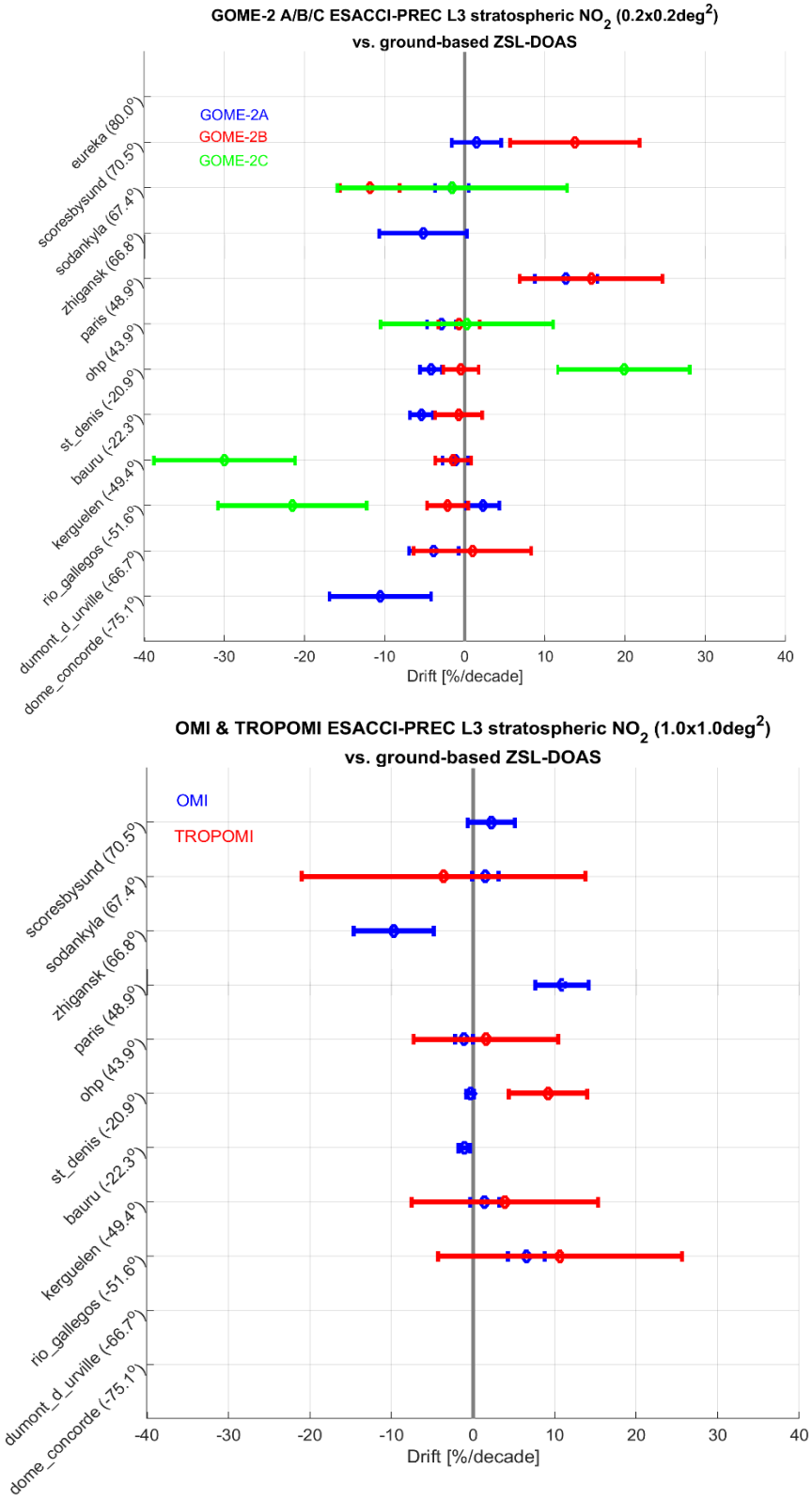


Figure 5-14: Similar to Figure 5-13 but for the three GOME-2 instruments (top panel) and for OMI and TROPOMI (bottom panel). Note that for GOME-2C and TROPOMI, the available time series are actually too short for a stable drift estimate.

5.3.2.3. Level-3 tropospheric NO_2 columns

Impact of sampling bias on L3 validation

As an illustration of the L3 tropospheric NO_2 VCD comparisons, Figure 5-15 shows the comparison between the OMI L3 data and the MAX-DOAS measurements at Xianghe, which is a sub-urban site in polluted conditions close to Beijing, China. It is clear when comparing to similar results obtained with the OMI QA4ECV L2 data (the baseline algorithm for the CCI+p L3), see Figure 5-16, that especially in the winter months the averaging of all the available ground-based data within the month (and around the mean overpass time-of-day) can strongly deviate from the value when considering only the days with common valid (i.e., clear-sky) GB-SAT pairs. The effect of sampling on the ground-based MAX-DOAS NO_2 data can be important, as MAX-DOAS also measure tropospheric NO_2 under (partially) cloudy conditions. This is less important for comparisons with Pandora, since these direct sun measurements are tilted towards cloud-free conditions. E.g., in winter 2013, the ground-based monthly median values for the L3 comparisons range from 60 to 80 $x10^{15}$ molec/cm², while in the L2 comparisons, only some of the days are identified as clear sky in the satellite dataset and thus retained, and then the ground-based monthly median is only around 40 to 60 $x10^{15}$ molec/cm². This sampling bias explains in part the worse regression results for the L3 comparison (Pearson correlation coefficient Rp=0.85 and Theil-Sen regression slope of 0.45 with an intercept of about 5x 10^{15} molec/cm²) compared to the L2 comparison (Rp=0.96, S=0.67 and I=2.8 $x10^{15}$ molec/cm²). This effect is even stronger for the morning sensors (SCIAMACHY, GOME-2A/B/C). This sampling bias was also discussed in [RD40:Compernolle2020].

Note that in [RD10:Glissenaar2025] the CCI+p L3 TROPOMI NO_2 validation was done using the L2 subsampled MAX-DOAS and Pandora data (using the operational L2 TROPOMI product) and therefore this sampling bias effect does not play here.

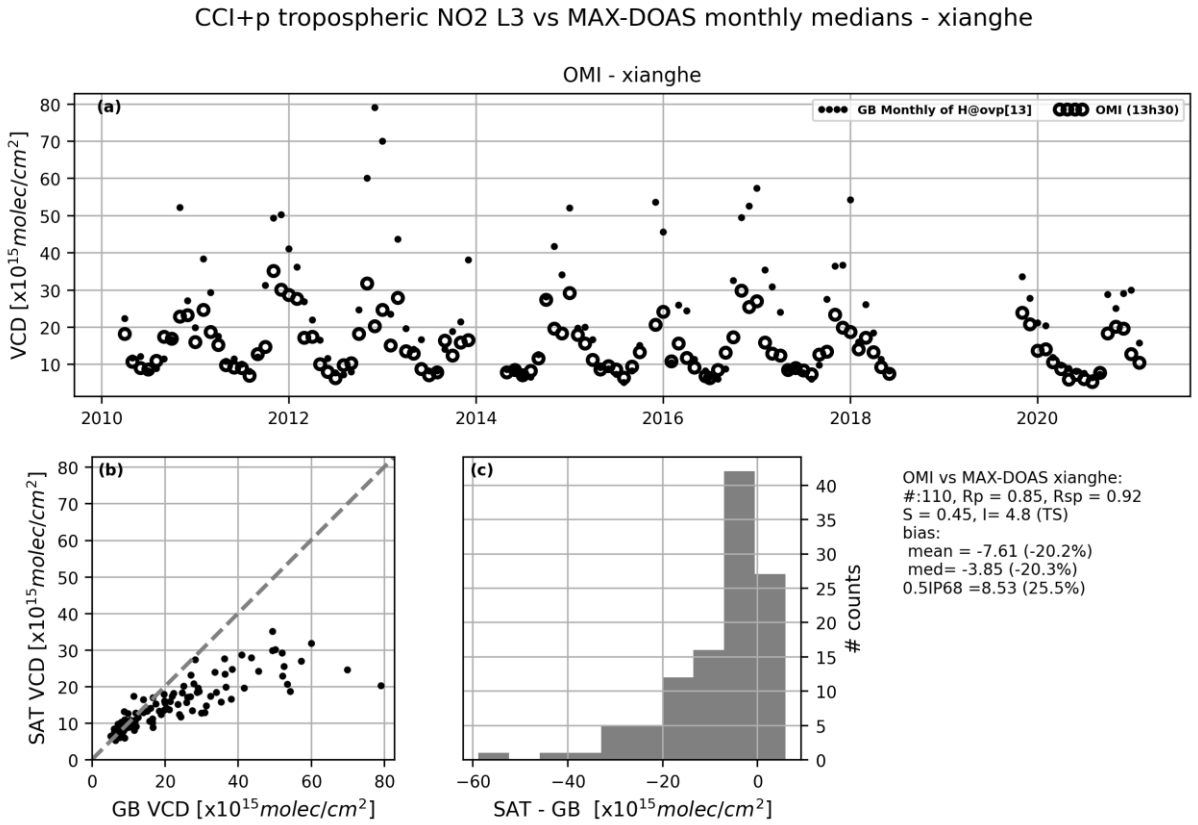


Figure 5-15: Illustration of the (a) time-series, (b) scatter plot and (c) histogram of the absolute differences for the OMI CCI+p L3 tropospheric NO_2 VCDs comparisons vs MAX-DOAS in Xianghe. The Theil-Sen regression, mean, median bias and dispersion statistics are given as inset.

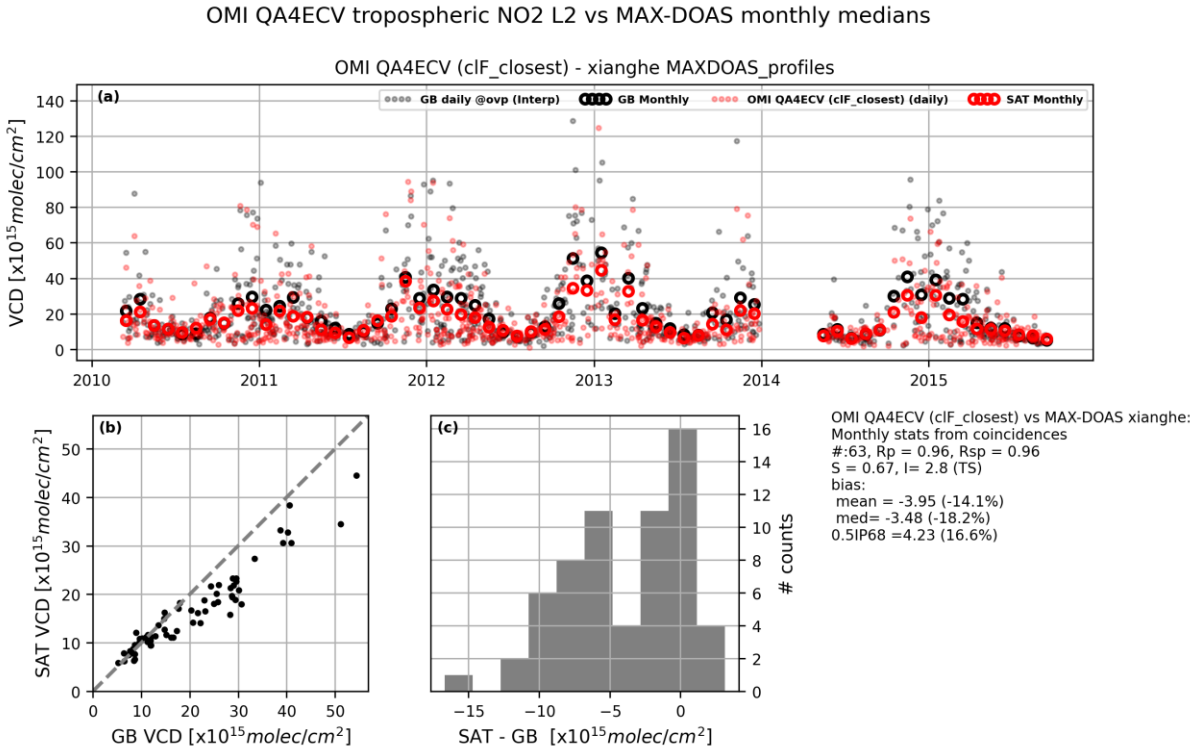


Figure 5-16: Illustration of the (a) time-series, (b) scatter plot and (c) histogram of the absolute differences for the OMI QA4ECV L2 tropospheric NO_2 VCDs comparisons vs MAX-DOAS in Xianghe. The Theil-Sen regression, mean, median bias and dispersion statistics are given as inset.

Main results

A summary of all per-sensor regression results when using all the available MAX-DOAS VCD datasets in a merged way (instead of per station) is presented in Figure 5-17 as a density plot of the satellite L3 VCDs vs the MAX-DOAS VCDs, for each sensor. The number of points and the number of datasets change significantly for each sensor, depending on the availability of the MAX-DOAS data but the behaviour is common for all the satellites. The (SAT-GB)/GB biases are generally of the order of -25% over the whole ensemble of comparison pairs. There is a tendency to strongly underestimate the ground-based data for high VCDs and some over-estimation in clean conditions, as it is highlighted by the slope smaller than one and a positive intercept (Figure 5-17) and in Figure 5-18 and Table 5-4. The specific statistics for each satellite are given in Figure 5-17 and are also summarized in Table 5-3. As discussed for the specific Xianghe case above, the L3 result are a bit worse but globally coherent with L2 validation results obtained in the past ([RD41:Pinardi2020], [RD39:Verhoelst2021], [RD40:Compernolle2020], [RD6:vanGeffen2002], , [RD31: Dimitropoulou2022], [RD42:ACSAF-VALNO2]) on OMI, GOME-2 and TROPOMI NO_2 data. The underestimation of the satellite VCD compared to MAX-DOAS can have multiple causes, among which a) inadequate satellite a-priori shape and b) the impact of horizontal NO_2 dilution within the satellite pixels, much larger than the MAX-DOAS field-of view. The first can be partly estimated here (see next subsection). The latter has been studied for L2 comparisons ([RD43:Chen2009]; [RD44:Ma2013]; [RD45:Chan2020]; [RD41:Pinardi2020]) and has been estimated to account for an underestimation of 15-25% in tropospheric column NO_2 over pollution hot-spots.

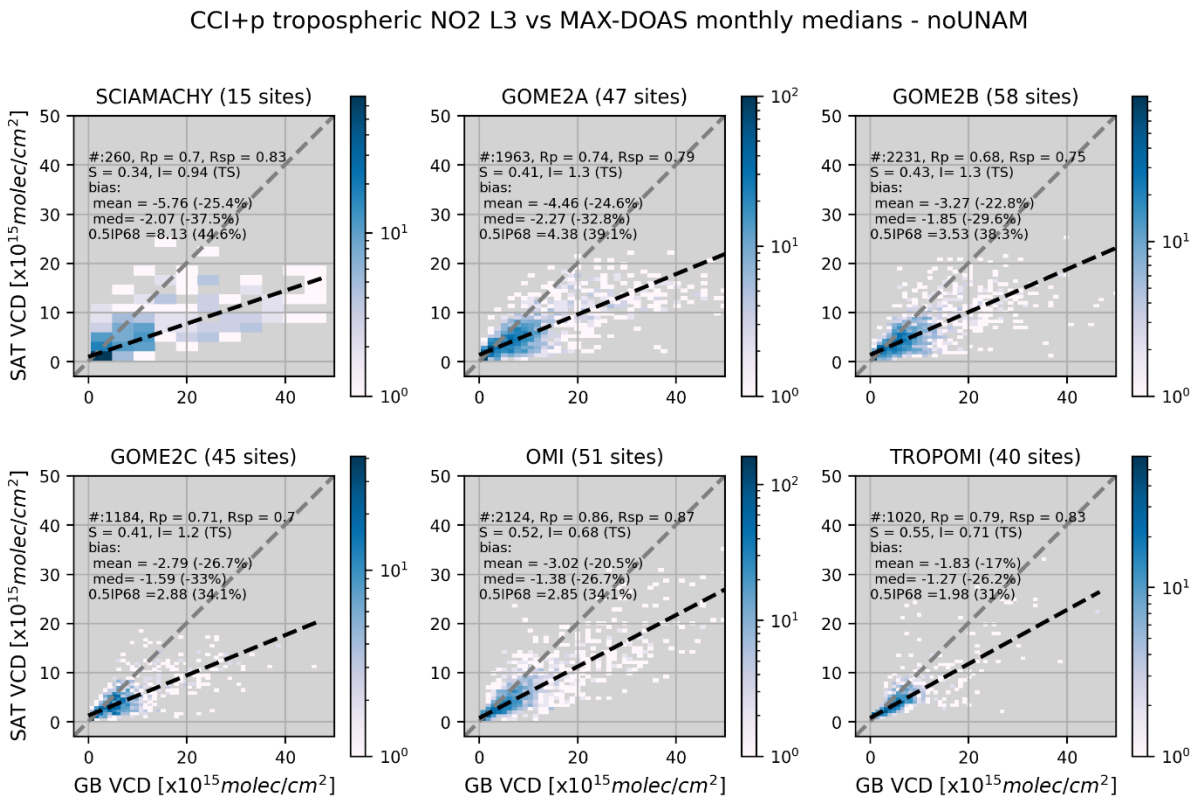


Figure 5-17: Scatter density plots for the different MAX-DOAS to L3 satellite tropospheric NO_2 VCDs comparisons. The number of sites for each comparison is given in each subplot title, and the Theil-Sen regression, as well as the following statistics: mean and median difference, dispersion (1/2IP68), and Pearson and Spearman correlation, are given as inset.

Table 5-3: Summary of the Quality Indicators derived from the comparison of the Level-3 tropospheric NO_2 VCD data sets to ground-based MAX-DOAS VCD measurements. The bias is computed as the mean or median difference of all the collocation pairs and the dispersion as their 1/2IP68 (i.e., half of the 68% central interpercentile). Rp and Rsp are Pearson and Spearman correlation respectively. Note that all these numbers include ground-based measurement and comparison uncertainties. The number of sites and number of measurement pairs are provided as well.

L3 NO_2 Product vs MAX-DOAS VCDs	Bias (Pmolec/cm^2)	Dispersion (Pmolec/cm^2)	Theil-Sen Regression Pearson & Spearman corr
GOME/ERS-2 L3 (0.2x0.2deg)	NA	NA	No available MAX-DOAS
SCIAMACHY/ENVISAT (0.2x0.2deg)	mean = -5.76 (-25.4%) med= -2.07 (-37.5%)	0.5IP68 =8.13 (44.6%)	15 sites, #:260 Rp = 0.7, Rsp = 0.83 S = 0.34, I= 0.94 (TS)
GOME-2/MetOp-A (0.2x0.2deg)	mean = -5.6 (-26.3%) med= -2.48 (-34.8%)	0.5IP68 =5.34 (38.7%)	47 sites, #:2063 Rp = 0.79, Rsp = 0.82 S = 0.39, I= 1.4 (TS)
GOME-2/MetOp-B (0.2x0.2deg)	mean = -4.71 (-25.1%) med= -2.1 (-31.7%)	0.5IP68 =4.27 (39.2%)	58 sites, #:2353 Rp = 0.71, Rsp = 0.78 S = 0.38, I= 1.6 (TS)
GOME-2/MetOp-C (0.2x0.2deg)	mean = -4.12 (-28.8%) med= -1.78 (-34.5%)	0.5IP68 =3.57 (35.1%)	15 sites, #:1249 Rp = 0.72, Rsp = 0.74 S = 0.36, I= 1.4 (TS)

OMI/Aura (0.2x0.2deg)	mean = -3.76 (-22.6%) med= -1.51 (-29%)	0.5IP68 =3.51 (34.6%)	51 sites, #:2221 Rp = 0.82, Rsp = 0.87 S = 0.48, I= 0.79 (TS)
TROPOMI/Sentinel-5P (0.2x0.2deg)	mean = -3.1 (-20.4%) med= -1.48 (-29.6%)	0.5IP68 =2.83 (31.2%)	40 sites, #:1129 Rp = 0.77, Rsp = 0.86 S = 0.5, I= 0.88 (TS)

Table 5-4: Bias and dispersion estimates for the different products, differentiated into 3 distinct (MAXDOAS-measured) tropospheric NO_2 VCD bins. For GOME, an insufficient number of MAXDOAS measurements were available. Per NO_2 VCD bin, median difference and relative difference (first line) and 1/2IP68 of difference and relative difference (second line) is provided.

Product	GND < 5 Pmolec/cm^2	5 < GND < 10 Pmolec/cm^2	10 Pmolec/cm^2 < GND
GOME/ERS-2 L3 (0.2x0.2deg^2)	NA	NA	NA
SCIAMACHY/ENVISAT L3 (0.2x0.2deg^2)	-0.1 (-9.3%) 0.8 (47.8%)	-3.2 (-38.6%) 3.3 (38.2%)	-11.3 (-62.4%) 10.4 (23.1%)
GOME-2/MetOp-A L3 (0.2x0.2deg^2)	0.1 (2.8%) 0.9 (45.4%)	-2.5 (-34.1%) 2.2 (28.6%)	-8.0 (-52.7%) 7.2 (23.5%)
GOME-2/MetOp-B L3 (0.2x0.2deg^2)	0.0 (-2.1%) 1.0 (42.5%)	-2.4 (-34.3%) 2.2 (29.6%)	-7.5 (-50.4%) 5.5 (25.3%)
GOME-2/MetOp-C L3 (0.2x0.2deg^2)	-0.3 (-12%) 0.9 (32.3%)	-2.6 (-39.3%) 2.2 (30.7%)	-8.4 (-56.6%) 5.3 (21.5%)
OMI/Aura L3 (0.2x0.2deg^2)	-0.2 (-10.5%) 0.9 (39.3%)	-2.3 (-33.3%) 2.0 (28.1%)	-6.8 (-46.1%) 6.4 (22.5%)
TROPOMI/Sentinel-5P L3 (0.2x0.2deg^2)	-0.3 (-13.2%) 0.9 (35.9%)	-2.2 (-31.9%) 1.7 (24.1%)	-6.0 (-41.9%) 4.2 (21.7%)

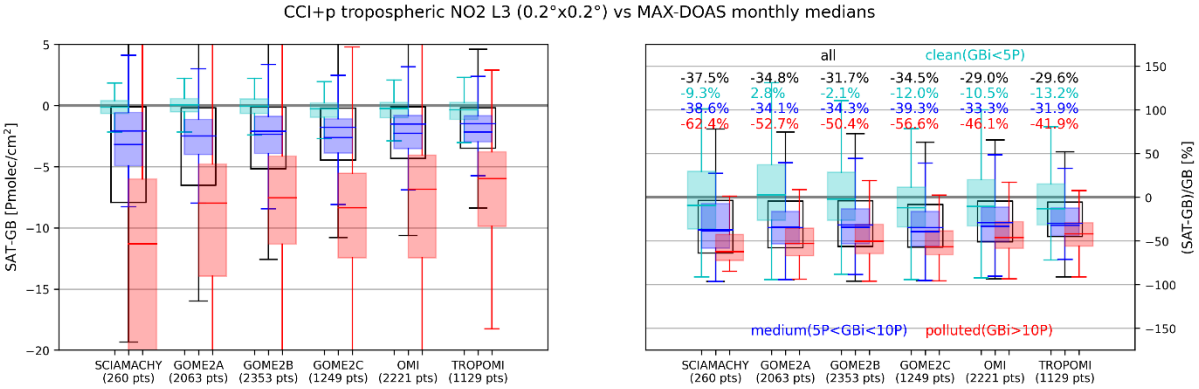


Figure 5-18: Box-whisker plots summarizing the tropospheric NO_2 comparison absolute (left) and relative (right) bias and dispersion for all the satellites instruments against MAX-DOAS VCDs. Different pollution conditions are overlayed over the results for all the points (black): cyan for clean situations (MAX-DOAS data <5 Pmolec/cm^2), blue for moderate pollution (between 5 and 10 Pmolec/cm^2) and red for polluted cases (>10 Pmolec/cm^2). Median biases are summarized in the corresponding colors for the relative differences in the right panel.

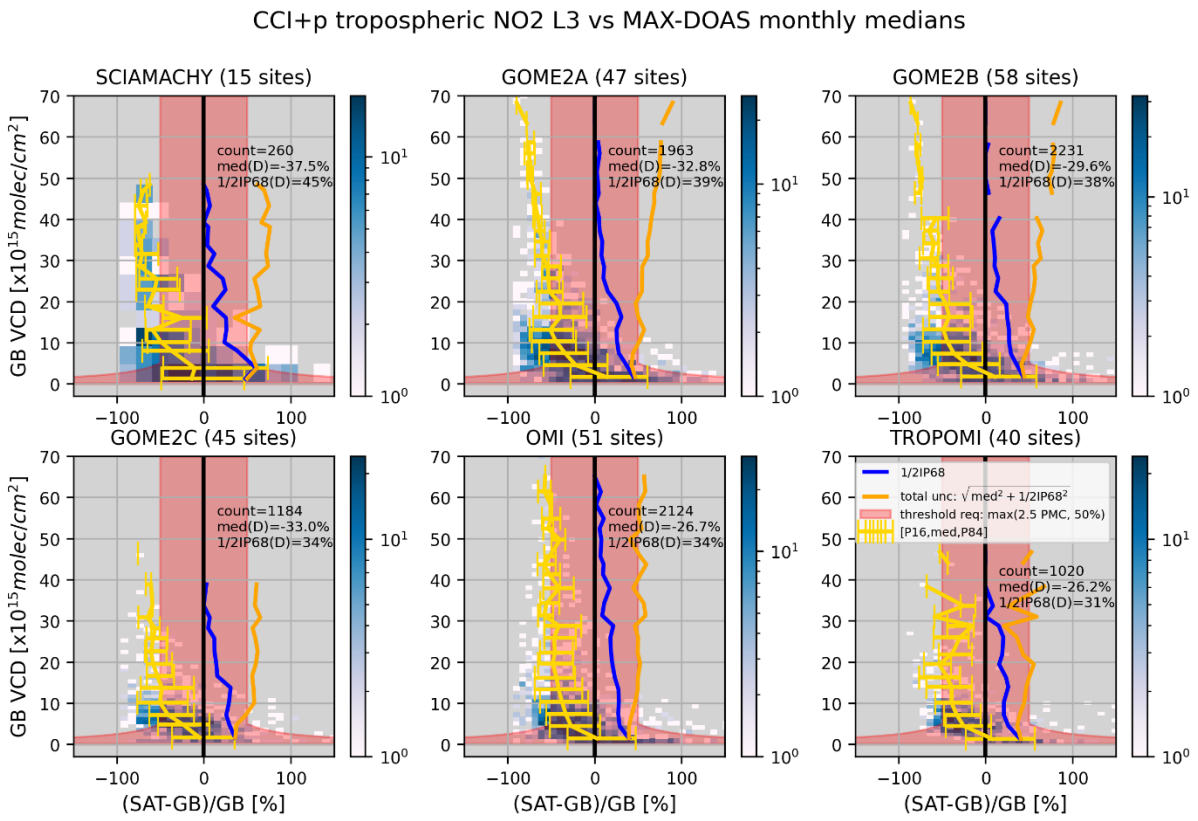


Figure 5-19: Density plots for the different MAX-DOAS to satellite tropospheric NO_2 VCDs comparisons. The relative (SAT-GB)/GB values are shown as a function of the MAX-DAOS VCD columns. The median and dispersion for bins of 3 Pmolec/cm² are overlayed (in yellow), while in blue and orange the dispersion and total uncertainties are reported. The number of sites for each comparison is given in each subplot title, and the median difference and dispersion statistics are given as inset. The red zone represents the GCOS threshold requirement (Max(50%, 2.5E15 molec cm-2)).

Impact of vertical prior – smoothing

As discussed above, for some of the sites used above, the ground-based PIs also provide vertical profiles in the low troposphere and averaging kernels. For those datasets (about 23 over 70) the smoothing of the ground-based profiles with the satellite averaging kernels (see discussion in Sect 5.2) is possible. An illustration of the results after smoothing is given in Figure 5-19 and a summary of the statistics before and after smoothing is given in Table 5-5.
It should be noted that in the currently assessed TROPOMI version (fv122) the averaging kernels variables were all filled with NaN. This has been corrected in a more recent version (fv125), but this could not be included yet in our validation results. The GOME-2 L3 files also provide the prior profiles, so additional options for smoothing, including prior harmonization and substitution, have been tested (not shown).
The impact of smoothing is not the same for all the satellites, not always reducing the bias/spread or improving the regression (as also found on L2 validation results and presented in [RD42:ACSAF-VALNO2] and ROCVR 2025 for GOME-2 and TROPOMI L2, respectively). For SCIAMACHY and OMI, the smoothing is typically reducing the bias, while for GOME-2 the impact is more variable, depending also on the type of smoothing applied.

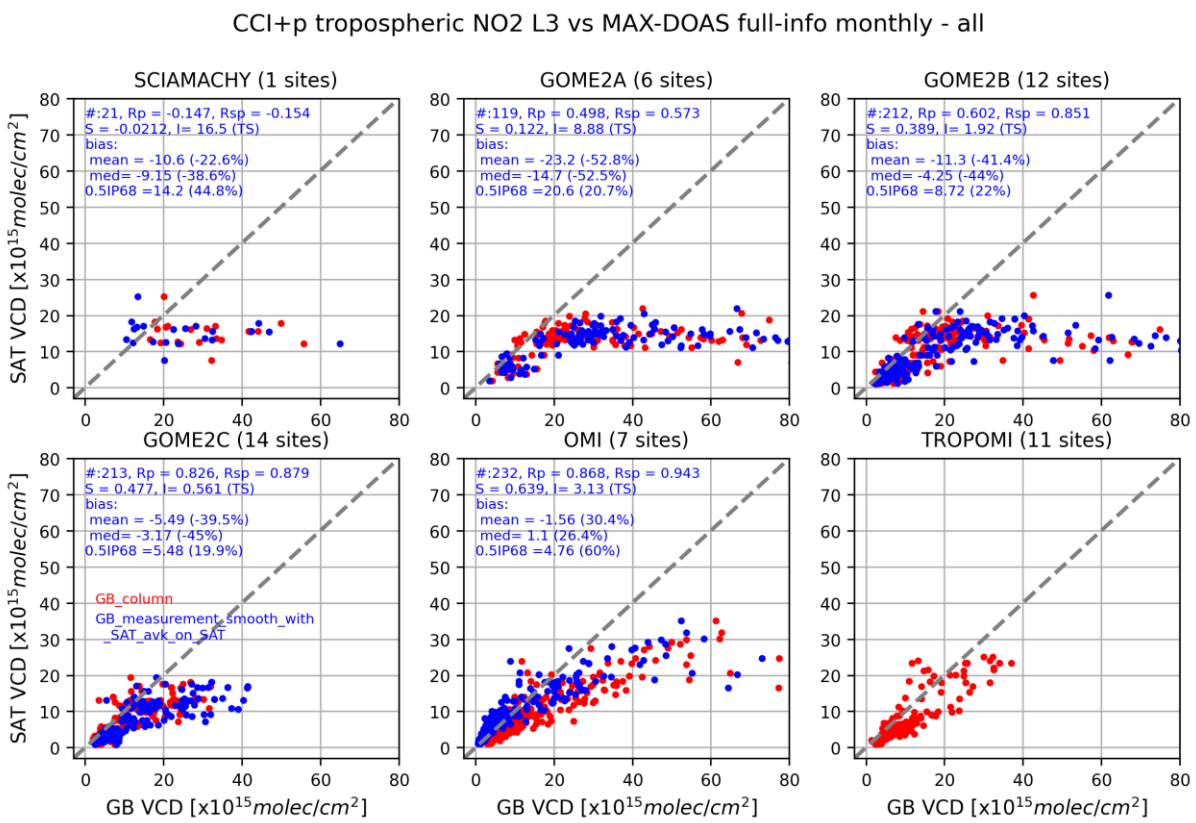


Figure 5-19: Scatter plots for the different MAX-DOAS to L3 satellite tropospheric NO_2 profiles comparisons. The different colors represent the original comparison (in red), the smoothed profiles with satellite averaging kernels (blue). The number of sites for each comparison is given in each subplot title, and the Theil-Sen regression, as well as mean, median bias and dispersion statistics for the smoothed cases are given as inset.

Table 5-5: Summary of the Quality Indicators derived from the comparison of the Level-3 tropospheric NO_2 VCD data sets to ground-based MAX-DOAS sites proving vertical profiles, and where smoothing is possible. The bias is computed as the mean or median (absolute or relative) difference over all the collocation pairs and the dispersion as the 1/2IP68 (i.e., half of the 68% interpercentile) of the differences. Note that all these numbers include ground-based measurement and comparison uncertainties. Note that since no valid AVK was present in L3 TROPOMI fv0122, smoothing could not be applied here.

L3 NO_2 Product vs MAX-DOAS profiles	Number of sites	Bias (Pmolec/cm2)	Dispersion (Pmolec/cm2)	Theil-Sen Regression Pearson correlation
GOME/ERS-2 L3 (0.2x0.2deg)	None	NA	NA	NA
SCIAMACHY/ENVISAT (0.2x0.2deg) Direct comparison *Smoothed*	1	mean = -15.2 (-42.7%) med= -14.6 (-48.6%) *mean = -10.6 (-22.6%)* *med= -9.15 (-38.6%)*	11.6 (22.7%) *14.2 (44.8%)*	#:21, Rp = -0.166, Rsp = -0.194 S = -0.0308, I= 16.8 (TS) *#:21, Rp = -0.147, Rsp = -0.154* *S = -0.0212, I= 16.5 (TS)*
GOME-2/MetOp-A (0.2x0.2deg) Direct comparison *Smoothed*	6	mean = -14.7 (-40.8%) med= -7.36 (-38.8%) *mean = -23.2 (-52.8%)* *med= -14.7 (-52.5%)*	14.2 (24.7%) *20.6 (20.7%)*	#:119, Rp = 0.438, Rsp = 0.538 S = 0.151, I= 9.05 (TS) *#:119, Rp = 0.498, Rsp = 0.573* *S = 0.122, I= 8.88 (TS)*

GOME-2/MetOp-B (0.2x0.2deg) Direct comparison *Smoothed*	12	mean = -7.49 (-33.1%) med= -3.24 (-37.1%) *mean = -11.3 (-41.4%)* *med= -4.25 (-44%)*	5.27 (27%) *8.72 (22%)*	#:212, Rp = 0.585, Rsp = 0.835 S = 0.479, I= 1.38 (TS) *#:212, Rp = 0.602, Rsp = 0.851* *S = 0.389, I= 1.92 (TS)*
GOME-2/MetOp-C (0.2x0.2deg) Direct comparison *Smoothed*	14	mean = -3.9 (-31.9%) med= -2.8 (-40.4%) *mean = -5.49 (-39.5%)* *med= -3.17 (-45%)*	3.53 (22.7%) *5.48 (19.9%)*	#:213, Rp = 0.815, Rsp = 0.859 S = 0.576, I= 0.178 (TS) *#:213, Rp = 0.826, Rsp = 0.879* *S = 0.477, I= 0.561 (TS)*
OMI/Aura (0.2x0.2deg) Direct comparison *Smoothed*	7	mean = -6.69 (-30.5%) med= -3.42 (-36%) *mean = -1.56 (30.4%)* *med= 1.1 (26.4%)*	6.31 (23.1%) *4.76 (60%)*	#:232, Rp = 0.857, Rsp = 0.897 S = 0.513, I= 1.52 (TS) *#:232, Rp = 0.868, Rsp = 0.943* *S = 0.639, I= 3.13 (TS)*
TROPOMI/Sentinel-5P (0.2x0.2deg) Direct comparison *Smoothed: not possible*	11	mean = -2.94 (-29.2%) med= -2.49 (-36%)	2.51 (22.4%)	#:198, Rp = 0.88, Rsp = 0.871 S = 0.648, I= -0.0751 (TS)

5.3.2.4. Level-3 total NO_2 columns

While stratospheric and tropospheric product validation tends to cover predominantly the ends of the clean-to-polluted spectrum, with most ZSL-DOAS instruments sampling clean background conditions and most MAX-DOAS instruments located near sources of tropospheric NO_2, the wide network of PGN instruments measuring total NO_2 columns covers the entire spectrum more continuously. Figure 5-20 contains an illustration for the most-polluted end of the spectrum: Vallejo in the megapolis of Mexico City. To assess the impact of the spatio-temporal gridding, the Level-3 comparisons to PGN are visualized here against PGN-comparisons for the underlying Level-2 data, revealing improved correlation, reduced dispersion, but also more pronounced negative biases for the Level-3 product. All of this is as expected: averaging brings the noise (instrumental and atmospheric variability) down, but the coarser horizontal resolution leads to more pronounced underestimation near pollution hot spots.

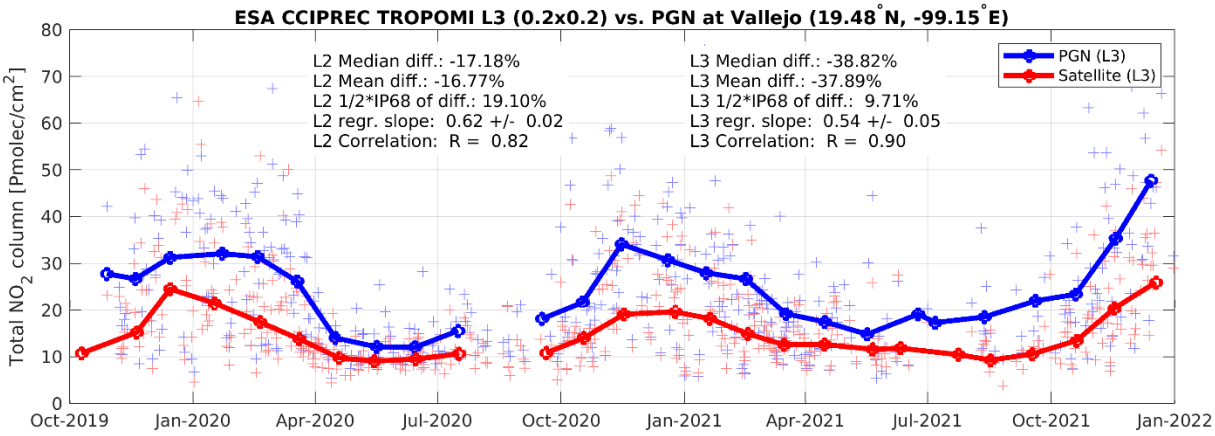


Figure 5-20: Time series of Level-3 and Level-2 TROPOMI total NO_2 column measurements (red solid lines and "+" markers respectively) and the corresponding PGN data (in blue), over the highly polluted site of Vallejo in Mexico City. The various quality indicators confirm the expected changes from Level-2 to Level-3: better correlation, lower dispersion, but also more pronounced underestimation due to increased spatial smearing of localized pollution.

An interesting intermediate case is visualized in Figure 5-21: at Charles City, Virginia, winter-time tropospheric pollution compensates winter-time stratospheric NO_2 lows, and stratospheric highs in summer compensate the cleaner troposphere, yielding a more-or-less flat total column time series. The consequence is that the loss of small temporal detail in the Level-3 monthly averages reduces the satellite-ground correlation w.r.t. the correlation seen in the Level-2 comparisons. Still, the dispersion in the comparisons is much lower at Level-3, and the more pronounced negative bias holds also here.

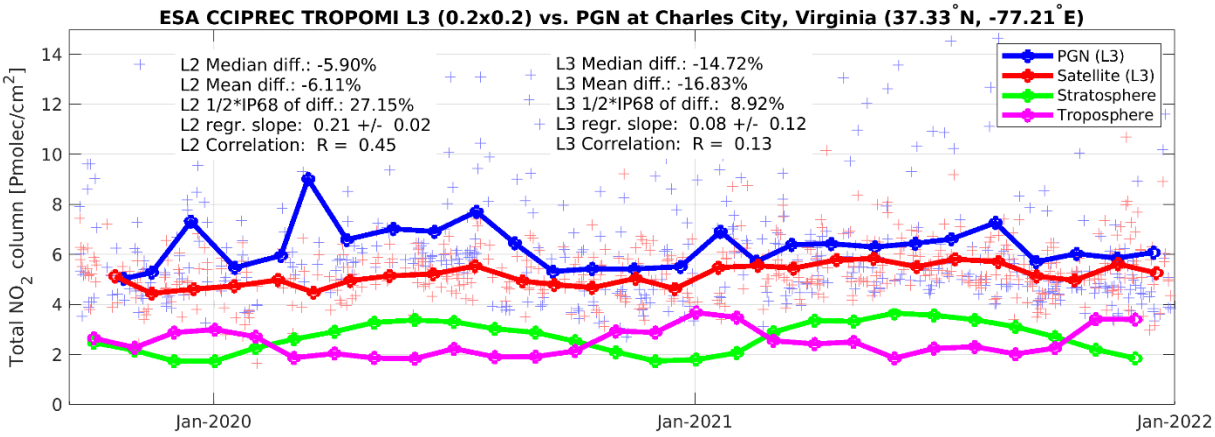


Figure 5-21: Time series of Level-3 and Level-2 TROPOMI measurements (red solid lines and "+" markers respectively) and the corresponding PGN data (in blue), over Charles City in Virginia. While the total column time series are remarkably flat, the green and magenta lines, representing the satellite stratospheric and tropospheric columns respectively, reveal that this is a consequence of winter-time tropospheric highs compensating winter-time stratospheric NO_2 lows, and vice-versa in summer.

Network-wide results for the three GOME-2 instruments and for OMI and TROPOMI are summarized in Figure 5-22 and Figure 5-23 respectively. Unfortunately, not enough Pandora measurements were made at the time of the GOME-ERS2 and SCIAMACHY-ENVISAT missions to be able to perform a meaningful, comprehensive total column validation. Another way to analyze these data, is to merge all comparison pairs per sounder into a single data set, i.e., no longer differentiated per station. This allows for a cleaner view of the dependence of bias and dispersion on vertical column (i.e., on pollution levels). This is elaborated in Figure 5-24. These summary graphs demonstrate similar behavior to the Level-2 validation results: good consistency between the various instruments, and a strong negative bias up to -50% at the most polluted sites, while the cleanest sites show no, or potentially even a slight positive bias. Note that some of the discrepancies between sensors are related to unequal temporal coverage in combination with strongly changing tropospheric NO_2 levels over the years. At very polluted sites, also differences in horizontal resolution of the underlying L2 data may cause different bias estimates. Overall, most sensors (just) satisfy the GCOS "threshold" requirements for total uncertainty at elevated NO_2 levels. The dispersion is often small enough to satisfy the "breakthrough" requirement, but the bias is too pronounced over most of the VCD range.

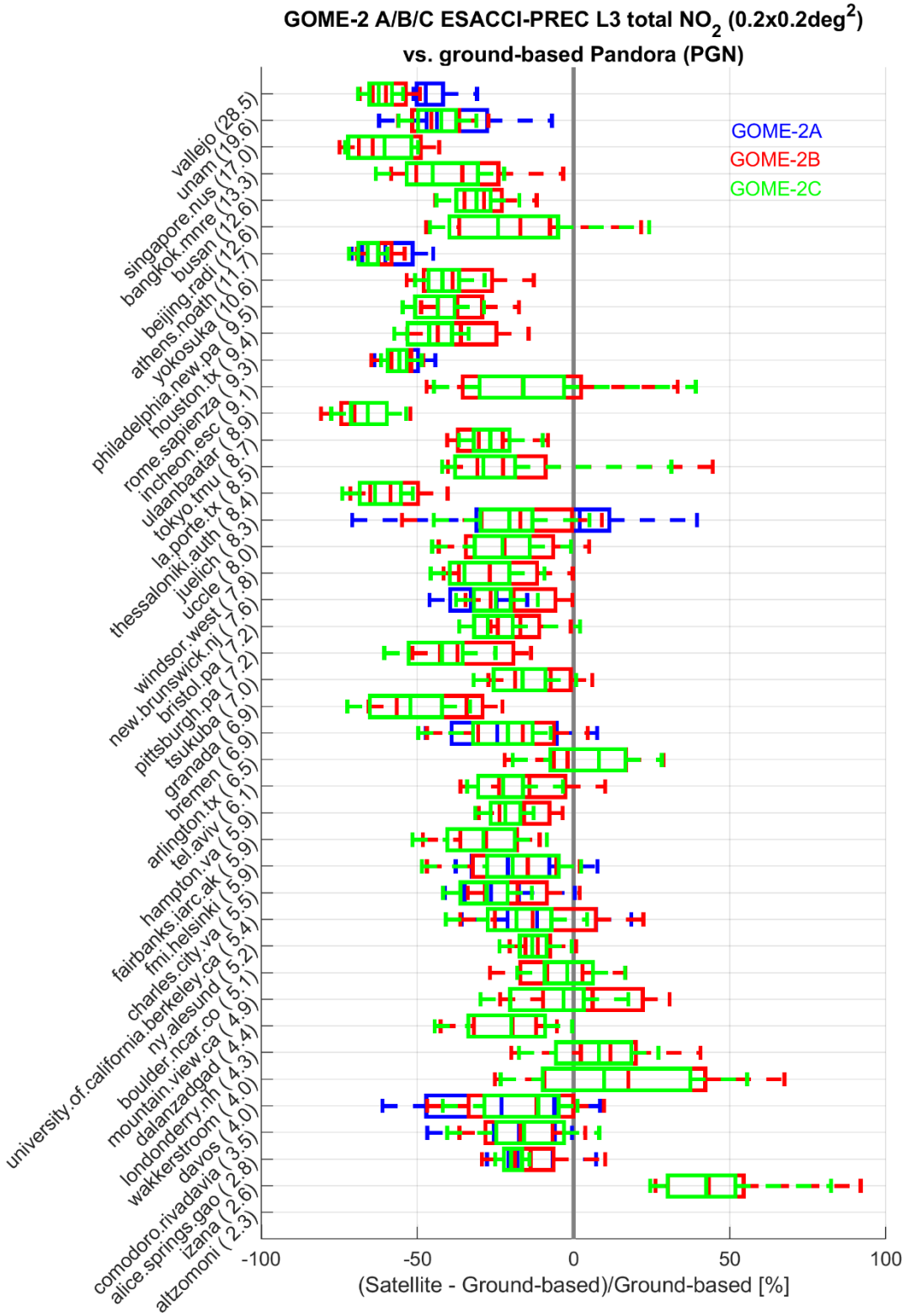


Figure 5-22: Box-whisker plots summarizing the total NO_2 comparison bias and dispersion for the three GOME-2 instruments against PGN. Stations are ordered by increasing long-term mean NO_2 levels (as measured by PGN). Note that not all sites produced data overlapping the time coverage of both sensors. In particular for GOME-2A, only a limited subset of PGN sites already produced measurements.

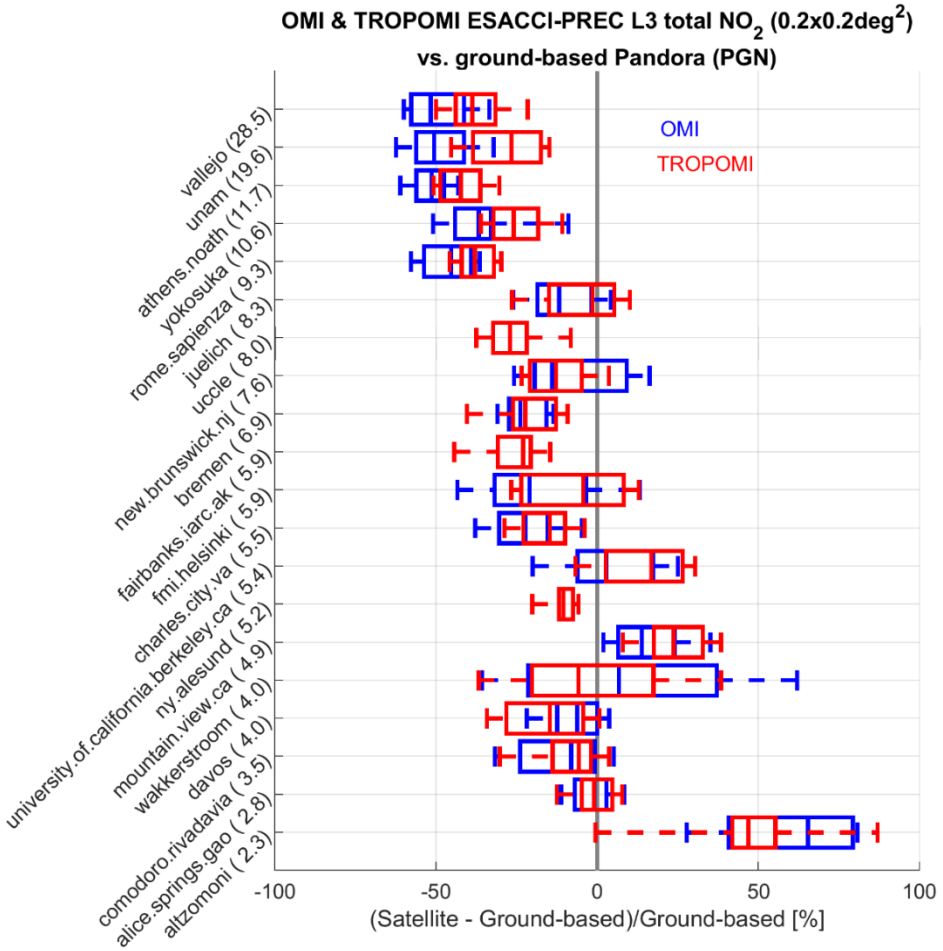


Figure 5-23: Box-whisker plots summarizing the total NO_2 comparison bias and dispersion for the Aura-OMI and Sentinel-5 Precursor TROPOMI instruments (red and blue respectively) against PGN. Stations are ordered by increasing long-term mean NO_2 levels (as measured by PGN). Note that not all sites produced data overlapping the time coverage of both sensors.

These bias levels are also summarized following 3 classes of NO_2 levels (low, medium, high) in Table 5-6, though it must be kept in mind that station-to-station scatter in the bias is very large and these numbers are to be used as coarse indications only.

Table 5-6: Bias estimates for the different products, differentiated by (PGN-measured) tropospheric NO_2 level into 3 distinct classes. For GOME and SCIAMACHY, an insufficient number of PGN measurements were available.

Product	GND < 5 Pmolec/cm²	5 < GND < 10 Pmolec/cm²	10 Pmolec/cm² < GND
GOME/ERS-2 L3 (0.2x0.2deg²)	NA	NA	NA
SCIAMACHY/ENVISAT L3 (0.2x0.2deg²)	NA	NA	NA
GOME-2/MetOp-A L3 (0.2x0.2deg²)	NA	-25%	-35%
GOME-2/MetOp-B L3 (0.2x0.2deg²)	-10%	-20%	-40%
GOME-2/MetOp-C L3 (0.2x0.2deg²)	-10%	-25%	-50%
OMI/Aura L3 (1.0x1.0deg²)	0%	-10%	-45%
TROPOMI/Sentinel-5P L3 (1.0x1.0deg²)	-0.1	-10%	-30%

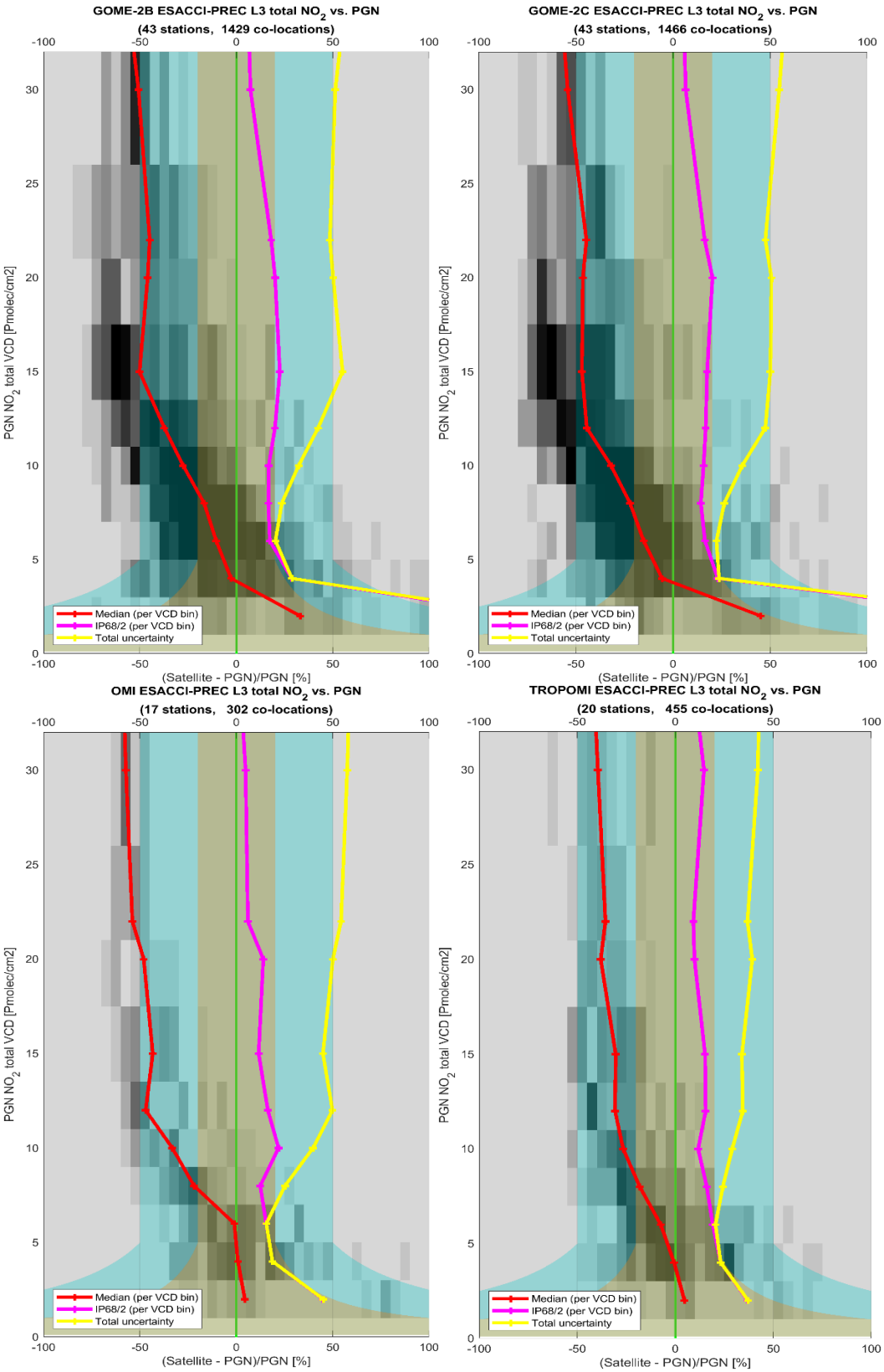


Figure 5-24: Bias (red), dispersion (magenta), and total uncertainty as a function of mean total NO_2 VCD using PGN data from all sites in a merged way, for all sensors yielding sufficient PGN co-locations to populate the matrix. Cyan and greenish zones represent GCOS threshold and breakthrough requirements respectively.

5.3.2.1. Compliance with user requirements

Table 5-7 compares the statistics for tropospheric NO2 VCD with the GCOS user requirements. It should be noted that ground-based uncertainties and comparison error also contribute to the obtained bias, dispersion and total uncertainty QI.

Table 5-7. Compliance with GCOS requirements for the tropospheric NO_2 L3 data. Bias, dispersion and total uncertainty are based on evaluation with MAX-DOAS. Regarding drift, stratospheric NO_2 VCD vs ZSL-DOAS is provided instead. Bias and dispersion are taken from Table 5-4, total uncertainty is then derived as $\sqrt{(bias^2+disp^2)}$.

Quantity	Compliance/evaluation	Remark
Horizontal resolution	0.2°x0.2° (All sensors)	T|B|G : 100|30|10 km
Temporal resolution	Monthly	T|B|G : 30 days | 1 day | 1 hour
Bias *VCD<5Pmolec cm^{-2}*	-0.3 to 0.1 Pmolec cm^{-2}	T|B|G: Max of (2.5|1|0.5 Pmolec cm^{-2}, 50%|20%|10%)
10>VCD>5Pmolec cm^{-2}	[-40%,-34%] (morning) ~-32% (afternoon)	Bias, dispersion and the total uncertainty QI are each compared with the total uncertainty requirement.
VCD>10Pmolec cm^{-2}	[-62%, -50%] (morning) [-46%,-42%] (afternoon)	
Dispersion *VCD <5Pmolec cm^{-2}*	~0.9 Pmolec cm^{-2}	Bias and dispersion increase with VCD.
10>VCD>5Pmolec cm^{-2}	38% (SCIA), 24% (S5P), ~30% (others)	
VCD>10Pmolec cm^{-2}	[22%,25%]	
Total uncertainty *VCD <5Pmolec cm^{-2}*	~0.9 Pmolec cm^{-2}	
10>VCD>5Pmolec cm^{-2}	54% (SCIA) [40%,50%] (others)	
VCD>10Pmolec cm^{-2}	47% (S5P) [51%, 67%] (others)	
Stability (**strat NO2**)	OMI, GOME-2B: Better than 5% at clean sites GOME2C, S5P: too short GOME, GOME2A, SCIA: larger drifts at sites, but no clear pattern.	T|B|G: max of (1 Pmolec cm^{-2}/decade, 20%|8%|4%) Trop NO2 drift was not assessed, we provide here strat NO2 drift as indication of instrument stability.

Color code:

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

6. Validation of HCHO Data Products

6.1. Scope and generalities

L2 products

HCHO L2 version 1 validation results were previously reported for TROPOMI operational product [RD47:Vigouroux2020][RD48:DeSmedt2021][RD38:ROCVR][RD50:Oomen2024]; QA4ECV OMI [RD49:Mueller2024], and GOME-2(A,B,C) AC SAF RPRO L2 GDP 4.8 [RD46:ACSAF-VALHCHO]. Regarding GOME-2(A,B,C), the main findings with respect to MAX-DOAS and FTIR instruments are summarized in [RD36:PVIRv1]. For OMI and TROPOMI the coherence with the L3 v1 data is demonstrated in [RD36:PVIRv1].

The L2 version 2 products are currently not released publicly. Their validation is outside scope of the current document.

L3 products

We show in this report the validation of the final version (v2) of the L3 datasets for all the instruments involved in the project (GOME, SCIAMACHY, GOME-2 A/B/C, OMI, and TROPOMI). The L3 datasets are validated using the FTIR network [RD32:Vigouroux2018] and MAX-DOAS data. Only MAX-DOAS products that include profile retrievals and averaging kernels (such that smoothing can be applied) are considered here.

For OMI and TROPOMI, both daily and monthly L3 products are provided and therefore validated. For the morning sensors, only monthly files are provided and validated.

Another complexity with the HCHO products is that if we want to check the consistency between OMI and TROPOMI, it is advised [RD48: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, for the sake of conciseness we limit ourselves in this document to the validation of “clear-sky” products (without the cloud correction applied) for all products for consistency.

6.2. Validation methodology

L2 validation

See [RD36:PVIRv1] for both the methodology and the results of the L2 v1 product. Validation of L2 v2 is outside scope of the current document.

L3 validation

Regarding the L3 validation, the following steps are applied:

FRM data selection

We excluded the FTIR station Altzomoni, because it seems to be an outlier for most of the validation exercises.
Likewise, MAX-DOAS stations Vallejo, Unam/Mexico city and Cuautitlan were not included, as they were an outlier in all validation exercises. For satellite products with older historical data (SCIAMACHY, OMI, GOME2A, GOME2B) the older bePRO processing was selected for MAX-DOAS Uccle and Xianghe as it has better temporal overlap. For the other satellite products (GOME2C, TROPOMI), the newer FRM4DOAS processing was selected.

Satellite pixel filtering
For all satellite products, we only keep the L3 data with the quality flag QA=1.

Co-location
- Space: The collocation is made with the satellite pixel containing the station.
- Time: all FTIR data within the same day (month) that are collocated to a valid satellite daily (monthly) pixel are averaged and compared to the satellite pixel value. We did test taking into account the local time of the satellite overpasses and averaging only FTIR data within ±3 hours of this satellite time (provided in the L3 v2 products), but it degraded all metrics (bias, correlation, median absolute deviation), so we finally use all data within the day (month).
- For MAX-DOAS, the same steps were followed, except that the ±3 hours window around satellite time was considered. For MAX-DOAS, this improved the bias for SCIAMACHY, GOME2A and -B compared to the full day mean.

The fact that the ±3 hours overpass window degraded the comparisons with FTIR but improved the comparisons with MAX-DOAS can be understood as follows: there are only a few FTIR measurements per day, therefore applying the small temporal window removes too much data. MAX-DOAS on the other hand measure at higher frequency. Applying the temporal window removes data that is temporally less representative but still leaves sufficient data for analysis.

Vertical harmonization
- The satellite averaging kernels are provided in the L3 files of all v2 products. So, the formalism of [RD57:Rodgers2003] is applied (with smoothing and a priori correction; see also [RD47:Vigouroux2020]).
- 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.
- The regridding/smoothing process is done using each FTIR or MAX-DOAS observation before daily or monthly averaging.

Drift estimation methodology: see section 4.4.

6.3. Validation results

6.3.1. Level-2 products

Validation results for version 1 of the L2 products (which are publicly available products) were provided for GOME-2 and summarized for other products in [RD36:PVIRv1]. Validation of L2 version 2 is considered out of scope.

6.3.2. Level-3 products

6.3.2.1. Product specification and file content checks

We noted the following:

- The format of the data files is largely compatible with the HARP conventions. We noted however that the vertical dimension is not called 'vertical' as required by HARP, but 'layer'.
- NO_2 and HCHO products can be relevant for the same users (e.g., study of NO_2/HCHO concentration ratio). Therefore, it would be good if the resolution of the NO_2 and HCHO products were aligned, which is currently not the case. E.g., the S5P NO2 monthly product has a spatial resolution 0.2°x0.2°, while the S5P HCHO monthly product has resolution 0.125°x0.125°.

6.3.2.2. Bias, dispersion, and uncertainties

Bias

Comparison with FTIR

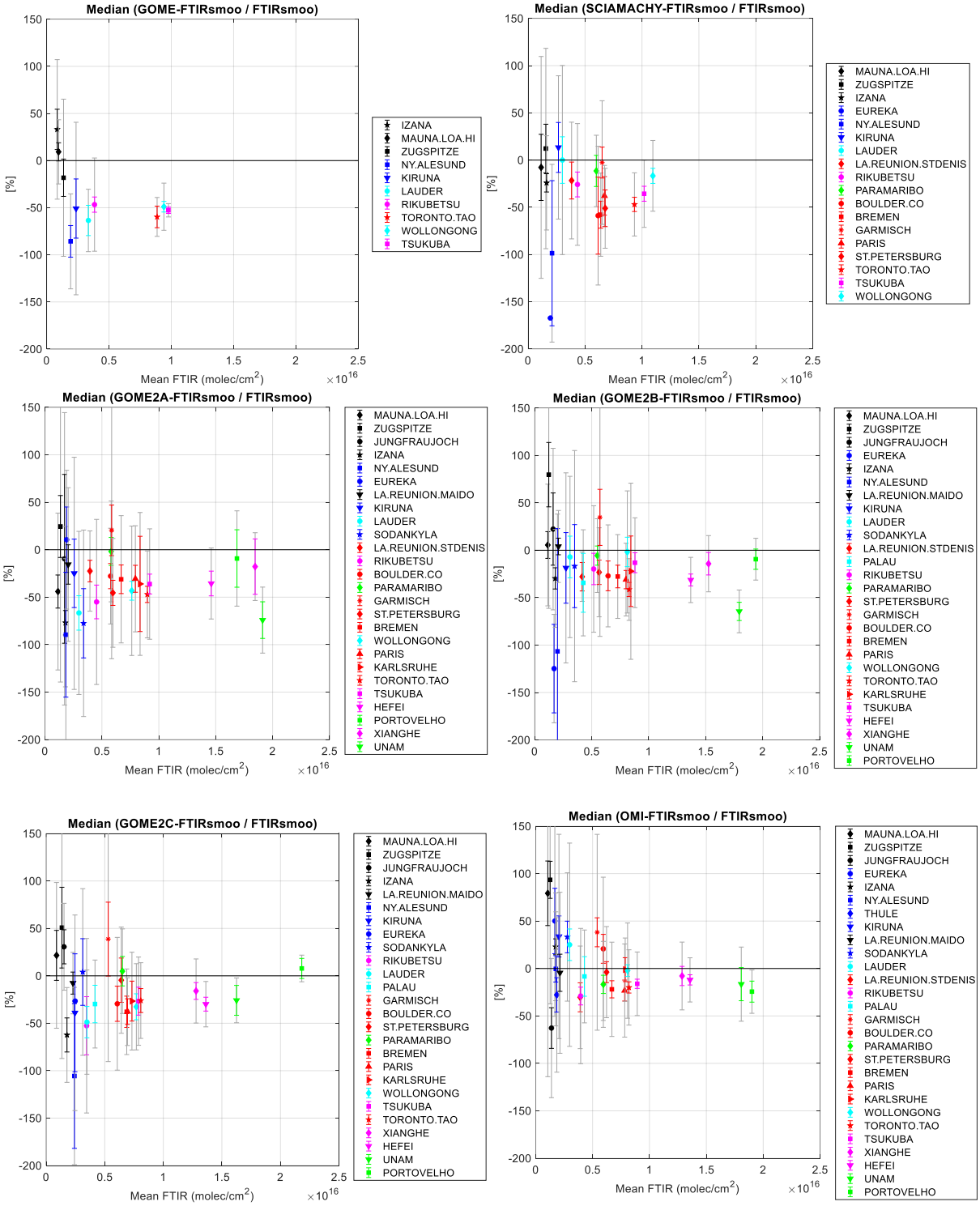
The bias of all v2 products versus FTIR has been improved compared to the validation of previous versions (see [RD36:PVIRv1]):

- We still see a positive constant bias (offset) for all products (Table 6-1: intercept of the regression), but smaller than for previous versions. It is between 0.18 and 0.66 Pmolec/cm^2 for morning sensors, and between 0.69 and 1.01 Pmolec/cm^2 for afternoon sensors.
- The proportional bias (slope of the regression) has improved for all products, ranging from 0.60 to 0.67 for morning sensors, except GOME. There is high consistency between GOME2B and GOME2C biases. GOME is showing a larger bias (slope=0.37). The monthly L3 products for afternoon sensors are also very consistent with slopes of 0.75 and 0.73 for OMI and TROPOMI, respectively. The TROPOMI daily L3 product shows a very similar regression than the L3 monthly product. The OMI daily L3 product is less consistent (slope=0.82 and the highest intercept of 1.10 Pmolec/cm2). Note that the correlation is also the weakest for that product.

The biases (median of the relative differences between all products and FTIR) at individual stations are shown in Figure 6-1. The biases of the v2 products are still negative for polluted sites, but lower than for previous versions.

Comparison with MAX-DOAS

The bias of the v2 products versus MAX-DOAS was investigated as well (see Table 6-2). Note that there are less cases with clean conditions, as opposed to the comparisons with FTIR. In polluted conditions, the bias is mostly negative in the polluted case (exception GOME2C) for the direct comparison, but this bias gets reduced or stays similar when applying the smoothing and is mostly less than 10%. For all sensors, it is of the order of 1PMC or less.



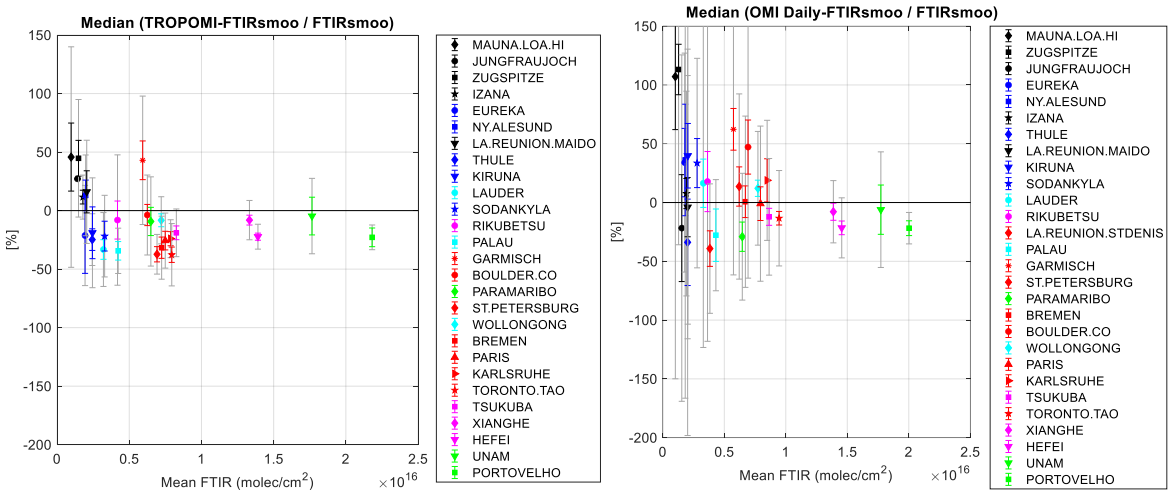


Figure 6-1: Biases for monthly L3 products "clear-sky" (i.e., no cloud correction applied) GOME, SCIAMACHY, GOME2A/B/C, OMI and TROPOMI versus FTIR, after smoothing the FTIR profile with the satellite averaging kernel. The biases for daily L3 OMI product is also shown. We do not show Daily TROPOMI biases, since they are similar to the monthly L3 ones. The grey bars are the dispersion of the differences, and the colored bars are the errors on the median (=2*dispersion/sqrt(nb coincidences)).

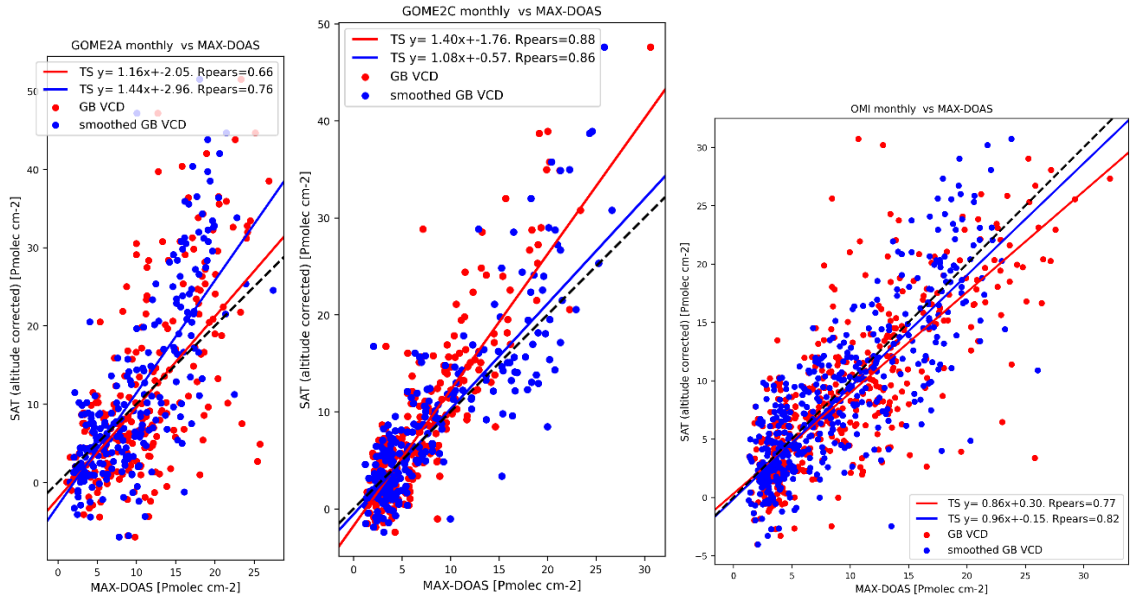


Figure 6-2. Correlation plots of GOME2A, GOME2C and OMI L3 monthly vs MAX-DOAS, without and with smoothing. Theil-Sen regression and Pearson-R are indicated as well.

Table 6-1: Overview statistics of the HCHO L3 files vs FTIR ground-based data. For bias and dispersion, the statistics are separated by pollution level: Clean<2.5E15 molec/cm^2; Polluted: >8E15molec/cm^2). PMC=E15 molec cm^{-2}.

L3 HCHO vs FTIR	Bias (median)	Dispersion (1/2IP68)	Total uncertainty √(med^2+1/2IP68^2)	Theil-Sen Regression Pearson correlation
GOME Monthly	Clean: -0.2PMC; -19.5%	Clean: 1.1PMC; 86.8%	Clean: 1.1PMC; 89.0%	y = 0.37x+0.53PMC R = 0.75

	Polluted: -6.3PMC; -55.6%	Polluted: 3.5PMC; 22.1%	Polluted: 7.3PMC; 59.8% χ^2_{red}=1.07 (random only) χ^2_{red}= 0.62 (total)	
SCIAMACHY Monthly	Clean: -0.3PMC; -14.8% Polluted: -3.7PMC; -32.0%	Clean: 1.1PMC; 88.7% Polluted: 4.1PMC; 34.9%	Clean: 1.2PMC; 90.0% Polluted: 5.5PMC; 47.4% χ^2_{red}=2.52 (random only) χ^2_{red}= 0.89 (total)	y = 0.60x+0.62PMC R = 0.63
GOME2A Monthly	Clean: -0.6PMC; -43.9% Polluted: -4.0PMC; -35.3%	Clean: 1.8PMC; 109.7% Polluted: 4.9PMC; 41.6%	Clean: 1.9PMC; 118.1% Polluted: 6.3PMC; 54.6% χ^2_{red}=0.84 (random only) χ^2_{red}=0.74 (total)	y = 0.60x+0.18PMC R = 0.56
GOME2B Monthly	Clean: -0.1PMC; -7.2% Polluted: -3.4PMC; -29.3%	Clean: 1.2PMC; 82.9% Polluted: 4.6PMC; 36.6%	Clean: 1.2PMC; 83.3% Polluted: 5.7PMC; 46.8% χ^2_{red}=1.23 (random only) χ^2_{red}=0.97 (total)	y = 0.67x+0.66PMC R = 0.69
GOME2C Monthly	Clean: -0.2PMC; -9.7% Polluted: -2.9PMC; -26.0%	Clean: 1.6PMC; 113.3% Polluted: 3.9PMC; 31.9%	Clean: 1.6PMC; 113.7% Polluted: 4.9PMC; 41.2% χ^2_{red}=2.04 (random only) χ^2_{red}=1.26 (total)	y = 0.67x+0.51PMC R = 0.78
OMI Monthly	Clean: 0.2PMC; +12.6% Polluted: -2.5; -20.2%	Clean: 1.7PMC; 113.0% Polluted: 4.1PMC; 34.1%	Clean: 1.7PMC; 113.7% Polluted: 4.8PMC; 39.6% χ^2_{red}=1.97 (random only) χ^2_{red}=0.59 (total)	y = 0.75x+0.85MC R = 0.76
TROPOMI Monthly	Clean: 0.3PMC; +18.2% Polluted: -2.3PMC; -19.9%	Clean: 0.8PMC; 60.3% Polluted: 3.2PMC; 24.2%	Clean: 0.9PMC; 63.0% Polluted: 4.0PMC; 31.3%	y = 0.73x+0.69PMC R = 0.90

			χ^2_{red}=7.35 (random only) χ^2_{red}=0.46 (total)	
OMI Daily	Clean: 0.7PMC; +46.9% Polluted: -1.6PMC; -11.6%	Clean: 4.4PMC; 354.8% Polluted: 8.6PMC; 63.4%	Clean: 4.4PMC; 357.9% Polluted: 8.7PMC; 64.4% χ^2_{red}=1.59 (random only) χ^2_{red}=1.14 (total)	y = 0.82x+1.01PMC R = 0.53
TROPOMI Daily	Clean: 0.4PMC; +26.1% Polluted: -2.7PMC; -19.1%	Clean: 1.5PMC; 108.4% Polluted: 3.8PMC; 25.1%	Clean: 1.6PMC; 111.5% Polluted: 4.7PMC; 31.5% χ^2_{red}=2.61 (random only) χ^2_{red}=0.70 (total)	y = 0.76x+0.71PMC R = 0.87

Table 6-2. Overview statistics of the HCHO L3 files vs MAX-DOAS ground-based data. For bias and dispersion, the statistics are separated by pollution level: Clean<2.5E15 molec/cm2; Polluted: >8E15molec/cm2). PMC=E15 molec cm-2. The second entry per cell is for the smoothed MAX-DOAS profile. 'Daily per month' means that monthly averages are taken for the daily product comparisons, to make it more intercomparable with the monthly product.

L3 HCHO vs MAXDOAS	Bias (median)	Dispersion (1/2IP68)	Total uncertainty √(med²+1/2IP68²)	Theil-Sen Regression Pearson correlation
SCIAMACHY Monthly Direct *Smoothed*	Clean: No data Polluted: -1.4PMC, -15% *-1.3PMC, -7%*	Clean: No data Polluted: 5.6PMC,45% *5.8PMC,45%*	Clean: No data Polluted: 5.8PMC,48% *6.0PMC,45%*	y=1.25x-4.56PMC R=0.76 *y=1.55x-6.54PMC* *R=0.83*
GOME2A Monthly Direct *Smoothed*	Clean: low data Polluted: -1.1PMC,-10% *0.1PMC,1%*	Clean: low data Polluted: 8.0PMC,83% *7.3PMC,78%*	Clean: low data Polluted: 8.0PMC,83% *7.3PMC,78%*	y=1.16x+-2.05PMC R=0.66 *y=1.44x+-2.96PMC* *R=0.76*
GOME2B Monthly Direct *Smoothed*	Clean: 0.6PMC,26% *-0.4PMC,-12%* Polluted: -0.7PMC,-8% *0.3PMC,3%*	Clean: 2.7PMC,126% *2.8PMC,102%* Polluted: 5.9PMC,67% *4.7PMC,62%*	Clean: 2.7PMC,129% *2.8PMC,103%* Polluted: 5.9PMC,67% *4.8PMC,62%*	y=0.93x+-0.01PMC R=0.68 *y=1.07x+-0.39PMC* *R=0.80*
GOME2C Monthly Direct *Smoothed*	Clean: -0.3PMC,-21% *-0.8PMC,-38%* Polluted: 1.2PMC,15% *0.0PMC,0%*	Clean: 1.6PMC,72% *1.7PMC,67%* Polluted: 4.1PMC,58% *3.7PMC,51%*	Clean: 1.7PMC,76% *1.9PMC,77%* Polluted: 4.2PMC,60% *3.7PMC,51%*	y=1.40x+-1.76PMC R=0.88 *y=1.08x+-0.57PMC* *R=0.86*

OMI Monthly Direct *Smoothed*	Clean: 0.4PMC,18% *0.1PMC,3%* Polluted: -1.0PMC,-10% *-0.5PMC,-6%*	Clean: 2.6PMC,137% *2.8PMC,150%* Polluted: 3.6PMC,45% *3.6PMC,48%*	Clean: 2.7PMC,139% *2.8PMC,150%* Polluted: 3.8PMC,46% *3.6PMC,48%*	y=0.86x+0.30PMC R=0.77 *y=0.96x+-0.15PMC R=0.82*
TROPOMI Monthly Direct *Smoothed*	Clean: 0.0PMC,1% *0.1PMC,4%* Polluted: -0.3PMC,-5% *-0.3PMC,-5%*	Clean: 0.8PMC,49% *0.8PMC,47%* Polluted: 1.9PMC,32% *2.1PMC,30%*	Clean: 0.8PMC,49% *0.8PMC,47%* Polluted: 1.9PMC,32% *2.1PMC,30%*	y=1.03x+-0.43PMC R=0.91 *y=0.89x+0.34PMC R=0.91*
OMI Daily Direct *Smoothed*	Clean: 0.2PMC,9% *-0.3PMC,-12%* Polluted: -0.8PMC,-9% *0.0PMC,0%*	Clean: 7.3PMC,382% *7.0PMC,343%* Polluted: 8.1PMC,97% *8.1PMC,102%*	Clean: 7.3PMC,383% *7.0PMC,343%* Polluted: 8.2PMC,98% *8.2PMC,102%*	y=0.78x+1.34PMC R=0.54 *y=0.90x+0.76PMC R=0.55*
OMI daily, per month Direct *Smoothed*	Clean: 1.5PMC,92% *1.1PMC,67%* Polluted: -0.6PMC,-7% *-0.0PMC,-0%*	Clean: 3.0PMC,206% *3.4PMC,200%* Polluted: 4.0PMC,51% *3.7PMC,50%*	Clean: 3.4PMC,225% *3.5PMC,211%* Polluted: 4.1PMC,52% *3.7PMC,50%*	y=0.79x+1.43PMC R=0.77 *y=0.87x+1.18PMC R=0.78*
TROPOMI daily Direct *Smoothed*	Clean: 0.2PMC,9% *-0.0PMC,-1%* Polluted: -0.6PMC,-9% *-0.8PMC,-13%*	Clean: 2.3PMC,136% *2.4PMC,126%* Polluted: 3.2PMC,49% *3.2PMC,48%*	Clean: 2.3PMC,136% *2.4PMC,126%* Polluted: 3.3PMC,50% *3.3PMC,49%*	y=0.88x+0.22PMC R=0.82 *y=0.84x+0.27PMC R=0.82*
TROPOMI daily, per month Direct *Smoothed*	Clean: -0.1PMC,-5% *-0.1PMC,-12%* Polluted: -0.5PMC,-7% *-0.6PMC,-9%*	Clean: 0.7PMC,39% *0.8PMC,40%* Polluted: 1.9PMC,27% *2.2PMC,29%*	Clean: 0.7PMC,39% *0.8PMC,42%* Polluted: 1.9PMC,28% *2.3PMC,30%*	y=0.93x+0.00PMC R=0.92 *y=0.85x+0.33PMC R=0.91*

Dispersion

Comparison with FTIR

The 1/2IP68 dispersion values for the comparison with FTIR are provided in Table 6-1. They range from 1.1 to 1.8Pmolec/cm^2 for clean conditions (HCHO < 2.5 Pmolec/cm^2) and from 3.2 to 4.9 Pmolec/cm^2 for polluted conditions (HCHO> 8 Pmolec/cm^2). The only exception is the OMI Daily product which shows a larger dispersion (4.4 and 8.6 Pmolec/cm^2 for clean and

polluted condition, respectively). The TROPOMI Daily L3 product is still good, with a dispersion similar to OMI monthly product.

The Pearson correlation coefficients range from 0.53 (OMI Daily) or 0.56 (GOME2A) to 0.9 for TROPOMI. We also see the good improvement of the correlation (and dispersion) from GOME2A to GOME2C.

Comparison with MAX-DOAS
The 1/2IP68 dispersion values for the comparison with MAX-DOAS are provided in Table 6-2. For polluted conditions, they range between 2 PMC (30%) for TROPOMI monthly and 8 PMC (80-100%) for OMI daily and GOME2A monthly. The dispersion improves with smoothing for GOME2A,-B,-C and stays similar otherwise. The dispersion is always the dominant term in the total uncertainty QI $\sqrt{(med^2+1/2IP68^2)}$.
The correlation (Pearson-R) improves with smoothing (SCIAMACHY, GOME2A,B,C) or stays similar and varies between 0.55 (OMI daily) and 0.91 (TROPOMI monthly). As for the FTIR comparison, we see the improvement of GOME2C over GOME2A and B for correlation and dispersion.

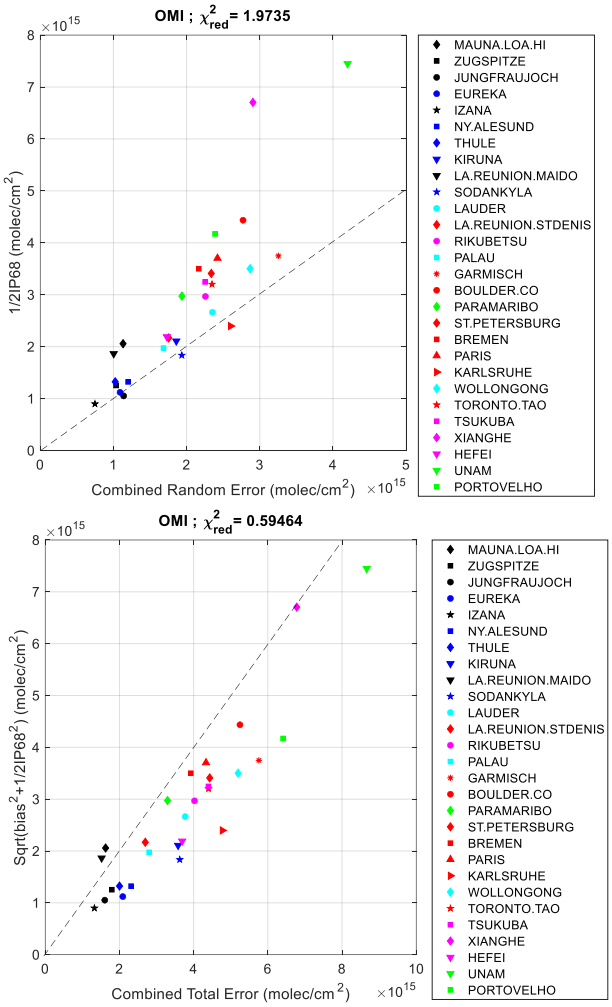
The OMI and TROPOMI daily products, with comparisons averaged per month, have reduced sampling mismatch compared to the OMI and TROPOMI monthly products, since MAX-DOAS data for days where the satellite does not contribute are removed in the first case. Notwithstanding this, the quality indicators have not improved.

Intercomparison of validation results vs FTIR and vs MAX-DOAS
The satellite validation results when using FTIR (Table 6-1) and when using MAX-DOAS (Table 6-2) exhibit differences.

- The satellite bias in polluted conditions is more negative when comparing with FTIR. This is consistent with [RD51:Pinardi2025], where it was found that MAX-DOAS retrieved lower HCHO VCD than FTIR, and this is related to their different vertical sensitivity.
- The dispersion is higher when comparing with MAX-DOAS. Among the likely reasons are (1) the higher sensitivity of MAX-DOAS close to the surface (as compared to FTIR and satellite) and (2) the lower degree of processing harmonization across the different MAX-DOAS datasets leading to higher inter-station differences.

Uncertainty
As described in Section 4.3, we validate the uncertainties provided in the L3 files. Both random and systematic uncertainties are provided. We therefore first analyse the random uncertainty by comparing it to the 1/2IP68 dispersion (see Figure 6-3a as an illustration) and providing the χ^2_{red} of this comparison (Sect. 4.3): except for GOME2A where χ^2_{red} is 0.84 (=small overestimation of the satellite random uncertainty) and GOME where χ^2_{red}=1.07 (=good estimation of the random uncertainty), the random uncertainty are usually underestimated, but with a reasonable χ^2_{red} between 1.23 and 2.61. Note also that the collocation uncertainty is not taken into account and is probably having an impact on the dispersion: Figure 6-3 shows a larger deviation from the 1:1 line for more polluted sites.

The only larger χ^2_{red} value (7.35) is for the TROPOMI monthly L3 product. When calculating the χ^2_{red} for total uncertainty (comparing it to the square root of the squared sum of bias and dispersion), we observe that is usually below 1, pointing to an overestimation of the systematic uncertainty (since the random one has been identified to be underestimated). Moving from Daily to Monthly L3 files, we observe that the underestimation of random uncertainties increases while the overestimation of the total (=systematic then) also increases. We can conclude that some uncertainties might have been propagated in the averaging process as being systematic while they actually average out.



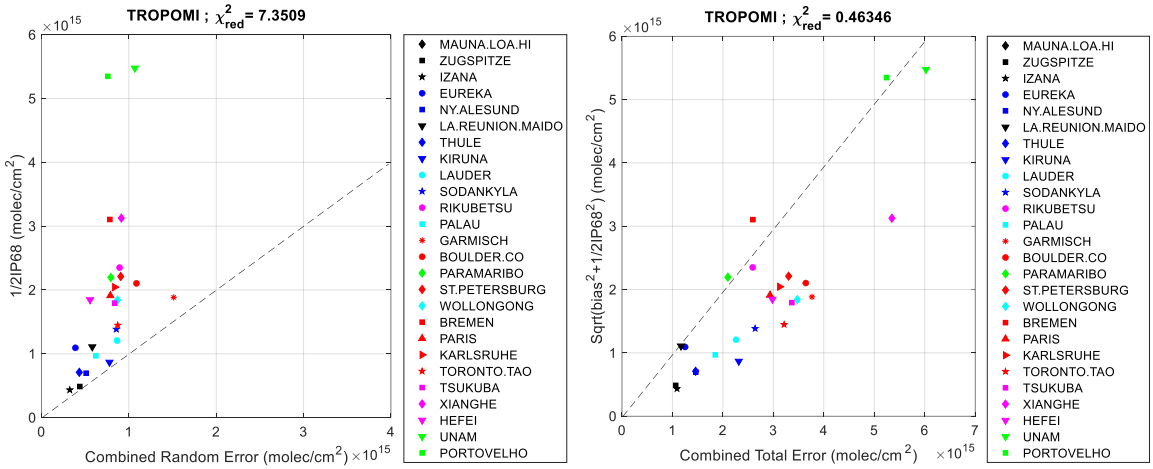
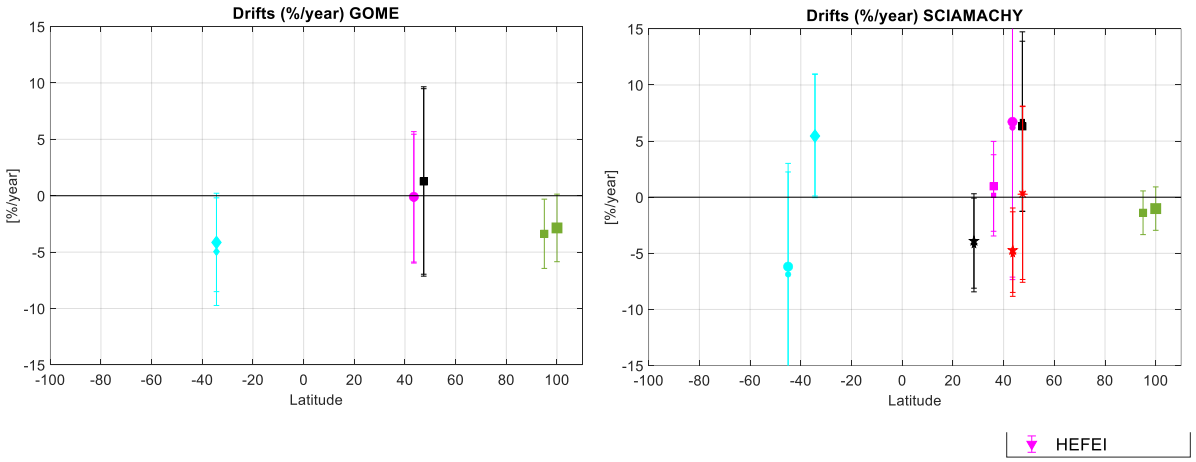


Figure 6-3: a) and b) Validation Monthly OMI L3 random and total uncertainty, respectively; c) and d) Validation Monthly TROPOMI L3 random and total uncertainty, respectively.

6.3.2.3. Stability

We apply the MLR model described above (Sect. 4.4) to the time-series of absolute differences SAT-FTIR at all sites. The trends (drifts) are given in Figure 6-4 in percent by dividing the absolute trends by the mean of FTIR values at each site. Only time-series with more than 60 coincidences are used for drift calculations (so no drift evaluation yet for GOME2C).

- The drifts at individual stations are usually within 50%/decade for all sensors.
- The high and significant negative drifts observed in the Southern Hemisphere (Lauder and Wollongong), in GOME2 A and GOME2 B for v1 (see [RD36:PVIRv1]) is no longer present in the v2 products.
- Although at individual stations a few significant drifts are observed, the overall drift (taking the mean for each month of the differences at all stations) is non-significant and below 10%/decade, except for GOME (smaller number of stations have data for the GOME time-period, see Figure 6-4).



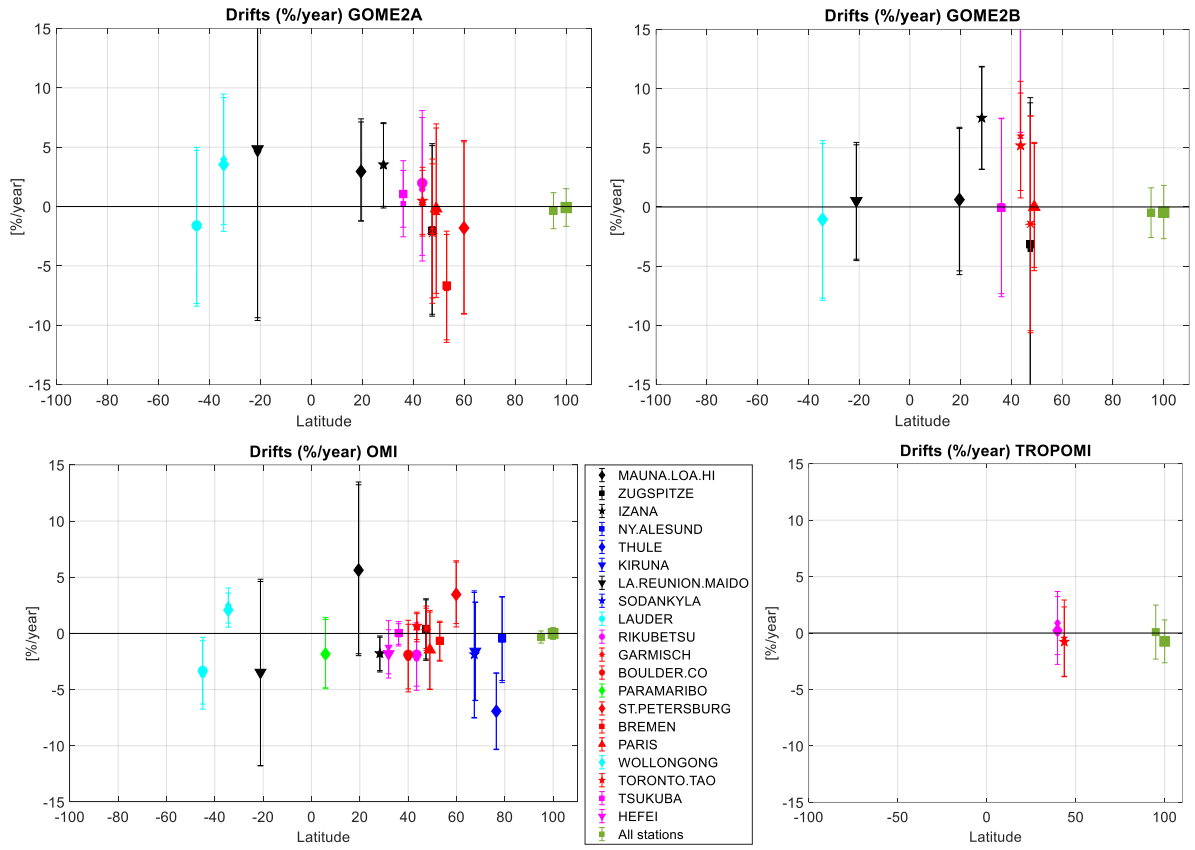


Figure 6-4: Drift (%/year) of the monthly L3 products (trend regression including seasonal cycle of the SAT–FTIR absolute differences; divided by mean of FTIR to obtain drift in %) (SAT= GOME, SCIAMACHY, GOME2-A/B/C, OMI, TROPOMI). The error bars are equivalent to 2-sigma uncertainty on the drifts. The larger markers are for the comparisons taking AK into account. Similar drifts are obtained using L3 Daily for OMI and TROPOMI (not shown).

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

In the current work, no validation of L2 v2 is performed. However, it was demonstrated in [RD36:PVIRv1], for v1 of the data, that the validation results of L2 and L3 are very consistent as can be seen in the GOME2-A and -B L2 and L3 validation results in that work. Also, the validation results of OMI and TROPOMI L3 data in [RD36:PVIRv1] are very consistent with past studies on L2 ([RD49:Mueller2024], [RD47:Vigouroux2020], respectively).

6.3.2.5. Compliance with user requirements

Table 6-3. 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 6-1.	Vs FTIR: Positive intercept and slope lower than 1 (negative columns for polluted conditions) for all sensors. Vs MAX-DOAS, after smoothing: Slopes between 0.8 and 1.1 for most sensors, and bias close to zero in polluted conditions.
Dispersion	Not specified	Check summary in Table 6-1	
Total uncertainty (**1-sigma**)	Absolute: 20e15|8e15|4e15 *molec/cm^2 Relative: 50%|20%|10%	**Clean (< 2.5E15 molec/cm^2)**: Vs FTIR: Monthly L3: 1.1-1.8E15 molec.cm^{-2} Daily L3 (afternoon sensors): 1.6-4.4E15 molec.cm^{-2} Vs MAX-DOAS: Monthly L3: 1-3E15 molec.cm^{-2} Daily L3 (afternoon sensors): 1-7E15 molec cm^{-2} **Polluted (> 8E15 molec/cm^2)** : Vs FTIR: Monthly L3: 4.0-6.3E15 molec.cm^{-2} Daily L3 (afternoon sensors): 4.7-8.7E15 molec.cm-2 Vs MAX-DOAS: Monthly L3: 2-7E15 molec cm^{-2} Daily L3 (afternoon sensors): 3-8E15 molec.cm-2	The "goal" is reached by all sensors in absolute values for clean conditions, and the "breakthrough" is reached for all sensors in absolute values for polluted conditions. Only the Daily OMI L3 product is just above the limits in both cases. (based on FTIR and MAX-DOAS) Vs FTIR: From χ^2_{red} values in Table 6-1: Random uncertainty are usually slightly underestimated (except TROPOMI daily high underestimation). Systematic uncertainty are usually overestimated.
Stability	Absolute: 8e15|8e15|8e15 molec cm^{-2}/decad Relative: 20%|8%|4% /decade	Drifts at individual stations 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. The drifts obtained using all stations together are small (<10%/decade) and non significant.	The requirement is always achieved in absolute value. The "breakthrough" is achieved in %/decade when all stations are used together (except for GOME).

Color code:

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

7. Validation of SO_2 Data Products

7.1. Scope and generalities

The following L2 and L3 datasets were available: the SO_2 column retrieved with the COBRA algorithm, for the sensors ERS2\GOME, S5P\TROPOMI, Aura\OMI (see Table 3-3). ERS2\GOME was not validated because of lack of FRM data within scope. The product for ENVISAT\SCIAMACHY was not yet available. The default COBRA column is cloud-corrected, but the product also has the clear-sky air mass factor which allows to calculate a column with clear-sky assumption.
Due to the limited availability of FRM data (MAX-DOAS and PGN Pandora), the validation is geographically and temporally limited in scope (see section 4.6.3).

7.2. Validation methodology

L2 data. Overpasses in HARP-compatible format [RD60:HARP] were generated from the L2 SO_2 data. The specific steps are provided below.

Satellite pixel filtering. Based on quality filter recommendations from the product provider, pixels were kept only if the qa_value>50%.
Ground-based filtering. Specifically, for PGN data: only data was kept with quality flag 0 or 10.
Co-location and comparison pair handling. Only satellite pixels within 50 km (OMI) or 10 km (TROPOMI) of the FRM sensor location are kept. Only overpasses with at least 5 pixels were kept and the pixels were averaged; as a result, the median number of pixels per overpass was 10 (OMI) or 9 (TROPOMI). Ground-based data within ±1 h (OMI) or ±0.5 h (TROPOMI) of an overpass are kept and averaged.
As the number of pixels averaged per overpass is similar for the OMI and the S5P product validation, their dispersions are evaluated in this sense on an equal footing.

L3 data. For both OMI and TROPOMI, the daily product on the 0.25° x 0.25° grid was considered. The specific steps are provided below.

Satellite data filtering. Data was kept only if qa_L3>0, and if the variable eff_fract_day was not NaN.
Spatial co-location. Only the grid cell covering the FRM sensor location was kept.
Temporal co-location. Common days of satellite and ground-based data were selected. Ground-based data was averaged within ±0.5 h of the satellite eff_fract_day.

The median number of pixels included per grid cell was 1 for OMI and 24 for TROPOMI. Therefore, it can be expected that random error (contributing to the dispersion in the validation) will be more reduced for TROPOMI.

Robust and non-robust validation metrics are calculated as defined in [RD34:PVPv2]. The robust validation metrics are used for discussion and to assess the compliance with user requirements. Both cloud-corrected (the default) and clear-sky columns are investigated here.

Considered over the temporal range of OMI, the MAX-DOAS stations Xianghe and Wuxi have the highest SO_2 pollution and therefore the strongest signal. Over the S5P temporal range however, Xianghe is one of the lesser polluted sites, demonstrating the SO_2 decrease in this Chinese city over the years. It follows therefore that for the S5P validation a less strong signal is available compared to the OMI validation.

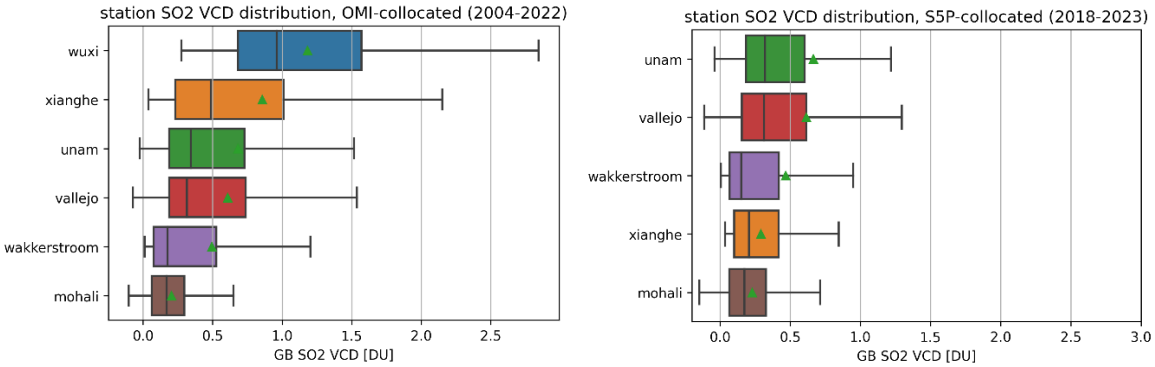


Figure 7-1. Ground-based SO_2 vertical column distribution, as collocated with OMI (left), and with S5P (right). Boxplots are ordered along mean SO_2 vertical column. Note that the Wuxi MAX-DOAS has no temporal overlap with S5P, and that Xianghe changes from a strongly polluted site within OMI temporal range, to a much less polluted site in S5P temporal range.

7.3. Validation results

7.3.1. Level-2 products

7.3.1.1. Bias and dispersion

Figure 7-2 (top) presents boxplots of the difference and relative difference per station, for cloud-corrected and clear-sky VCDs from L2 OMI, sorted by mean station SO_2 level. Overall, the bias (median difference) is negative, being ~-0.1 DU (-30% to –70%) for the four least polluted sites, but higher for the two highly polluted MAX-DOAS sites Xianghe and Wuxi (-0.4 DU and -0.8 DU, or ~-80%). The difference dispersion (IP68/2) increases with station mean pollution level, from 0.2 to 0.6 DU, but in relative terms it decreases with pollution level. There is not much difference in validation results between the cloud-corrected and the clear-sky SO_2 columns.

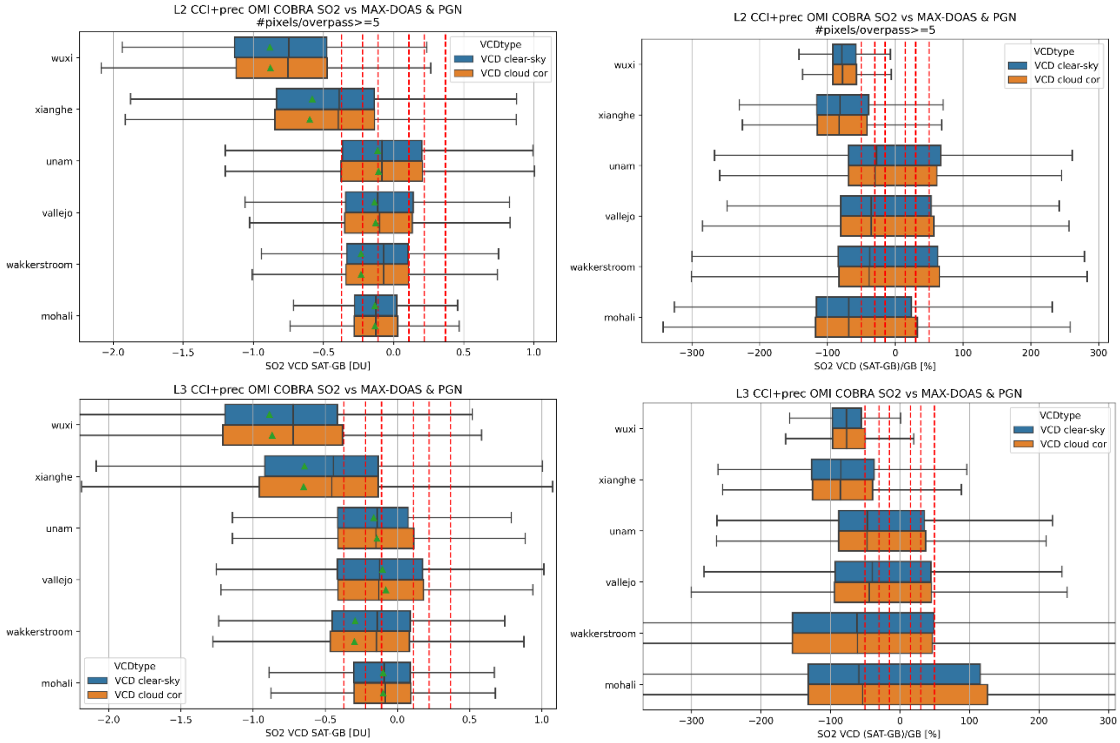


Figure 7-2. Boxplots of difference (left) and relative difference (right) for L2 CCI+p OMI (top) and L3 CCI+p (bottom), for cloud-corrected and clear-sky VCD. Green marker is the mean difference. Boxplots are ordered by mean station SO_2 level (bottom most clean). In red dashed line the total uncertainty requirements (threshold, breakthrough, goal) are indicated.

For the lower polluted stations (Mohali to Unam) the bias is of the order of the goal total uncertainty requirement (0.11 DU), but it exceeds all requirements for the higher polluted sites (Xianghe and Wuxi). The dispersion at Mohali to Vallejo is below or on the order of the 0.22 DU breakthrough requirement, but exceeds the requirements at higher polluted stations (Unam to Xianghe), except at Wuxi where it is below the 50% the threshold relative requirement. For the combined bias and dispersion, the threshold requirement is only met at Mohali and Wakkerstroom.

Figure 7-3 (top) presents the boxplots of the difference and relative difference for S5P vs PGN and MAX-DOAS. The bias is negative and in the order of -0.14 to -0.2 DU (-40% to -75%), which is within the breakthrough total uncertainty requirement of 0.22 DU. The dispersion is 0.3 to 0.4 DU, i.e., within or about equal to the threshold total uncertainty requirement (0.37 DU). The clear-sky VCD have slightly higher dispersions than the cloud-corrected ones.

While maximum bias and dispersion appear better for the S5P than for the OMI product, one should remember that the range of validation is more limited: the maximum pollution levels are much lower than for OMI. If we neglect the stations with high pollution in OMI range) (Wuxi and Xianghe), bias and dispersion are slightly higher for the S5P product.

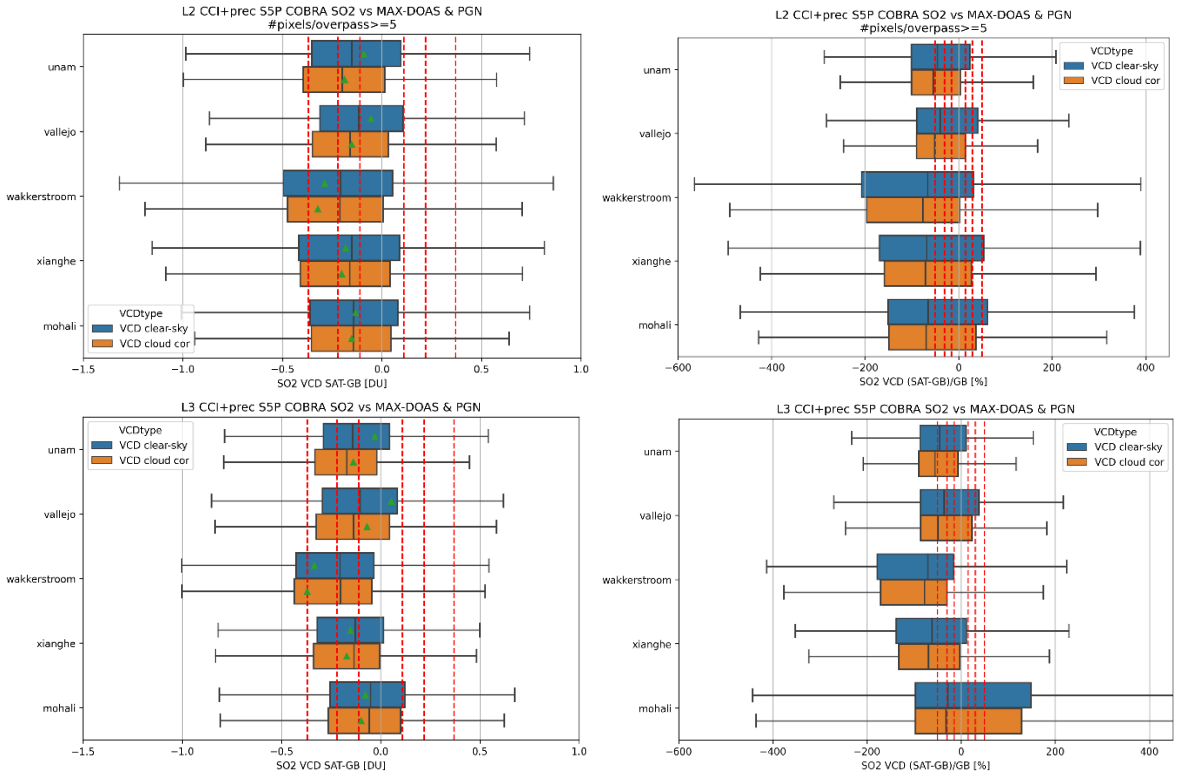


Figure 7-3. Boxplots of difference (left) and relative difference (right) for L2 CCI+p S5P (top) and L3 CCI+p S5P (bottom), for cloud-corrected and clear-sky VCD. Green marker is the mean difference. Boxplots are ordered by mean station SO_2 level (bottom most clean). In red dashed line the total uncertainty requirements (threshold, breakthrough, goal) are indicated.

Figure 7-4 presents the correlation plot of L2 OMI daily and L2 S5P daily vs ground-based data at the different stations. Regarding the OMI cloud-corrected product, per station, Theil-Sen slopes range between 0.26 and 0.4, while normed Kendall-Tau correlations are between 0.41 and 0.56. The clear-sky product has very similar regressions and correlations (not shown). Regarding the S5P cloud corrected product, the slopes range between 0.2 and 0.7, while the correlations are between 0.15 and 0.66. The clear-sky S5P product has similar correlations, but higher slopes (between 0.2 and 0.9, not shown).

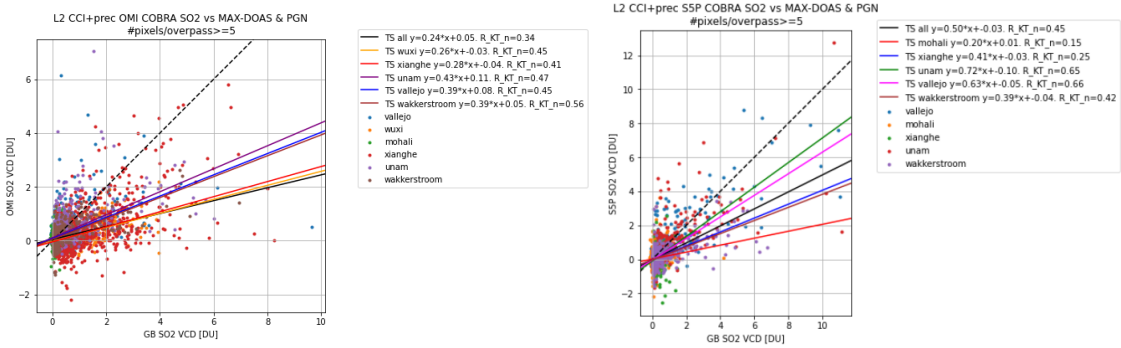


Figure 7-4. Left: L2 OMI SO_2 vs MAX-DOAS and PGN correlation plots. Theil-Sen robust regression is used, and normed Kendall-Tau correlation is indicated. Right: similar but for L2 S5P SO_2.

7.3.2. **Level-3 products**

7.3.2.1. **Product specification and file content checks**

We noted the following:

- After removing quality flagged data, still a small part of the data had NaN-filled values for the eff_fract_day variable. For the validation these data points were removed. We recommend to look into this issue.
- The top-level netCDF attributes 'datetime_start' and 'datetime_stop' had values not compatible with HARP conventions, preventing the use of some HARP tools. We recommend to fix this issue.

7.3.2.2. **Bias and dispersion**

Figure 7-2 (bottom) presents boxplots of the difference and relative difference per station, for cloud-corrected and clear-sky VCDs from L3 OMI, sorted by mean station SO_2 level. The bias (median difference) of L3 OMI is similar as for the L2 product: ~-0.1 DU for the lower polluted stations (Mohali to Unam), and ~-80% for the higher polluted sites Xianghe and Wuxi. The dispersion (1/2 IP68) is 0.3 to 0.7 DU which is larger than for the corresponding L2 product validation, i.e., often larger than the threshold requirement. This can be understood: the median pixel number per grid cell in the L3 product is only 1, while for the 50 km radius overpasses of the L2 product it is 10. Noise reduction due to pixel averaging is therefore more important in the L2 validation.

Figure 7-3 (bottom) presents boxplots of the difference and relative difference per station, for cloud-corrected and clear-sky VCDs from L3 S5P, sorted by mean station SO_2 level. The biases of L3 S5P (median difference) per station range between -0.06 and -0.2 DU and are similar as for the L2 product, except at Mohali where the bias is smaller. Dispersions are ~0.3 DU and smaller than for the L2 validation, which can be attributed to the higher number of pixels used per L3 grid cell (median number of 24 pixels) as compared to the 10 km overpasses (median number of 9 pixels) used for the L2 validation. The random error contribution in the L3 validation will therefore be smaller.

Focusing on the lower polluted stations (Mohali to Unam) for a fair comparison, the dispersion of S5P is clearly lower than OMI, which reflects the higher number of pixels per 0.25°x0.25° grid cell for the S5P product.

Figure 7-5 presents the bias, dispersion and the derived total uncertainty QI based on all combined data pairs of all stations together, in function of ground-based SO_2 column bin. The OMI bias is outside the total uncertainty threshold requirement except at the lowest SO_2 bin, while the OMI dispersion is within the requirement. The S5P bias and dispersion are (just) within the requirement, but the derived total uncertainty QI is just outside the threshold requirement, except at the lowest SO_2 bin. Bias, dispersion and total uncertainty QI are proportionally related to SO_2 column. Note that most data is confined to the lowest SO_2 bin.

The fact that the OMI dispersion per station often exceeds the threshold requirement (Figure 7-2, bottom) while it complies in the SO_2 column resolved plot (Figure 7-5) is not a contradiction: in the per-station boxplot of Figure 7-2 of e.g., Xianghe, the SO_2 columns is represented from low to high value. The bias proportional to the SO_2 column, as uncovered in Figure 7-5, therefore collapses to a dispersion contribution in Figure 7-2.

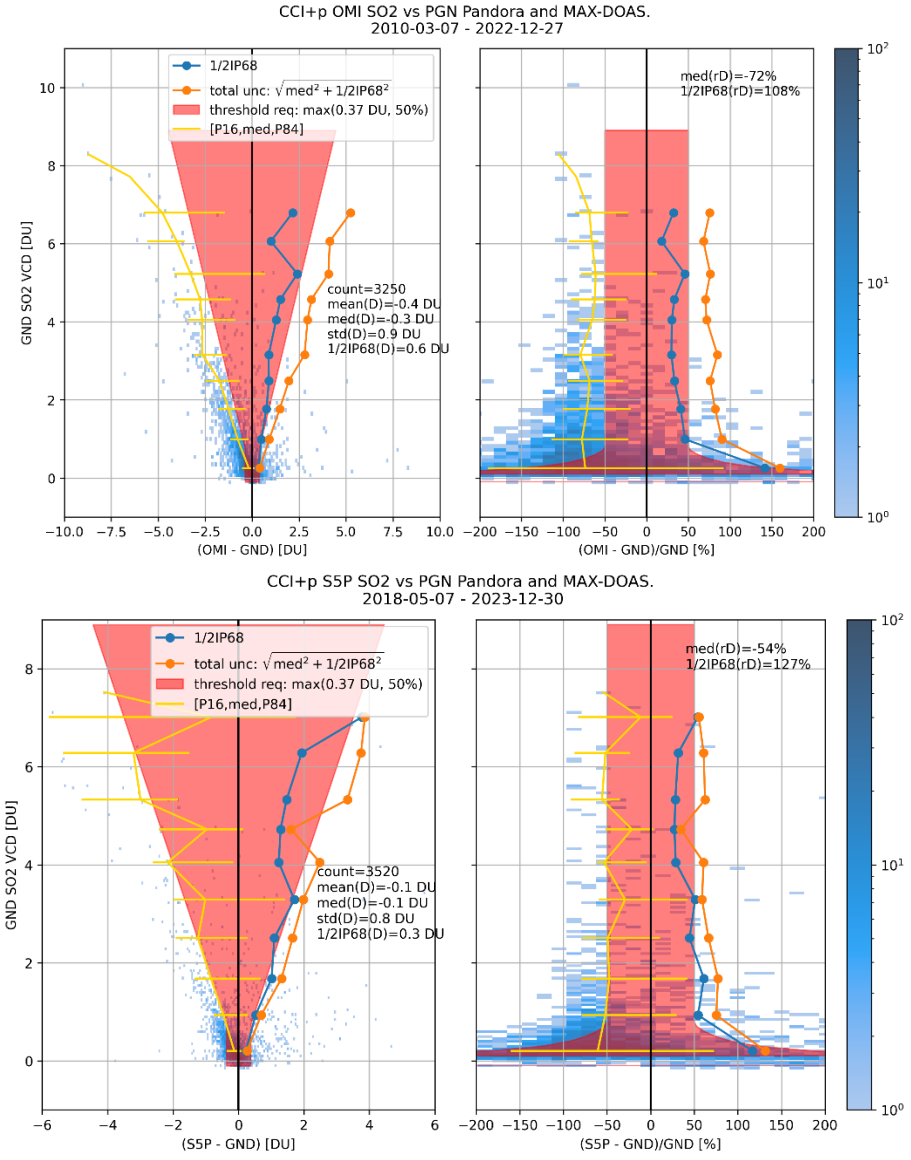


Figure 7-5. Difference and relative difference versus ground-based value for OMI (top) and S5P (bottom), for all PGN Pandora and MAX-DOAS stations combined. Per bin, bias (median difference), dispersion (1/2IP68) and the their combination as total uncertainty QI are indicated, and the GCOS threshold requirement on total uncertainty is indicated as a red zone. The bins are in steps of 0.73 DU; at the lowest bin, the absolute requirement of 0.37 DU holds and in the higher bins the relative requirement of 50% holds.

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

Given the similar biases and explainable differences in dispersion (see discussion section 7.3.2.2), we conclude that the L2 and L3 validation results are consistent for both OMI and TROPOMI.

7.3.2.4. Uncertainty budget of the comparison

The diagnostic total uncertainty QI of the L3 products was compared with the prognostic uncertainties of the satellite and ground-based data. To this end, first the total satellite uncertainty was derived from the random and systematic components by adding them in quadrature. Next the ground-based uncertainties were also added in quadrature to obtain the total uncertainties, which were then averaged per site. For Xianghe and Wuxi, the literature recommendations of 23% and 31% were taken [RD34:PVPv2], which are higher than the values found in the files, for the other stations the uncertainty from the data files was used. The results are shown in Figure 7-6.

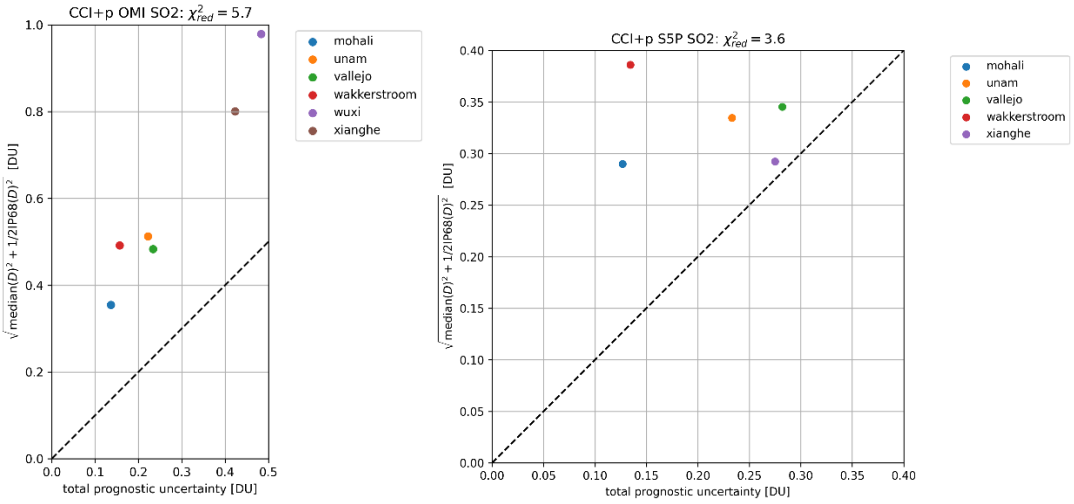


Figure 7-6. Comparison of diagnostic and prognostic total uncertainties. The derived reduced chi^2 value is also shown.

The prognostic uncertainty is lower than the diagnostic one, with quite high reduced chi^2 values, especially for the OMI comparison, pointing to an underestimated prognostic uncertainty of ground-based or satellite data and/or a large contribution of the comparison error, which is not taken into account here.

7.3.2.5. Intercomparison with alternative EO data sets

In [RD35:PVASR] validation was also performed for variant products of OMI COBRA: using TM5 as prior profile. Using TM5 leads to a larger proportional error [RD35:PVASR].

7.3.2.6. Compliance with user requirements

Table 7-1 provides an overview of the compliance with GCOS user requirements for the L3 products.
The products reach breakthrough regarding horizontal and temporal resolution.

[RD3:GCOS_Req] does not formulate separate requirements for bias and dispersion; here we compare these with the total uncertainty requirement. At low VCD (<0.73 DU), the bias of OMI and S5P is at breakthrough and goal respectively, while their dispersion is at threshold and breakthrough. The total uncertainty QI meets the threshold requirement for both products at low VCD.

At high VCD (>0.73 DU), the S5P bias meets the threshold requirement but the OMI bias exceeds it. The dispersion of the S5P and OMI products meet the threshold requirement, but is lowest for OMI (see also Figure 7-5). The total uncertainty QI exceeds the threshold requirement for both products at high VCD. It should be noted that uncertainties in the ground-based data, and comparison error, will contribute to bias, dispersion and the total uncertainty QI.

Stability was not assessed in this report.

Table 7-1. Compliance with GCOS user requirements (Table 4-3) for the L3 COBRA SO_2 products. Bias is calculated as median difference, dispersion as ½ IP68, and total uncertainty as √(bias²+dispersion²).

Quantity	Compliance/evaluation	Remark
Horizontal resolution	0.25°x0.25°	T|B|G : 100|30|10 km
Temporal resolution	Daily, <daily at high lat	B: 1 day, G: 1 hour
Bias *VCD <0.73 DU*	-0.2 DU (OMI), -0.1 DU (S5P)	T|B|G: Max of (0.37|0.22|0.11 DU, 50%|30%|15%)
VCD >0.73 DU	~-70% (OMI), ~-50% (S5P)	Bias, dispersion and the total uncertainty QI are each compared with the total uncertainty requirement. Bias and dispersion increase with VCD.
Dispersion *VCD <0.73 DU*	0.36 DU (OMI), 0.2 DU (S5P)	
VCD >0.73 DU	~35%(OMI), ~50% (S5P)	
Total uncertainty *VCD <0.73 DU*	0.4 DU (OMI), 0.3 DU (S5P)	
VCD >0.73 DU	~80% (OMI), ~60% (S5P),	
Stability	Not assessed	

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 [RD53: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 satellite pixel satellite co-located with the FTIR station is compared to the mean of 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 [RD53: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 (IASI merged)

This product has been validated in [RD36:PVIRv1]. No change has been made in the current report for the bias and dispersion estimates. We have added a paragraph on the validation of IASI merged product uncertainty and updated the stability section.

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.
Although validation results showed that a better spatial resolution could be used for IASI CO, the 1°x1° resolution is kept since the final product is a merging with MOPITT, which does not provide a better resolution.

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 a priori profile had to be hardcoded 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, dispersion, and uncertainty

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 Figure 8-1). 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. The Spearman correlation is very good (0.96).

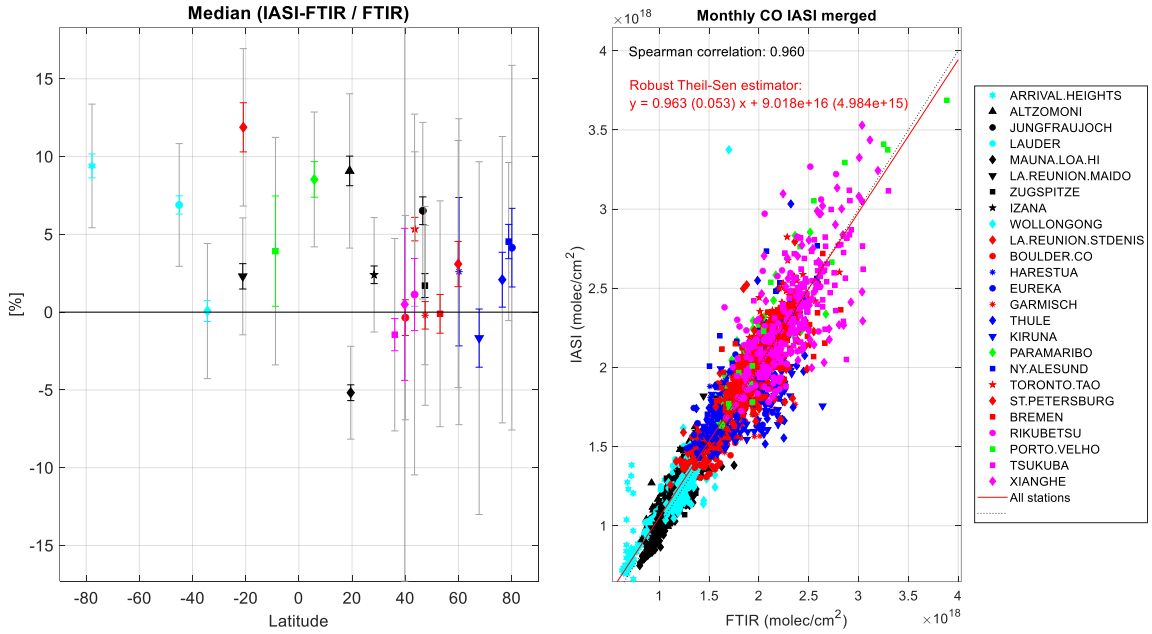


Figure 8-1: a) Left: Individual biases at all sites as a function of latitude; The grey bars are the dispersion of the differences (1-sigma), and the colored bars are the errors on the median (=2*dispersion/sqrt(nb coincidences)). b) Right: the scatter plot of the comparisons.

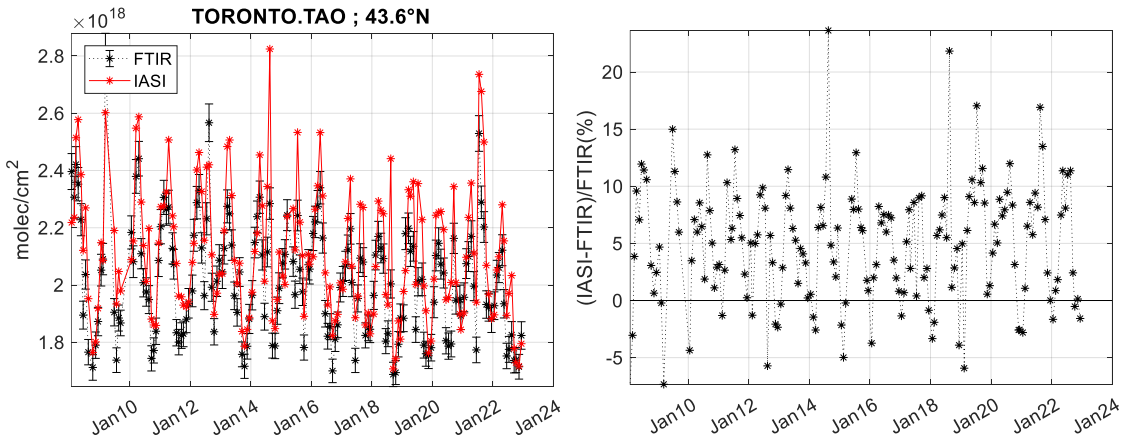


Figure 8-2: Left: CO time-series from IASI L3 multi-platform data set and FTIR at Toronto. Right: their relative differences.

Uncertainty

Only the total uncertainty is provided in the L3 CO products, so we evaluate it by comparing with the square root of the squared sum of the bias and the dispersion using the χ^2_{red}.

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, as we can observe with the higher dispersion at Xianghe or Porto Velho compared to the combined uncertainty (Figure 8-3). However, the total uncertainty seems also underestimated at clean Arctic stations (blue markers), where the impact of collocation should be smaller.

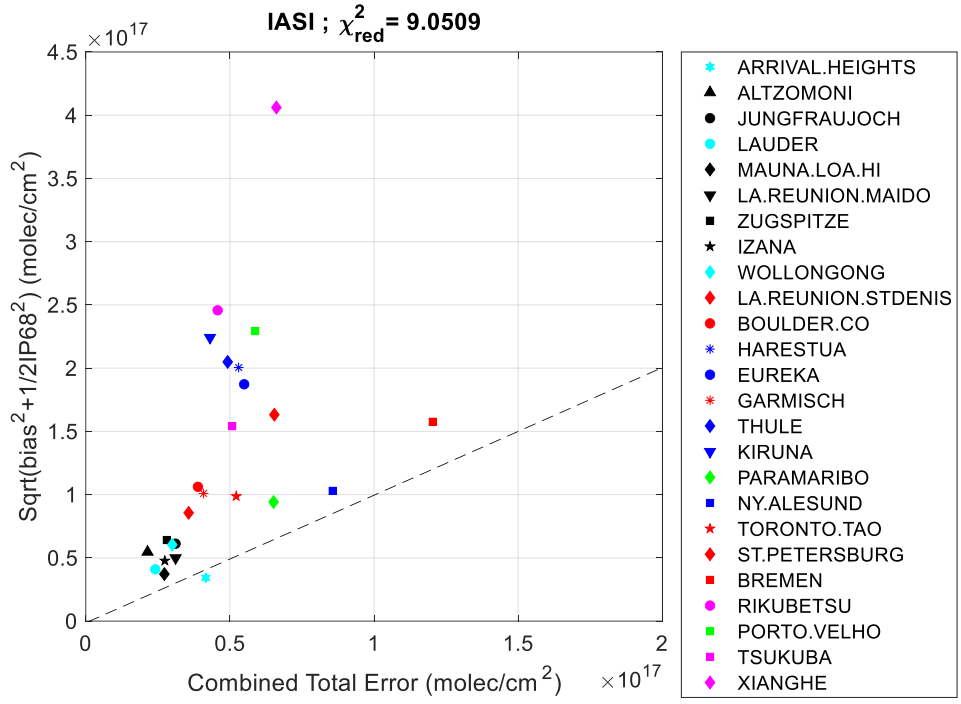


Figure 8-3: Validation of Monthly IASI L3 total uncertainty.

8.3.2.3. Stability

We apply the MLR model described above (section 4.4, Eq(3)) to the time-series of absolute differences IASI-FTIR at all sites. The trends (drifts) are given in Figure 8-4a 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 used for drift calculations.

Usually, IASI CO L3 shows no significant drift, except at 3 sites (positive drift at Lauder, negative drift at Boulder and Rikubetsu). Except for these sites and Ny-Alesund, the drifts are within 2%/decade. We have added the drift when all stations are taken into account by calculating the trend on the time-series of the mean differences between the satellite and the FTIR

stations for each month (Figure 8-4b). We obtain a very small drift of 0.0±0.7%/decade. A clear seasonal cycle in IASI product is confirmed as shown by the MLR model in red.

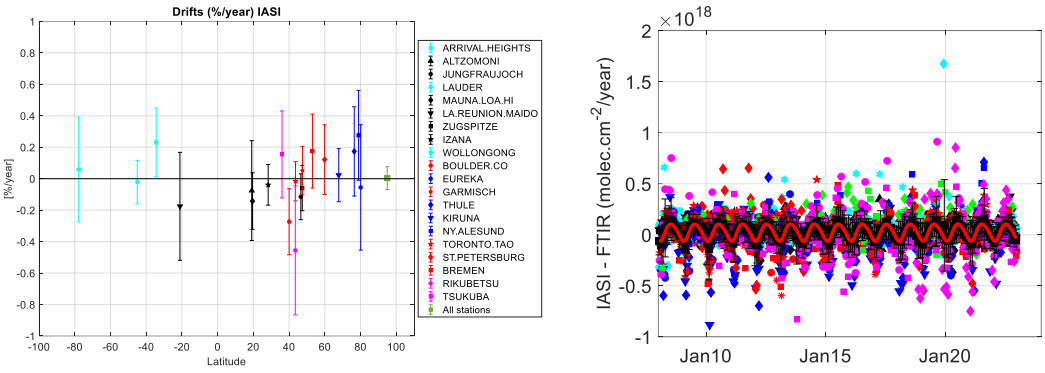


Figure 8-4: a) Drift (%/year) of the IASI CO L3 – FTIR absolute differences. The error bars are equivalent to 2-sigma uncertainty on the drifts. b) Colors: Monthly means of absolute differences at individual stations; black squares and errors bars: Mean of the differences obtained at all stations for each month and standard deviation; Red line: regression model (trend and seasonal cycle, see Sect. 4.4)

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. 8.3.2.2) 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 [RD53: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	Overall: +2.72%. Individual sites: from -5.2% to +12%	

Dispersion	dispersion separately. We compare with total uncertainty requirement	Overall: 1.1E17 molec/cm^2 (7.4%).	- 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.
Total uncertainty (**1-sigma**)	5|2.5|0.5 ppb → ~ 5|2.5|0.5 %	Mean of Sqrt(median^2+1/2IP68^2)= 1.16E17 molec/cm^2 (7.85%). χ^2_{red}= 9.05	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). From χ^2_{red} and Figure 8-3: underestimation of the total uncertainty, especially in the Arctic.
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 5%/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 IASI-MOPITT merged product**

8.3.3.1. **Product specification and file content checks**

The IASI-MOPITT CO L3 product has a monthly temporal resolution and a 1°x1° spatial resolution. It covers the 2008-2023 period. See Table 3-4.
Several tests have been conducted for the merging of IASI and MOPITT, for which we have performed validation. The results of validation have been shared in PMs and with the data provider (M. George) and helped in the choice of the final version. We show in the present report only results for the final product (v.2.6.2).

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 a priori profile had to be hardcoded 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.

8.3.3.2. Bias, dispersion, and uncertainties

Bias

The overall (all stations together) bias (median of relative difference) is +3.98%. The biases at individual sites vary from -4.5% to +12.8% (See Figure 8-5). This overall bias is larger than for IASI alone (2.72%). This is confirmed by the Theil-Sen regression (see Figure 8-5b) giving a smaller slope of 0.95. The intercept is also larger (1.2E17 molec/cm^2).
As for IASI alone, although the median biases at individual sites are usually within 5%, the individual relative differences are larger, and usually show a seasonal cycle, but smaller than for IASI alone (see Figure 8-7 for illustration).

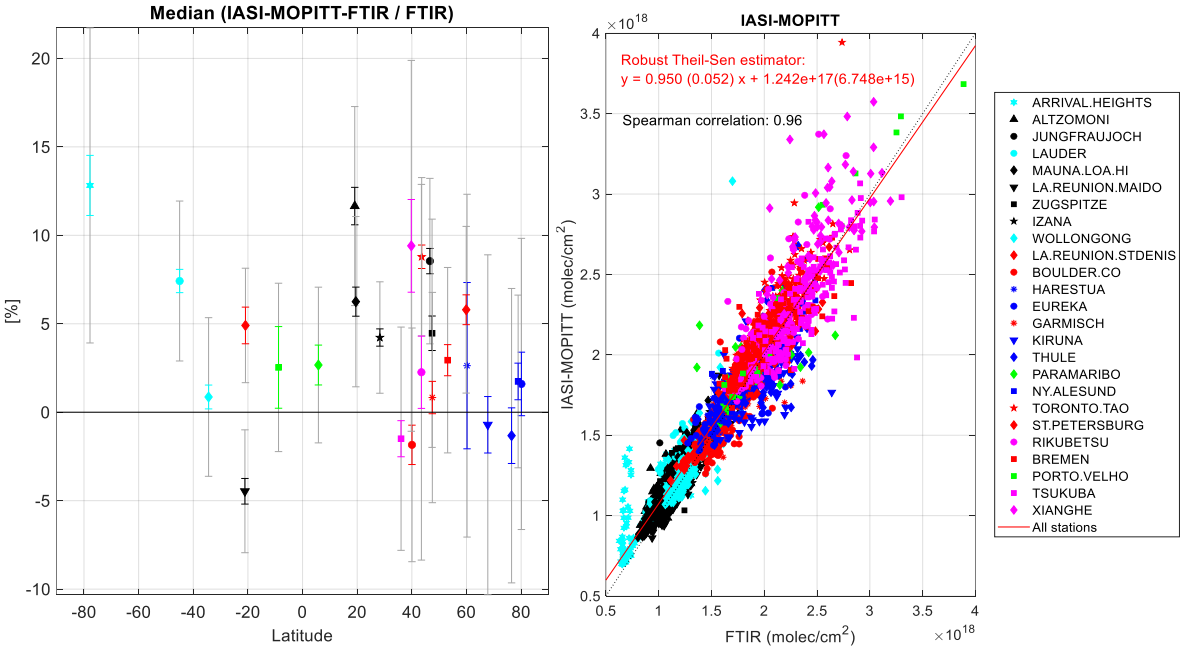


Figure 8-5: a) Left: Individual biases at all sites as a function of latitude; The grey bars are the dispersion of the differences (1-sigma), and the colored bars are the errors on the median (=2*dispersion/sqrt(nb coincidences)). b) Right: the scatter plot of the comparisons.

Dispersion

The overall dispersion (1/2IP68) of the IASI-MOPITT comparisons with FTIR is 1.1E17 molec/cm^2 (or 7.4%) which is the same as for the IASI alone comparison (see previous section). However, it is much larger at the cleanest site Arrival Heights (1.1E17 molec/cm^2; 15%), where the disagreement occurs in Spring. The larger absolute dispersion is 3.4E17 molec/cm^2 (13.2%) at Xianghe. The Spearman correlation is the same as for the IASI alone product (0.96).

Uncertainty

Only the total uncertainty is provided in the L3 CO products, so we evaluate it by comparing with the square root of the squared sum of the bias and the dispersion using the χ^2_{red}.
The underestimation of total uncertainty that we observed in IASI alone product is not present in the IASI-MOPITT merged product (Figure 8-6): the χ^2_{red} is very close to 1, given a high confidence on the provided uncertainty.

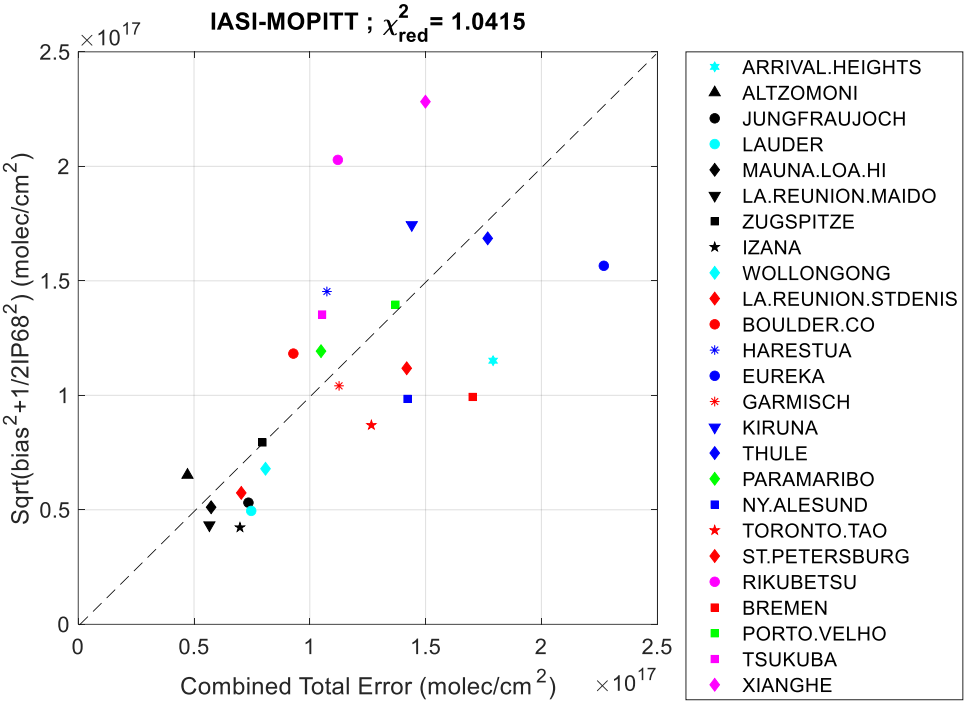


Figure 8-6: Validation of Monthly IASI-MOPITT L3 total uncertainty.

8.3.3.3. Stability

We apply the MLR model described above (Eq(3)) to the time-series of absolute differences IASI-MOPITT -FTIR at all sites. The trends (drifts) are given in Figure 8-7a 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 used for drift calculations.

IASI-MOPITT shows no significant drift, except at 3 sites (negative drift at Arrival Heights, positive drift at Bremen and Ny-Alesund). The drifts are mostly within 2-3%/decade. However, the drift using all stations (see Sect. 8.3.2.3) is significantly positive: 1.3±0.8 %/decade. This

higher drift compared to IASI alone seems to be due to two factors. First, the time period: IASI alone ends in Dec. 2022 while IASI-MOPITT ends in Nov. 2023. If we limit our IASI-MOPITT comparison to the same time period, the drift is lower (0.8±0.8 %/decade). Second, the bias at Xianghe is larger with IASI-MOPITT, and this station starts the measurements only in 2018. Therefore the drift can also be attributed partly to the sampling of the comparisons.

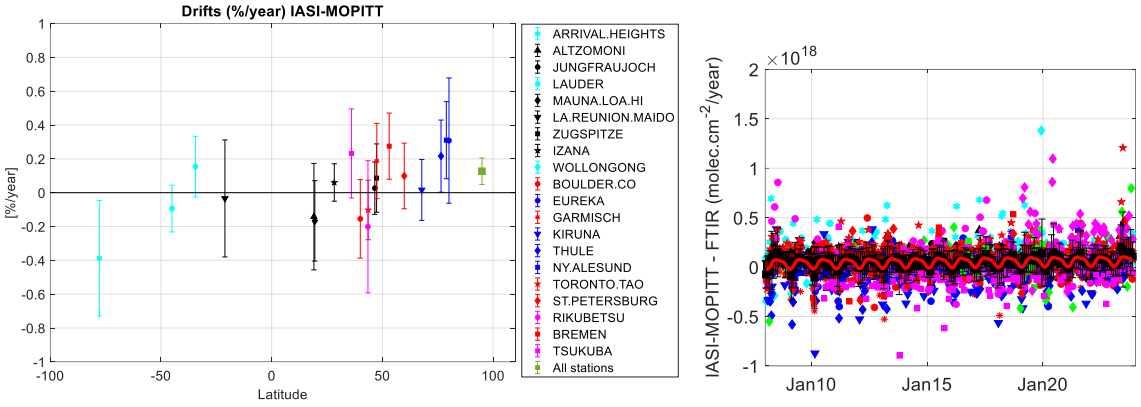


Figure 8-7: a) Drift (%/year) of the IASI-MOPITT CO L3 – FTIR absolute differences. The error bars are equivalent to 2-sigma uncertainty on the drifts. b) Colors: Monthly means of absolute differences at individual stations; black squares and errors bars: Mean of the differences obtained at all stations for each month and standard deviation; Red line: regression model (trend and seasonal cycle, see Sect. 4.4)

8.3.3.4 Compliance with user requirements

Table 8-2. Evaluation, compliance with GCOS user requirements (Table 4-4) for L3 IASI-MOPITT 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	**Spatial resolution limited by MOPITT**
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: +3.98%. Individual sites: from -4.5% to +12.8%	- Still a seasonal cycle on the relative differences but smaller than IASI alone. - 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 %	Mean of Sqrt(median^2+1/2IP68^2)= 1.23E17 molec/cm^2 (8.36%). χ^2_{red}= 1.04	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). From χ^2_{red} and Figure 8-6: good estimation of the total uncertainty in the L3 IASI-MOPITT product.
Dependencies			Advise to **provide influence quantities** in L3 files
Stability	2|1|<1 ppb/dec → ~ 2|1|<1 %/dec	Drift are within 4%/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

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 [RD18: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; <0.5; and <1.
The best validation results according to bias/dispersion/correlation when the co-location is made with the nearest pixel is obtained with the threshold 1 (results are however quite similar using 0.5). Only these validation results are shown in this report.

Co-location.
The version v1 of NH_3 L3 had a fine spatial resolution (0.125°x0.125°). It was discussed in [RD36:PVIRv1] that using the **50km collocation criteria** improved the validation metrics, and the use of a coarser grid was advised.
The version 2 of NH_3 L3 has now a spatial resolution of 0.2°x0.2°. We can therefore use in the current report the unique collocated pixel (as done for HCHO and CO). The present report shows only the validation of v2 NH_3 products (see PVR1 for v1).

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: although the surface altitude is provided in L3 files v2, the apriori profiles are still not provided, so a correction of the altitude is not possible without using external ancillary data, which is beyond the scope tof this project. 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).

Comparison pair handling
Within the month where a collocated valid IASI pixel exists, we average all FTIR data that are measured +/- 2 hours from the IASI overpass time, according to the short lifetime of NH_3 and its diurnal cycle that can be large.

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 [RD18: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

There are four different L3 v2 products to validate: IASI A (2007-2021), B (2013-2023), and C (2019-2023), and the stitched product IASI A/B/C. The stitched product is not a merging of IASI individual data sets, it is a concatenation of only the most recent satellite when several satellites are flying. Note that the v1 of independent A, B, and C products has been validated in [RD36:PVIRv1]. We don't repeat this in the present report which contains only v2 validation.

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, dispersion, and uncertainty

Bias

The individual and overall (all stations together) biases (median of relative difference) for the four products (IASI A, B, and C, and the stitched product) are shown in Figure 9-1a. 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 CDMX= Mexico City, and PortoVelho but there are only 5 coincidences only at this site). This is also seen in the scatter plots (Figure 9-2), for the four IASI products, we observe a small positive constant bias (offset) and a negative proportional bias.

We therefore give bias for clean and polluted conditions separately, an overall bias for all sites being irrelevant for NH_3 (as for HCHO). For individual IASI products:

- **Below 2E15 molec/cm^2**: the respective biases for clean conditions are +68%, + 56%, and + 149%, + 82% for the three Metop A, B, C, and the stitched A/B/C (about 0.4 to 1.1E15 molec.cm^{-2}).

- **Above 4E15 molec/cm^2**: the respective biases for these conditions are -33%, -26%, and -47%, and -33% for the three Metop A, B, C, and the stitched A/B/C (about -2.1 to -3.9E15 molec.cm^{-2}).

The biases for v2 products are similar for Metop A and B to the biases obtained for v1 (See [RD36:PVIRv1]). However, the validation shows worse bias for Metop C v2. This is also visible in the scatter plots below (Figure 9-2), showing a worse slope for Metop C than Metop A or B. In v1, the slope of Metop C was even better than Metop A or B. This needs to be investigated. Note that Metop C has much less coincidences with ground-based data.

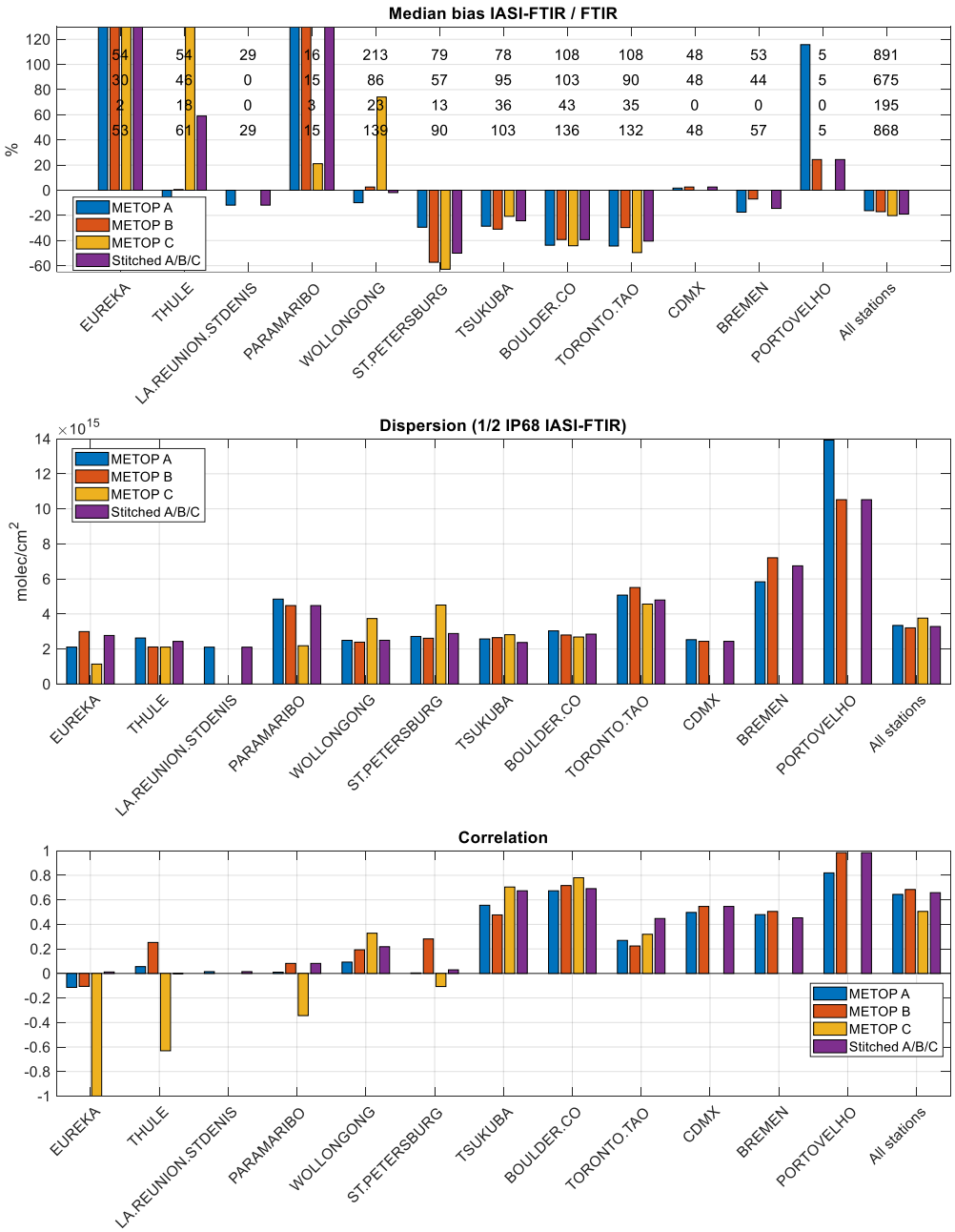


Figure 9-1. For the three IASI sensors and the stitched product: a) The bias (median of the relative differences, in %; b) the dispersion (1/2IP68), in molec/cm^2; c) the Pearson correlation between IASI and FTIR. The number

of coincidences is provided inside a): from top (Metop A) to bottom (Stitched A/B/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 clean sites can be up to 700% but the axis goes up to 130% for clarity.

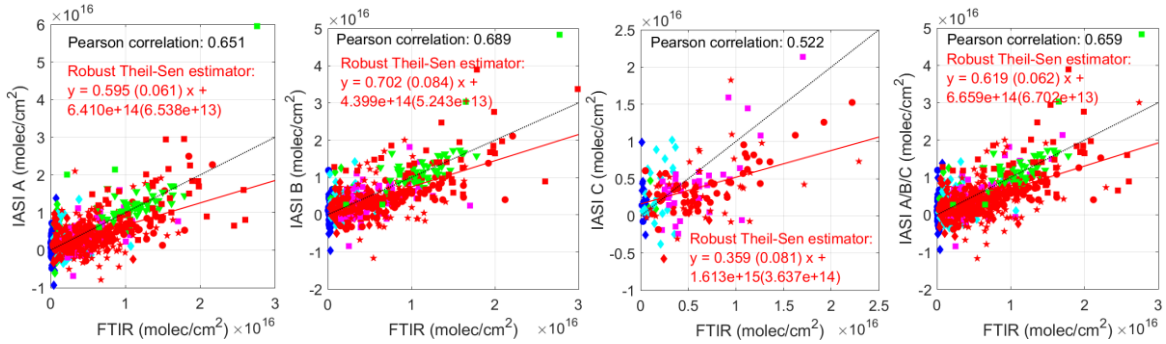


Figure 9-2: Scatter plots between IASI A, B, C, stitched A/B/C and the FTIR stations. Correlation coefficients are provided as well as the regression slope and the intercept.

Dispersion

The individual and overall dispersion (1/2IP68) are shown in Figure 9-1b.

The correlation and the dispersion are larger for v2 than for v1 (see [RD36:PVIRv1]) for all products, but the present validation uses the collocated pixel only, while in [RD36:PVIRv1] a mean of several pixels (within 50 km of the station) was used.

The overall dispersion is about 3.3E15 molec.cm^{-2} for Metop A and B and for the stitched A/B/C, and 3.8E15 molec.cm^{-2} for Metop C. Again, the v2 Metop C is worse than v1: it is now the worse dispersion of the three independent satellites, while it was the best in v1 ([RD36:PVIRv1]).

If we separate for pollution conditions, we obtain:

- **Below 2E15 molec/cm^2**: the dispersion is 2.5 E15 molec.cm^{-2} for Metop A and C (415-500%), 2.8 E15 molec.cm^{-2} for Metop B, and 2.7 E15 molec.cm^{-2} for stitched A/B/C.
- **Above 4E15 molec/cm^2**: the dispersion is 3.6 E15 molec.cm^{-2} for Metop A, B , and stitched A/B/C (41-46%) and seems improved in Metop C: 2.9 E15 molec.cm^{-2} (35%). Therefore, the worsening of Metop C v2 compared to v1 seems to come from clean conditions.

Uncertainty

The uncertainties are validated following Sect. 4.3. Both random and systematic uncertainties are provided in L3 NH_3 files. We first analyse the random uncertainty by comparing it to the 1/2IP68 dispersion (see Figure 9-3a for stitched A/B/C as an illustration; all individual sensors show a similar behavior). Except for Metop C for which χ^2_{red}=1.07 (= very good estimation of the random uncertainty, in line with the decreased dispersion of Metop C for polluted conditions, see above), the random uncertainties are usually slightly underestimated (but not for all sites, see Figure 9-3a), with a reasonable χ^2_{red} between 1.4 and 1.8. Note also that the collocation uncertainty is not taken into account and is probably having an impact on the dispersion: as for HCHO, Figure 9-3a shows a larger deviation from the 1:1 line for the most polluted sites (and Porto Velho has very few coincidences).

When calculating the χ^2_{red} for total uncertainty (comparing it to the square root of the squared sum of bias and dispersion), we observe that is usually below 1 (from 0.74 to 0.84), pointing to a slight overestimation of the systematic uncertainty.

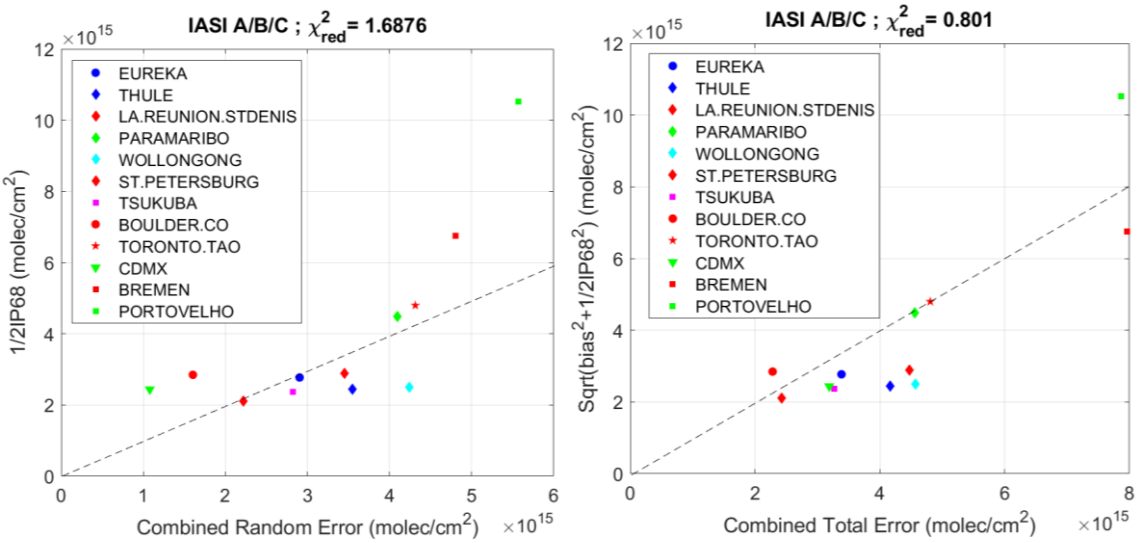
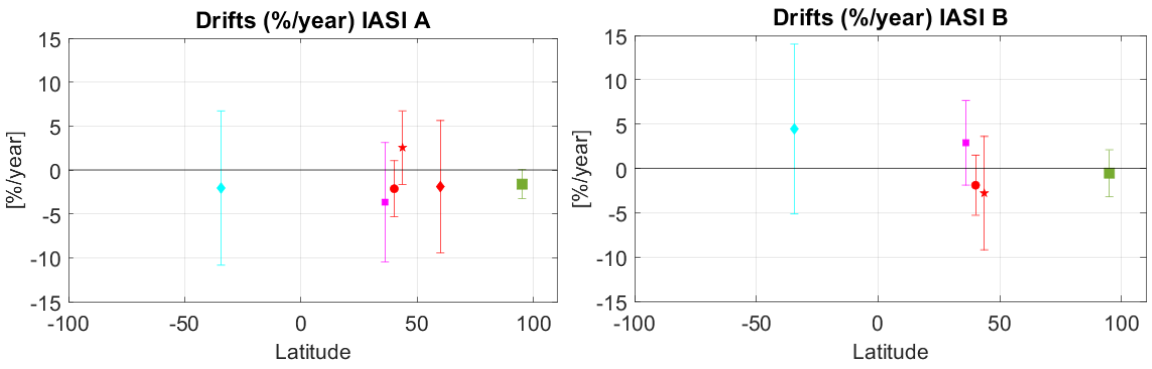


Figure 9-3. a) and b) Validation of Monthly stitched IASI A/B/C L3 random and total uncertainty, respectively.

9.3.2.3. Stability

We apply the MLR model described above (Eq(3)) to the time-series of absolute differences IASI-FTIR at all sites. The trends (drifts) are given in Figure 9-4 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 used 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.

Usually, IASI NH_3 L3 show no significant drift, except at 1 site for stitched A/B/C (Boulder). Except for Metop C, the drifts are within 50%/decade and 2E10^{15} molec.cm^{-2}/decade. A drift was observed at Bremen in v1 ([RD36:PVIRv1]), but the present collocation criteria (nearest pixel; filtering number=1) does not allow for enough coincidences.



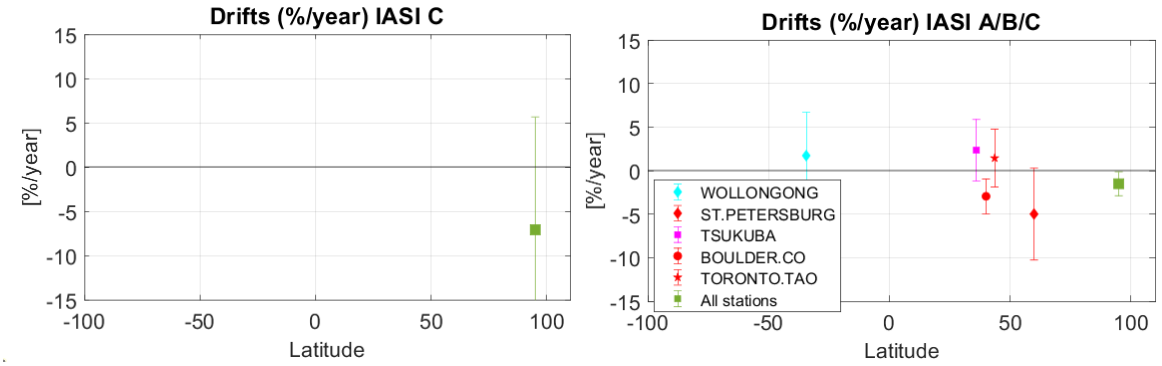


Figure 9-4: Drift (in %/year) of the NH_3 L3 – FTIR absolute differences for IASI A, B, C, and stitched A/B/C; error bars are equivalent to 2-sigma uncertainty on the drifts.

9.3.2.4. Influence quantities

As no influence quantities are provided in the L3 files, we cannot 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: 0.2°x0.2°	The spatial resolution requirement is larger than the goal. However, the validation results of PVR1 (v1 was 0.125°x0.125°) showed that a coarser resolution would be better.
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 +56% to +149% (0.4 to 1.1E15 molec.cm^{-2}) **> 4E15 molec/cm²**: from -26% to -47% (-2.1 to -3.9E15 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. - The bias of Metop C v2 is worse than v1 while Metop A and B v2 are similar to v1. - The dispersion of Metop C is worse in v2 than in v1 for clean conditions. - The dispersion given here is larger than in [RD36:PVIRv1] but only the nearest pixel is compared (in [RD36:PVIRv1]= average within 50 km).
Dispersion	Not specified	**< 2E15 molec/cm²**: from 2.5 to 2.8 E15 molec.cm^{-2} (415-500%). **> 4E15 molec/cm²**: from 2.9E15 molec.cm^{-2} for Metop C (35%) to 3.6 E15 molec.cm^{-2} for the other products (41-46%).	
Total uncertainty (**1-sigma**)	Absolute: 5 PMC | 1.25 PMC Relative: 50% | 20%	Sqrt(median^2+1/2IP68^2)= **< 2E15 molec/cm²**: 2.5, 2.8, 2.6, 2.7 E15 molec.cm^{-2} for A, B, C, and stitched A/B/C, respectively; (400-500%).	From χ^2_{red} and Figure 9-3: good estimation of the random and total uncertainty.
		Sqrt(median^2+1/2IP68^2)= **> 4E15 molec/cm²**: 4.2, 4.1, 4.8, 4.4E15 molec.cm^{-2} for A, B, C, and stitched A/B/C, respectively. (55-59%).	
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 50%/decade and 2E10^{15} molec.cm^{-2}/decade for individual stations. However, when all stations are used together, the drift is within 15%/dec. (exc. For Metop C which has a short time-series, so larger uncertainty on the drift).	More FTIR stations would be needed to better evaluate long-term drift.

Color code:

x>Threshold	Threshold≥x>Goal	Goal≥x

10. Validation of CHOCHO Data Products

10.1. Scope and generalities

Glyoxal retrievals on satellite data have much larger uncertainties than other products such as NO_2 or HCHO. The reasons are a) the relatively small optical depth of the absorption signal at typical concentrations and b) cross-correlations of the glyoxal absorption cross-section with other absorbers (H_2O, O_4, liquid H_2O absorption over water bodies, NO_2 temperature dependence) as well as surface spectral reflectance. There also is a known interference in the presence of clouds. Therefore, glyoxal satellite data are usually averaged over months for validation, the exception being large signals from intense wild fires which can be detected in individual overpasses.
Only the TROPOMI product was ready for validation.

10.2. Validation methodology

The only independent source of validation data for glyoxal are ground-based MAX-DOAS measurements. Unfortunately, these measurements suffer from similar problems as the satellite measurements, and in particular at small glyoxal columns, the ground-based data are prone to offsets and noise. Therefore, validation of winter observations is often not possible. The number of MAX-DOAS stations providing glyoxal observations is limited, and time-series do not always cover the full TROPOMI data set.
In order to add additional verification to the limited validation possible, the precursors-CCI product is also compared to two other TROPOMI glyoxal data products: The GLYRETRO PAL product, which is the predecessor of the precursors-CCI product and the IUP-UB glyoxal product [RD19:Alvarado2020].

Filtering

- The precursors-CCI product was read from the level-3 monthly gridded data product, where quality assurance is already applied so no further filtering was necessary.
- For the PAL product, only data having a qa-value >= 0.5 were used and the solar zenith angle (SZA) was limited to 85°.
- For the IUP-UB product, the SZA was limited to 75°, a cloud radiance threshold of <= 0.5 was applied and a rms filter of 8E-4 used.
- MAX-DOAS data was filtered for fitting rms < 1E-4.

Co-location

- As the analysis was performed on gridded monthly lv3 data having a spatial resolution of 0.125°x0.125°, the grid cell closest to the geolocation of the ground-based station was selected for the comparisons without further averaging.
- Ground-based data was filtered for the mean time of satellite overpass +- 2 hours. The large time window was used to reduce the noise of the reference data. All matching MAX-DOAS measurements for a month were then averaged without further weighting.

Harmonization

Data from the different MAX-DOAS stations has not been harmonised. Fit settings as well as conversion to vertical columns varies as the final products were used as provided by the data originators.

- Data for Chiba were provided by Hitoshi Irie from Chiba University and contain vertical columns
- Data for Uccle and Xianghe were provided by BIRA and contain vertical profiles and columns
- Data for Athens, Orleans, and Incheon were provided by IUP-UB and in this version of the report are based on geometric vertical columns based on the 15° elevation viewing direction

10.3. Validation results

10.3.1. Level-2 products

No validation was performed for level-2 products in this version of the PVIR.

10.3.2. Level-3 products

10.3.2.1. Verification with other TROPOMI products

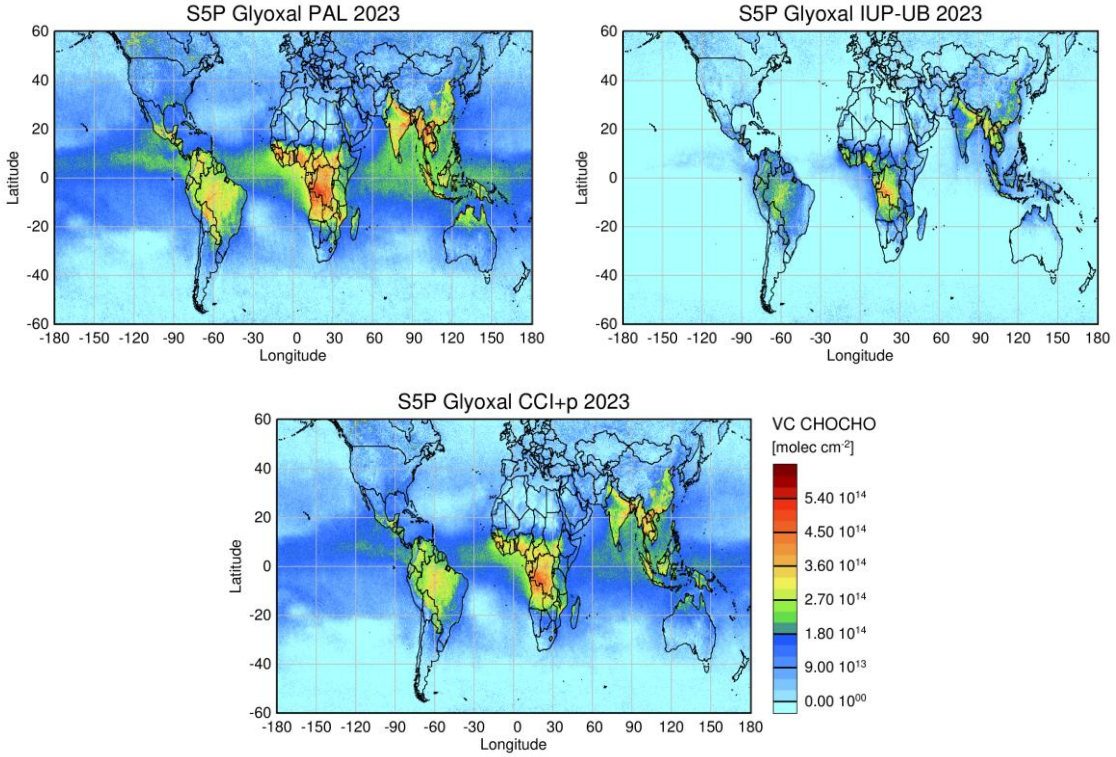


Figure 10-1: Comparison of annual means for 2023 from the three glyoxal data products; the GLYRETRO PAL product (top left), the IUP-UB product (top right) and the new precursors-cci product (bottom).

In Figure 10-1, a direct comparison is shown for the three TROPOMI glyoxal data products on the annual mean for 2023. All three products show similar hotspot regions, but the PAL product is slightly larger than the precursors-cci product while the IUP-UB product is clearly

lower, in particular over the oceans. The low bias over the oceans in the IUP-UB product stems probably from incomplete offset correction.
Comparing the three products of selected regions (Figure 10-2) shows again the offsets between the three data sets. The pattern of the seasonality is very similar, the IUP-UB product having a tendency of larger seasonalities.

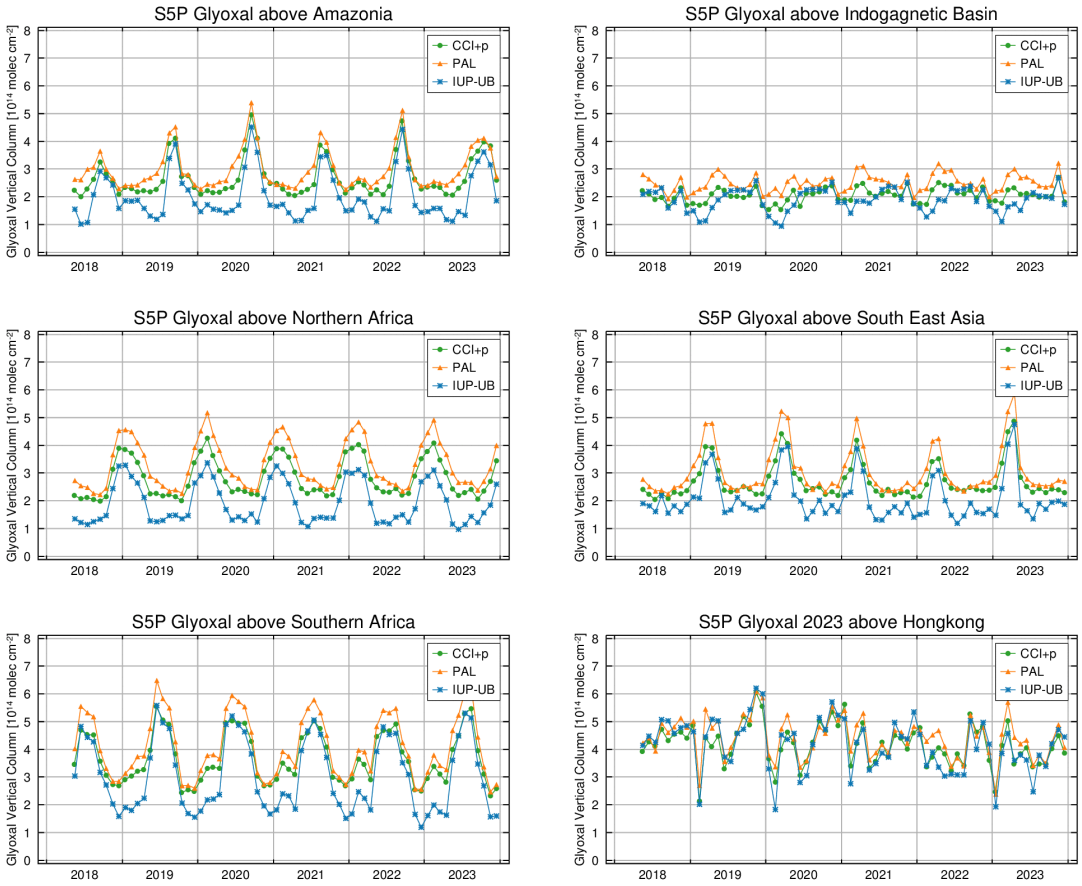


Figure 10-2: Time series of TROPOMI glyoxal columns over six selected regions.

Evaluating the interannual variability, there is no visible difference between the precursors-cci and the IUP-UB products. This is interesting as the latter uses a static climatology for the a priori, always repeating the profiles from 2018, while the other two products are based on model runs for the specific day and year. This illustrates that the interannual variations of the a priori are not very important at least for monthly averages over larger regions.
For the location of HongKong, the agreement between all three products is excellent. This could be a coincidence but it is also worthwhile to note, that this is the only anthropogenically dominated source region in this comparison.

10.3.2.2. Comparison to ground-based MAX-DOAS

In Figure 10-3, time series are shown for the six glyoxal TROPOMI products above the ground-based stations, together with MAX-DOAS observations. In all cases, monthly averages are plotted. Comparing the satellite products among each other, again the close resemblance between the precursors-cci product and the PAL data is apparent, with systematically lower values in the precursors-cci product. The IUP-UB product shows larger scatter, in particular for

Athens and Uccle, and in general lower values than the other two products. The order of magnitude of the ground-based and satellite columns agrees, but the correlation for the month-to-month changes is limited.

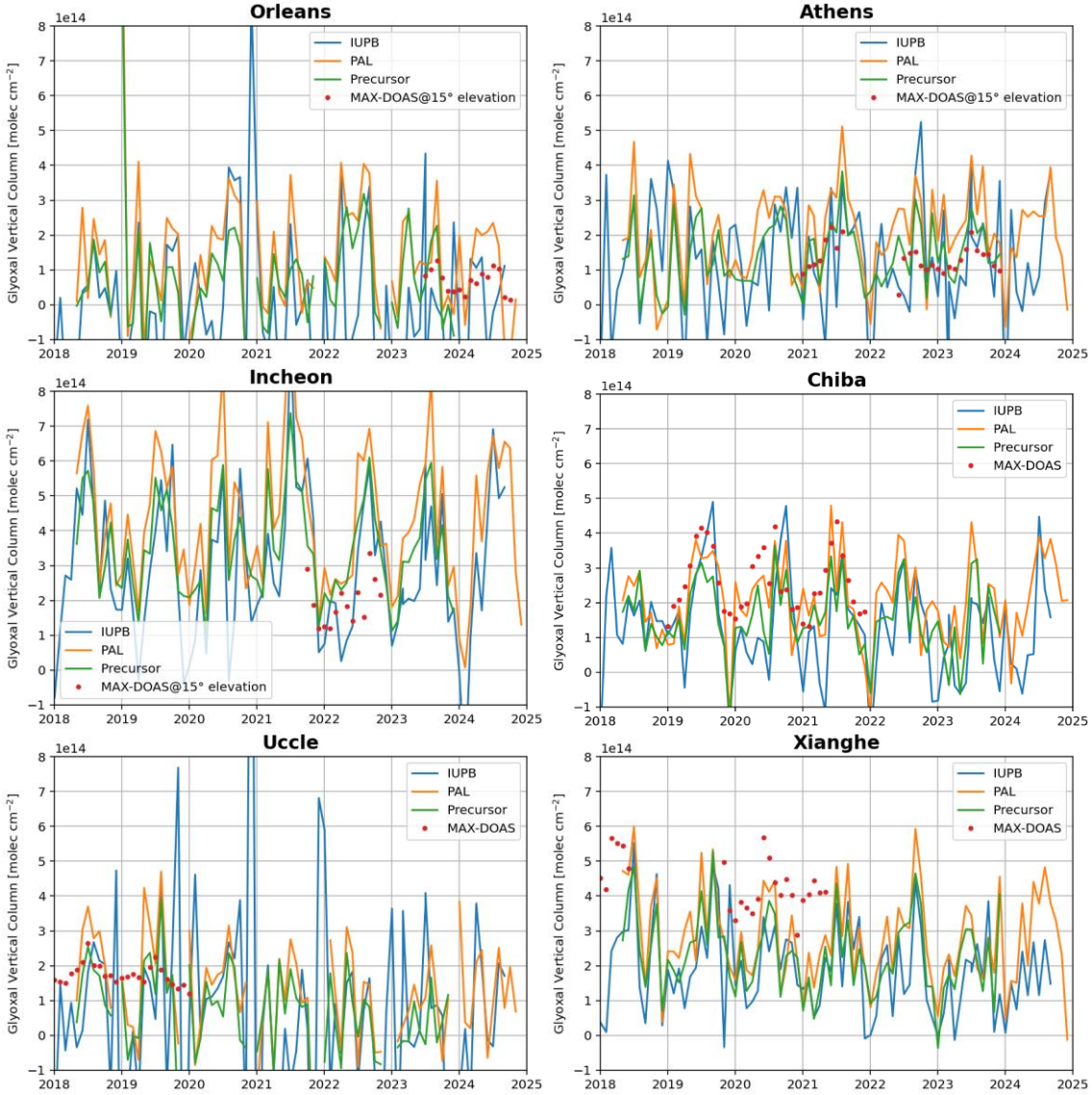
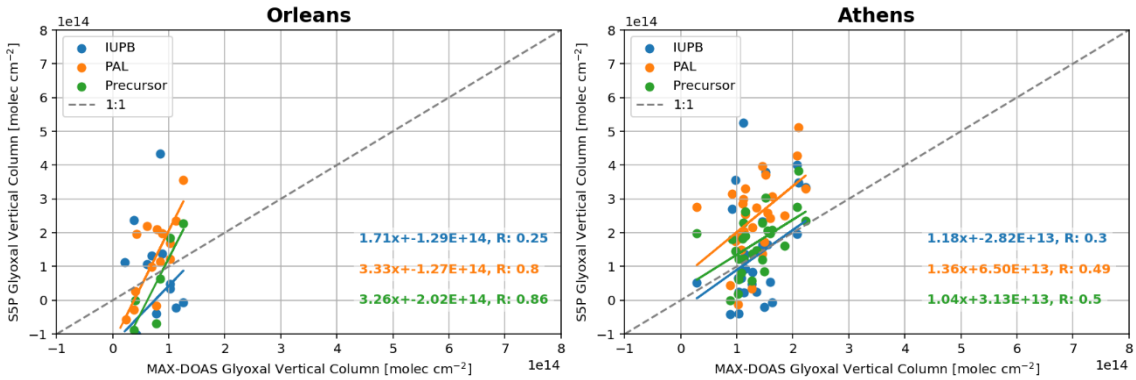


Figure 10-3: Timeseries of TROPOMI glyoxal columns over ground-based MAX-DOAS stations.



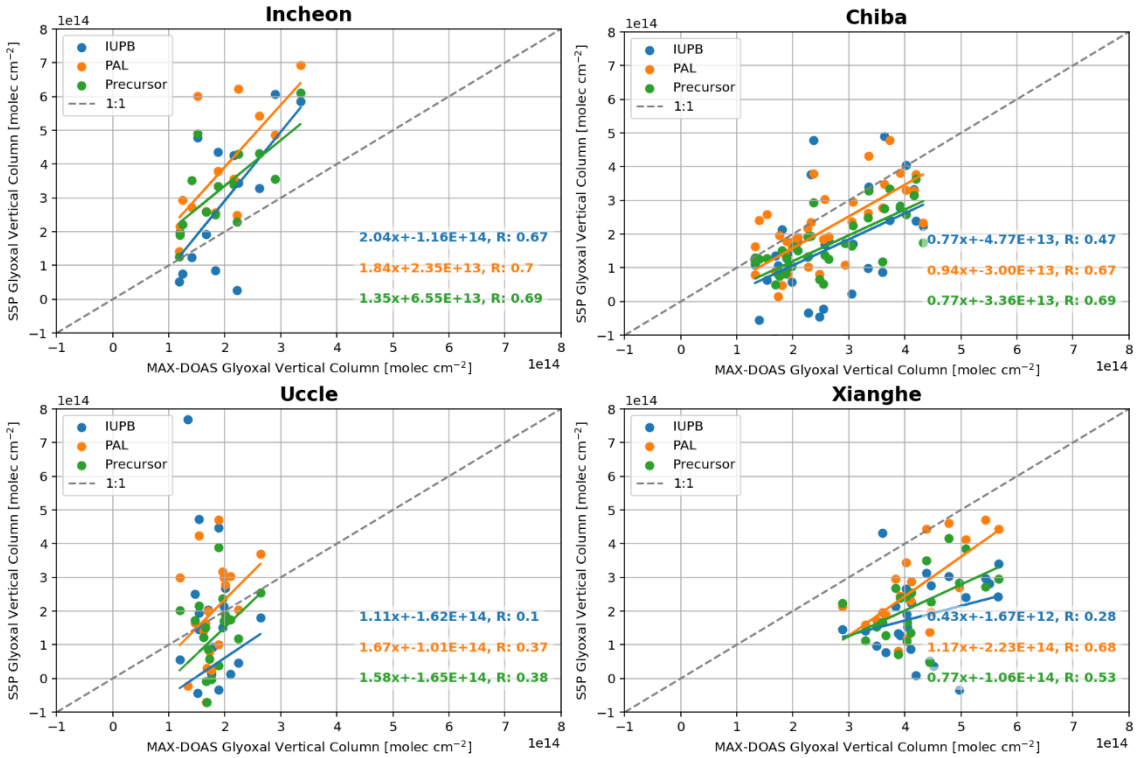


Figure 10-4: As Figure 10-3 but as scatter plots with regression lines.

The same data as in Figure 10-3 is shown in Figure 10-4 as scatter plots. While in Orleans and Athens the values of satellite and ground-based measurements are consistent, TROPOMI measurements over Incheon are systematically larger than the MAX-DOAS data. As shown in Figure 10-3, this is mainly because of a more pronounced seasonality in the satellite data. In Orleans, Athens, Incheon and Uccle the slopes are larger than one, while in Chiba the slope is lower than one and Xianghe the slopes are different and depend on the product. In the case of Orleans, the slopes are much larger than 1, indicating a stronger increase of satellite data at large glyoxal columns. For Uccle the slopes of the PAL and precursors-cci product are much larger than one with correlation coefficients of about 0.4, while the IUPB product shows a slope of 1.1 with a correlation coefficient around 0.1. In Xianghe, the PAL product shows a slope larger than 1, while the IUP-UB product and the precursors-cci products have slopes below 1. The correlation coefficients are 0.28 for the IUP-UB product, 0.68 for the PAL product and 0.53 for the precursors-cci product.
Comparing the results from the three TROPOMI glyoxal products, the precursors-cci product has less bias to the ground-based stations than the PAL product, and clearly less scatter than the IUP-UB product. However, overall, the differences over the stations are not large.

Per station, the bias (median difference) of the precursors-cci product versus MAX-DOAS ranges from -2E14 molec cm^{-2} (Xianghe) to 1E14 molec cm^{-2} (Incheon) and is mostly better than 0.4E14 molec cm^{-2}, while the relative bias (median relative difference ranges from -60% (Orleans) to +60% (Incheon). The dispersion (1/2 IP68 of the differences) ranges from 0.6 to 1E14 molec cm^{-2}. The total uncertainty QI (calculated as $\sqrt{(bias^2+dispersion^2)}$) is mostly around 1E14 molec cm^{-2}, except for Xianghe where it is 2E14 molec cm^{-2}.

10.3.2.3. Compliance with user requirements

Table 10-1. Compliance with GCOS requirements for the S5P CHOCHO L3 data.

Quantity	Requirements T| G	Compliance/evaluation		Remark
Horizontal resolution	100|10 km	L3 daily: 0.25°x0.25° grid	L3 monthly: 0.125°x0.125°	The spatial resolution threshold requirement is achieved. Monthly product close to goal.
Temporal resolution	30 days |1h	L3: Monthly	L3: Daily	User requirement T is reached.
Bias	Max of 4e14|2e14 *molec/cm^2 50%|10%	Bias no more than 2e14 molec/cm^2, mostly better than 0.4e14 molec/cm^2		Agreement depends on station. FRM data not harmonized. Bias: median difference, dispersion: 1/2IP68
Dispersion		0.6 to 1.0 *e14 molec/cm^2		
Total uncertainty (**1-sigma**)		Around 1.0 e14 molec/cm^2 in most cases, 2 e14 molec/cm^2 at most.		Calculated as √(bias^2+dispersion^2)
Stability	Max of 2e14|1e14 *molec cm^{-2}/10y 25%|10%	Not assessed		

Color code:

x>Threshold	Threshold≥x>Goal	Goal≥x

11. 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 11-1. 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 ,8.2,9.2	
(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, 10.3 Uncertainty budget 6.3.2.1, 7.3.2.4, 8.3.3.2, 9.3.2.2	
(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 5.3.2, 6.3.2.4, 7.3.2.3, 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, 5.3.2.2, 6.3.2.3, 8.3.2.3, 8.3.3.3, 9.3.2.3	SO_2, CHOCHO: 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: [RD36:PVIRv1], 6.3.1.4 SO_2: 7.3.2.5	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. CHOCHO limited scope: 10.1 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).	5.3.1.4,5.3.2.1,6.3.2.5,7.3.2.6,8.3.2.6,9.3.2.6, 10.3.2.3	

(11) identify limitations in the CRDP, e.g. anomalies, or conditions under which the product quality is lower than expected.	5.3.1.1	SCIA & GOME-2A QA4ECV high-lat issue
(12) provide recommendations to the EO developers on aspects of CRDP quality that are priorities for future improvement	5.3.2.1,6.3.2.1,7.3.2.1,8.3.2.6,8.3.3.4,9.3.2.6	Optimal grid size, separate random/systematic unc, more ancillary data, format remarks
(13) provide overall conclusions to users on the CRDP quality, and recommendations for use regarding any limitations identified.	5.3.1.4,5.3.2.1,6.3.2.5,7.3.2.6,8.3.2.6,9.3.2.6, 10.3.2.3	

12. Terms, abbreviations and definitions

12.1. Terms and definitions

In Table 12-1 terms and definitions as recommended by the CEOS WGCV and by standardization bodies have been reproduced.

Table 12-1. 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

12.2. 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