Precursors_cci+
ESA Climate Change Initiative (CCI)



Product Validation and Algorithm Selection Report (PVASR) – Version 1.1

Title	Product Validation and Algorithm Selection Report (PVASR)
Reference	Precursors_cci+_D2.3_PVASR_01_01
Issue	01
Revision	01
Status	Final
Date of issue	29/04/2024
Document type	Deliverable

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















	FUNCTION	NAME	DATE
LEAD AUTHOR(S)	Authors	Andreas Richter Folkert Boersma Isabelle De Smedt Nicolas Theys Gaia Pinardi Klaus-Peter Heue Sora Seo Athina Argyrouli Tijl Verhoelst Steven Compernolle	29/02/2024
REVIEWED BY	Science Leaders	M. Van Roozendael Folkert Boersma	01/03/2024
	ESA Project Officer ESA Climate Applications Scientist	Simon Pinnock Claire Macintosh	02/04/2024
DISTRIBUTED TO	ESA Precursors_cci+ Consortium		

Table of Contents

EXECUTIVE SUMMARY 10
1. PURPOSE OF DOCUMENT 11
2. NO_2 11
2.1. Data and Algorithms used 11
2.2. Slant Column Intercomparison 12
2.3. Stratosphere Troposphere Separation 18
2.4. AMF Impacts 23
2.4.1. LER 23
2.4.2. A priori profiles 26
2.4.3. Clouds 28
2.5. Validation 36
2.6. Algorithm Selection 40
3. HCHO 43
3.1. Data and Algorithms used 43
3.2. Slant Column Intercomparison 43
3.3. Background correction 46
3.4. AMF Impacts 47
3.4.1. LER 47
3.4.2. A priori profiles 50
3.4.3. Clouds 53
3.5. Validation 56
3.5.1. Cloud-corrected vs clear-sky columns 56
3.5.2. Prior profile choice 58
3.6. Algorithm Selection 59
4. SO_2 60
4.1. Data and Algorithms used 60
4.2. Fitting approaches 60
4.3. AMF Impacts 63
4.3.1. LER 63
4.3.2. A priori profiles 65
4.3.3. Clouds 67
4.4. Validation 70
4.5. Algorithm Selection 71
5. SUMMARY AND CONCLUSIONS 73
6. REFERENCES 75
ANNEX: VALIDATION METHODOLOGY 77
A1 Stratospheric NO2 QA4ECV & AC SAF prototype validation methodology 77
A2 GOME-2A AC SAF HCHO validation methodology 78
A3 GOME-2A QA4ECV HCHO validation methodology 79

List of Figures

Figure 2.1: Time-series of monthly averages of NO_2 slant columns from three algirithms (AC-SAF operational (black), AC-SAF prototype (green), QA4ECV (orange)) over the clean Pacific region (20°S - 20°N, 110 - 150°W) for the period of 2007 to 2017. 14

Figure 2.2: Same as Figure 2.1, but for sub-Saharan Africa (15 – 0°S, 10 – 35°E). 14

Figure 2.3: Same as Figure 2.1, but for anthropogenically polluted Europe (35 – 50°N, 10°W–20°E). 15

Figure 2.4: Distribution of the deviation the GOME-2A NO_2 slant columns from the mean slant column within each small-box in a histogram for the AC-SAF opertational (black), AC-SAF prototype (green), and QA4ECV (orange) algorithms. The histogram includes contributions from all boxes within the Pacific reference sector (60°S – 60°N, 150 – 180°E) from the 15th day of each month in the period 2007-2017. The fitted Gaussian function is also plotted as solid line. 16

Figure 2.5: Temporal evolution of GOME-2A NO_2 slant column uncertainties from the AC-SAF operational (black triangle), AC-SAF prototype (green square), and QA4ECV (orange circle) algorithms using deviations of NO_2 slant columns from box mean values over the Pacific reference sector. The solid lines are linear regressions fitted to the data. 17

Figure 2.6: Time-series of GOME-2A geometric NO_2 columns from the AC-SAF prototype algorithm with offline (black) and reprocessed (blue) level 1b v6.3 data, along with the differences in geometric NO_2 columns between the two products (red). 18

Figure 2.7: Comparison of GOME-2A stratospheric corrections. Upper left: total NO_2 column from QA4ECV. Upper middle: stratospheric NO_2 column from data assimilation and STREAM in QA4ECV data product. Upper right: stratospheric NO_2 from STREAM in AC-SAF product. Lower panels: tropospheric NO_2 column in QA4ECV product (lower left), in QA4ECV when using STREAM (lower center), and AC-SAF when using STREAM (lower right). 20

Figure 2.8: Zonal average NO_2 columns on 5 November 2008 between 44-46°N from GOME-2A. The tropospheric columns shown are from the QA4ECV and AC-SAF data products, calculated as the tropospheric slant column divided by the geometrical air mass factor. 20

Figure 2.9: Timeseries of GOME-2A stratospheric and tropospheric NO_2 over the Pacific sector. 21

Figure 2.10: As Figure 2.9 but for sub-Saharan Africa. 22

Figure 2.11: As Figure 2.9, but for Europe. Tropospheric NO_2 column values (indicated in red and green) have been divided by a factor of 4. 22

Figure 2.12: As Figure 2.9, but for eastern China. Tropospheric NO_2 column values (indicated in red and green) have been divided by a factor of 4. 23

Figure 2.13: Relative change in the tropospheric NO_2 clear-sky AMF when using the GOME2 DLER min instead of the GOME2 LER min surface reflectance on GOME-2A data for 4 selected days. 24

Figure 2.14: Relative change in the tropospheric NO_2 clear-sky AMF when using the OMI min instead of the GOME2 LER min surface reflectance on GOME-2A data for 4 selected days. . 25

Figure 2.15: Relative change in the tropospheric NO_2 clear-sky AMF when using the S5P LER min instead of the GOME2 LER min surface reflectance on GOME-2A data for 4 selected days. 25

Figure 2.16: Differences in tropospheric NO_2 columns between different models. Shown are the differences between the CAMS o-suite and the TM5-MP (left) and the CAMS-reanalysis (right), both for Janaury 2022 27

Figure 2.17: Differences in tropospheric NO_2 columns between GOME-2C retrievals and CAMS o-suite (left) and TM5-MP (right) for January 2022. 27

Figure 2.18: Ratio of tropospheric NO_2 AMFs when replacing the CAMS o-suite a priori by TM5-MP (left) and CAMS reanalysis (right) for Janaury 2022. 28

Figure 2.19: Difference of tropospheric NO_2 columns when replacing the CAMS o-suite a priori by TM5-MP (left) and CAMS reanalysis (right) for Janaury 2022. 28

Figure 2.20: Time series of the monthly means of Cloud Fraction over Europe: ACSAF_OFF (red line) versus FRESCO+ v7 (blue line) 31

Figure 2.21: Time series of the monthly means of Cloud Fraction over Europe: ACSAF_RPRO (red line) versus FRESCO+ v8 (blue line) 31

Figure 2.22: Time series of the monthly means of Cloud Height over Europe: ACSAF_OFF (red line) versus FRESCO+ v7 (blue line) 32

Figure 2.23: Time series of the monthly means of Cloud Height over Europe: ACSAF_RPRO (red line) versus FRESCO+ v8 (blue line) 32

Figure 2.24: Time series of the monthly means of Cloud Fraction over Eastern China: ACSAF_OFF (red line) versus FRESCO+ v7 (blue line) 33

Figure 2.25: Time series of the monthly means of Cloud Fraction over Eastern China: ACSAF_RPRO (red line) versus FRESCO+ v8 (blue line) 33

Figure 2.26: Time series of the monthly means of Cloud Height over Eastern China: ACSAF_OFF (red line) versus FRESCO+ v7 (blue line) 34

Figure 2.27: Time series of the monthly means of Cloud Height over Eastern China: ACSAF_RPRO (red line) versus FRESCO+ v8 (blue line) 34

Figure 2.28: Time-series of the ratio of total tropospheric NO_2 vertical columns to clear-sky tropospheric NO_2 vertical columns applying the OCRA/ROCINN (used in AC-SAF OFFL) and the FRESCO+ v7 (used in QA4ECV) cloud products for clean Pacific sector 35

Figure 2.29: Same as Figure 2.28, but for Europe. 35

Figure 2.30: Time-series of the absolute differences between the satellite-based stratospheric NO_2 VCDs and those measured with the ground-based SAOZ ZSL-DOAS instruments (LATMOSv3 processing). Top to bottom: QA4ECV DA, QA4ECV STREAM, AC SAF prototype. 37

Figure 2.31: Box-whisker plots corresponding to the comparisons shown in Figure 2.5-1. Comparisons at Paris and Bauru are affected by tropospheric pollution in the ZSL-DOAS signal and the large negative deviations are therefore not a reason for concern. Most noteworthy are the different signals at the highest latitudes: data assimilation (left-most panel) leads to slight positive biases not seen in the STREAM data..... 38

Figure 2.32: Histograms of the differences between satellite and ground-based stratospheric NO_2 column measurements at Dumont d'Urville (East coast of Antarctica), in this case for SCIAMACHY with either QA4ECV DA (left-hand panel) or QA4ECV STREAM (right-hand panel) processing. Note the bimodal distribution for QA4ECV DA..... 39

Figure 2.33: Scatter plot for the SCIAMACHY QA4ECV DA comparisons at Dumont d'Urville (cf. the left-hand panel of Figure 2.32). Markers are coloured by tropospheric VCD. 40

Figure 3.1: GOME-2A HCHO background-corrected slant columns retrieved with the QA4ECV and the GDP-4.8 algorithms. The first panel shows time series of the HCHO bc-SCDs averaged over the reference sector. The upper and lower right panels show global daily maps on 2007-07-30, while the lower left panel shows the corresponding SCD differences. 44

Figure 3.2: Same as Figure 3.1, but showing the residuals of the fits (RMS)..... 44

Figure 3.3: Histograms of the background-corrected HCHO SCDs over the Pacific Ocean for two selected days, 2007-07-30 and 2015-10-17. The right plots show the histograms of the differences between the three retrievals (QA4ECV, AC SAF GDP-4.8 and AC SAF Prototype). 45

Figure 3.4: Time series of the corrected HCHO SCD of the two orbits over the Pacific for 2015-10-17. The line shows the running mean over 11 Pixels. Towards the end of the orbit, the green line (QA4ECV) deviates towards lower values. 46

Figure 3.5: Comparison of March min LER values at 340 and 440 nm in the TROPOMI, OMI and GOME-2 databases. 48

Figure 3.6: TROPOMI HCHO VCD clear in Jul. 2020 obtained using the OMI (a) or TROPOMI (b) min LER databases. Figures (c) and (d) show the absolute differences between (b) and (a) and the histogram of the relative differences, respectively. Figures (e) and (f) show the absolute differences zoomed over the US and Southeastern Europe..... 49

Figure 3.7: (a) Differences of TROPOMI HCHO VCD clear in Jan. 2020, as in Figure 3.6c. (b) Differences of TROPOMI HCHO VCD clear in Jul. 2020, obtained using TROPOMI DLER or min LER. 50

Figure 3.8: Annual mean of model HCHO columns in 2021 in the afternoon: TM5, MAGRITTE and CAMS REA, and comparison with TROPOMI HCHO columns..... 51

Figure 3.9: Comparison of HCHO model profiles over Porto Velho (Brazil) and Xianghe (China) respectively in August and in February 2012. Insets provide the values of the model integrated

columns, the profile height (pressure under which 70% of the model column lies), and the corresponding satellite AMF. 51

Figure 3.10: Absolute differences of satellite HCHO columns obtained when replacing TM5(-MP) profiles by MAGRITTE (first column) or CAMS REA (second column) profiles. 52

Figure 3.11: Time series of the monthly means of Cloud Fraction from AC SAF (red line) and FRESCO (blue line) over (a) Amazon, (b) SE US. 54

Figure 3.12: Time series of the monthly means of Cloud Height from AC SAF (red line) and FRESCO (blue line) over (a) SE US, (b) Northern China 55

Figure 3.13: Monthly average of HCHO VCDs obtained from GOME, SCIAMACHY, GOME-2A, B, C, OMI, and TROPOMI measurements over Northern China and the Amazon regions. The upper plots present the HCHO VCD clear [units: 10^{15} molec/cm^2]. The lower plots present the ratio of VCD/VCD_clear [no unit] as a quantification of the cloud correction impact in each product. 56

Figure 3.14: HCHO comparison of cloud-corrected (left) and not cloud corrected (right) GOME-2A AC SAF columns with ground-based FTIR measurements [Pmolec/cm^2=1e15molec/cm^2]. 57

Figure 3.15. HCHO comparison of cloud-corrected (left) and not cloud corrected (right) GOME-2A QA4ECV columns with ground-based MAX-DOAS observations [Pmolec/cm^2=1e15molec/cm^2]. 57

Figure 3.16. Seasonally averaged HCHO profiles at Xianghe: model profiles TM5, CAMS, MAGRITTE and MAX-DOAS profiles retrieved with different settings: bePRO, MMF and MAPA (the latter two being developed for FRM4DOAS). 58

Figure 4.1: OMI 0.1°x0.1° monthly averaged SO_2 slant columns (in DU) retrieved with the (left) COBRA, (center) PCA and (right) DOAS techniques, for August 2006 over China. 61

Figure 4.2: Scatter plots of (gridded) OMI SO_2 slant columns from COBRA (y-axis) and PCA (x-axis) for 4 months (Feb, May, Aug, Nov) in 2006, for the geographical region shown in Figure 4.1 (10-60°N, 90-150°E). The slope and offset of the regression line and the correlation coefficient are given as an inset for each plot. 62

Figure 4.3: Comparison of April and July minimum LER values at 328nm from the OMI and TROPOMI albedo climatologies 63

Figure 4.4: Relative differences of SO_2 clear-sky air mass factors obtained using TROPOMI and OMI minimum LER, for the months of January, April, July, and October 2020. Pixel selection criteria: SZA less than 65°, cloud fraction (CRB) lower than 30% and TROPOMI rows 26– 424. 64

Figure 4.5: Comparison of TROPOMI monthly averaged SO_2 vertical columns using TROPOMI (red) and OMI (black) albedo climatologies as input of the (clear sky) AMF calculation, for January, April, July, and October 2020. The VCD results and relative differences are presented for 4 selected regions (Northeast India, Northeast China, South Africa, and Mexico). The same pixel selection criteria are used as in Figure 4.4. 65

Figure 4.6: Comparison of OMI monthly averaged SO_2 vertical columns using CAMS and TM5 SO_2 apriori profiles, as input of the (clear sky) AMF calculation. The VCD results and relative differences are presented for 2012 and for the 4 selected regions used in Figure 4.5. The same pixel selection criteria are used. 66

Figure 4.7: Comparison of TROPOMI monthly averaged SO_2 vertical columns (October 2020) using CAMS and TM5 SO_2 apriori profiles, as input of the (clear sky) AMF calculation. The maps of VCDs (top left: TM5, top right: CAMS) and relative differences (bottom left) are shown over the region covering the middle East and India [0-40°N, 30-100°E]. The histogram of the relative differences is displayed on the bottom right panel. The same pixel selection criteria are used as previously. Fo the relative differences, only SO_2 VCDs (based on TM5 profiles) larger than 0.3 DU are considered. 67

Figure 4.8: Comparison of TROPOMI daily averaged SO_2 vertical columns using cloud corrected AMFs based on different cloud products (see text) and and clear-sky AMFs, for 4 days and for 4 selected regions (Northeast India, Northeast China, South Africa, and Mexico). Pixel selection criteria: SZA less than 65°, cloud fraction (OCRA/ROCINN CRB) lower than 30% and TROPOMI rows 26– 424. 69

Figure 4.9: Comparison of OMI and TROPOMI monthly averaged SO_2 vertical columns, applying cloud correction (OMI: OMCLDO2, TROPOMI: OCRA/ROCINN scaled) or not ('clear') in the AMF calculation. The VCD results are presented for 4 months in 2020 and for 4 selected regions (Northeast India, Northeast China, South Africa, and Mexico). 70

Figure 4.10: Comparison of OMI COBRA SO_2 with MAX-DOAS. Top: product with CAMS reanalysis as prior. Bottom: product with TM5 reanalysis as prior. Left: direct comparison. Middle: after replacing satellite prior with MAX-DOAS retrieved profile. Right: direct comparison, using the clear-sky product. Ordinary (ORD, blue) and robust Theil-Sen regressions (TS, red), Pearson (RPr) and Spearman (RSp) correlation coefficients are indicated. Points are coloured by satellite cloud fraction. 71

List of Tables

Table 2-1: Overview on the algorithms used in the Round Robin for SCDs for GOME-2A 11

Table 2-2: Overview on the algorithms used in the Round Robin for stratosphere-troposphere separation for GOME-2A. 12

Table 2-3: Overview on algorithms used in the Round Robin for tropospheric AMF calculations for GOME-2A. 12

Table 2-4: Spectral fit settings for GOME-2A NO_2 algorithms evaluated in this work. 13

Table 2-5: Trends in GOME-2A NO_2 slant column uncertainty estimates for the sub-periods before and after the second throughput test (September 2009) for the three algorithms.... 17

Table 2-6: Regions used for the NO_2 cloud analysis..... 29

Table 2-7: Algorithm choices for the tropospheric NO_2 CDR 41

Table 2-8: Summary of NO_2 sensitivity analysis for each of the retrieval tests. 42

Table 3-1: Summary of the 2 step HCHO background correction..... 47

Table 3-2. Comparison statistics of GOME-2A QA4ECV variant products vs bePRO MAX-DOAS in Xianghe, year 2012, no vertical smoothing. Variant products are: (1) prior=TM5, (2) prior=TM5, clear-sky assumption, (3) prior=CAMS reanalysis, (4) prior=MAGRITTE. 58

Table 3-3: Summary of HCHO sensitivity analysis for each of the retrieval tests..... 60

Executive Summary

The product Validation and Algorithm Selection Report (PVASR) provides a complete report of the results of the algorithm Round Robin intercomparison as the basis for further algorithms development during the Precursors CCI+ project.

The Round Robin was performed for three products: NO_2, HCHO, and SO_2. For all trace gases, the same approach was taken in the evaluation by analysing all critical aspects of the algorithms:

1. Slant column retrievals (fitting range, cross-sections, for SO_2 also fitting approach)
2. Postprocessing (background correction, stratospheric correction)
3. Air mass factor inputs
    a. Surface reflectance data bases
    b. Model a priori profiles
    c. Cloud parameters
4. Comparison of selected algorithm versions to limited validation data

The options evaluated are mostly based on results from the QA4ECV project, the current developments in the AC-SAF, and scientific retrievals performed by the participating groups. The aim was to combine the best of the available inputs and algorithms and to achieve the highest possible consistency among sensors and with existing long-term data records.

The main findings of the Round Robin are that over the last years, there has been significant development for all three products, leading to improved consistency between different algorithms for the same quantity when using the latest data sets and retrievals. This is in particular true for inputs such as the various cloud products, the surface reflectances, and the stratospheric corrections. Therefore, the choice of a particular option has often only limited impact on the quality of the climate data record produced.

For SO_2, the COBRA algorithm, which is a relatively new development, was selected as it provides the lowest noise and least bias in the data. Comparison with the existing DOAS and PCA algorithms shows good agreement in the spatial distributions and absolute values. Therefore, consistency with existing data sets can be achieved in spite of using a novel retrieval approach.

Based on the results of the Round Robin, choices have been made and documented on all major retrieval options and input data sets, and uncertainties related to these choices have been quantified. The details of the implementation are described in the respective ATBDs.

1. Purpose of Document

This document repots on the results of the Round Robin exercise for the NO_2, HCHO, and SO_2 products performed in the first phase of the Precursors CCI+ project. The aim of the Round Robin was to evaluate the differences between existing products, to study the sensitivity to specific choices made for the algorithms and input data and to take a decision on which options to choose for the climate data records produced in the project. While some limited validation with independent measurements was performed as part of the Round Robin, the main part of the validation will be done on the CDR products and will be reported in the Product Validation and Intercomparison Report (PVIR).

The document is organised by product, starting with the section on NO_2 (section 2), followed by HCHO (section 0) and finally SO_2 (section 0). Each section follows a similar approach covering first the algorithm part concerning the creation of slant columns and, if necessary, their corrections, followed by an evaluation of choices impacting the air mass factors and closing with validation. Each section then outlines the choices taken for the respective product. The details on the consolidated algorithms can be found in the ATBDs of the respective products.

2. NO_2

2.1. Data and Algorithms used

The development of a climate data record for tropospheric NO_2 can build on a strong heritage of existing operational and scientific retrievals of this quantity for all existing satellite sensors. The two data sets most relevant for the project are the QA4ECV data set covering GOME, SCIAMACHY, GOME-2A and OMI and the AC-SAF data set covering GOME-2 data.

The tables below summarize the specifications of the level-1 and ancillary data, as well as the processor versions used for the 3 algorithm steps in the retrieval of NO_2 data included in the Round Robin exercise.

Table 2-1: Overview on the algorithms used in the Round Robin for SCDs for GOME-2A

Algorithm	Level-1 data set	Fitting window	Fit parameters	Wavelength calibration	Reference
AC-SAF operational	v5-v6.3	425-450 nm	NO_2, O_3, H_2O_{vap}, O_2-O_2, Ring	sao2010_solref_vac	Valks et al. (2011)
AC-SAF prototype	v6.3	425-497 nm	NO_2, O_3, H_2O_{vap}, O_2-O_2, H_2O_{liquid}, Ring, Eta correction	sao2010_solref_vac	Liu et al. (2019)
QA4ECV v1.1	v5-v6.2	405-465 nm	NO_2, O_3, H_2O_{vap}, O_2-O_2, Ring, H_2O_{liquid},	Along with fit	Zara et al. (2018)

Table 2-2: Overview on the algorithms used in the Round Robin for stratosphere-troposphere separation for GOME-2A.

Algorithm	Method	Reference
AC-SAF prototype	STREAM	Beirle et al. (2016), Liu et al. (2019)
QA4ECV v1.1	Data assimilation of SCDs in TM5	Boersma et al. (2018)
QA4ECV v1.1	STREAM	Boersma et al. (2018)

Table 2-3: Overview on algorithms used in the Round Robin for tropospheric AMF calculations for GOME-2A.

Algorithm	AMF calculation	AMF wvl	Cloud parameters	Surface albedo	NO_2 a priori profile
AC-SAF prototype	Box AMF LUT with VLIDORT	461 nm	OCRA cloud fraction, ROCINN cloud pressure	GOME-2_MetOp-AB_MSC_025x025_surface_LER_v3.0.hdf5	TM5 1° ×1°
QA4ECV v1.1	Box-AMF LUT with DAK	437.5 nm	FRESCO effective cloud fraction and pressure	GOME2_MetOp-A_MSC_surface_LER_product.hdf5 (v1.6)	TM5 1° ×1°

2.2. Slant Column Intercomparison

The slant column retrieval is important since it dominates the overall retrieval uncertainty of total and tropospheric NO_2 columns as an initial step. In particular, within the context of the generation of the NO_2 climate data record, it is important to find an optimal spectral fit setting that has low NO_2 slant column uncertainties and reduces inter-instrument biases. Accordingly, we performed intercomparisons of existing NO_2 slant column retrievals for different GOME-2A algorithms and evaluated the temporal evolution of NO_2 slant column uncertainties along with the stability of the long-term data record.

Table 2-1 summarises the settings of GOME-2A NO_2 slant column retrieval from three existing products assessed in this Round-Robin activity. As shown in the table, there are differences in NO_2 spectral retrieval settings, including fitting window, treatment of slit function, wavelength calibration and other fitted parameters.

Table 2-4: Spectral fit settings for GOME-2A NO_2 algorithms evaluated in this work.

	QA4ECV (v1.2)	AC-SAF operational (GDP 4.8)	AC-SAF prototype
Fitting window	405 - 465 nm	425 - 450 nm	425 - 497 nm
Fitting method	Optical depth fit	Optical depth fit	Optical depth fit
Fitted parameters	NO_2, O_3, H_2O_{vap}, O_2-O_2, Ring, H_2O_{liq}, I_{offset}	NO_2, O_3, H_2O_{vap}, O_2-O_2, Ring, I_{offset}	NO_2, O_3, H_2O_{vap}, O_2-O_2, Ring, H_2O_{liq}, I_{offset}, Eta correction[1], resolution correction
Polynomial	4th order	3rd order	5th order
Reference spectrum	Daily solar irradiance	Daily solar irradiance	Daily solar irradiance
Slit fuction	Preflight	Preflight	Stretched preflight
Wavelength calibration	Along with fit	Chance and Kurucz, 2010	Chance and Kurucz, 2010
Reference	Zara et al. (2018)	Valks et al. (2011)	Liu et al. (2019)

We compared the GOME-2A NO_2 slant columns retrieved from the QA4ECV, AC-SAF operational and AC-SAF prototype algorithm. Figure 2.1 shows a time-series of monthly mean GOME-2A NO_2 slant column densities from the three different algorithms over the clean Pacific region for the period from 2007 to 2017. Since they were retrieved with different spectral fit settings, the monthly averages of NO_2 slant columns are not the same, but they are consistent and show similar value ranges before June 2015. However, after June 2015, GOME-2A NO_2 slant columns from the 405-465 nm window (used in the QA4ECV algorithm) show substantial sensitivity to the updated version 6.1 (and later versions) of the level 1b data, resulting in stronger seasonality and jumps in NO_2 slant columns over the clean background sector. The inconsistent temporal variability of GOME-2A NO_2 slant column retrievals for the QA4ECV algorithm compared to the AC SAF algorithms is identified not only in the Pacific region but also in biomass burning and anthropogenically polluted regions (see Figure 2.1 to Figure 2.3).

[1] Eta and Zeta from GOME-2 calibration key data (EUMETSAT, 2009)

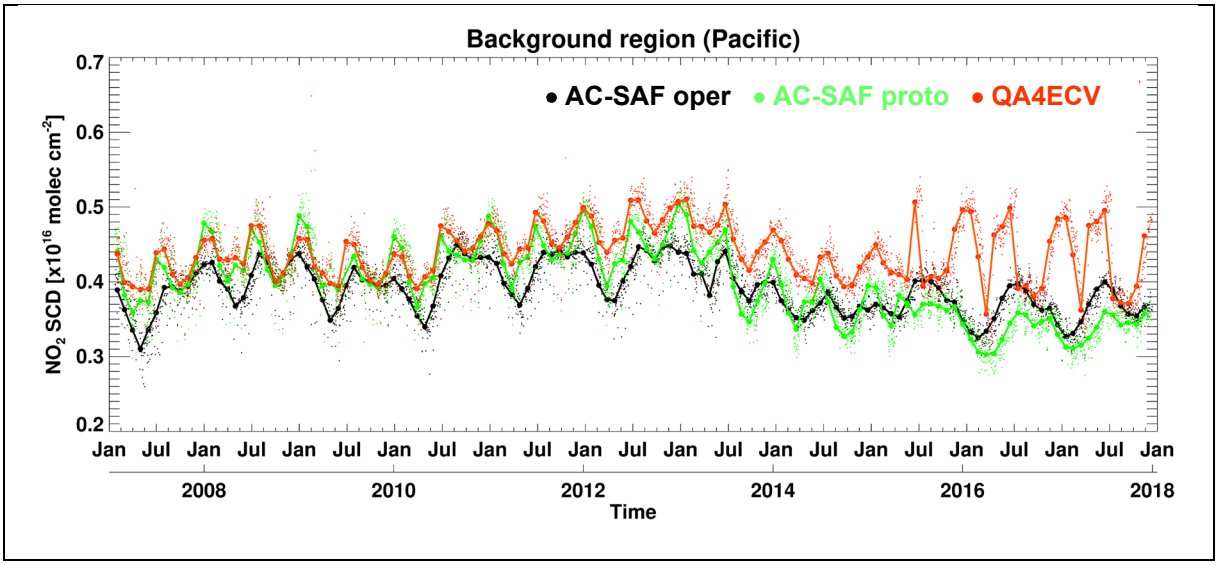


Figure 2.1: Time-series of monthly averages of NO_2 slant columns from three algirithms (AC-SAF operational (black), AC-SAF prototype (green), QA4ECV (orange)) over the clean Pacific region (20°S - 20°N, 110 - 150°W) for the period of 2007 to 2017.

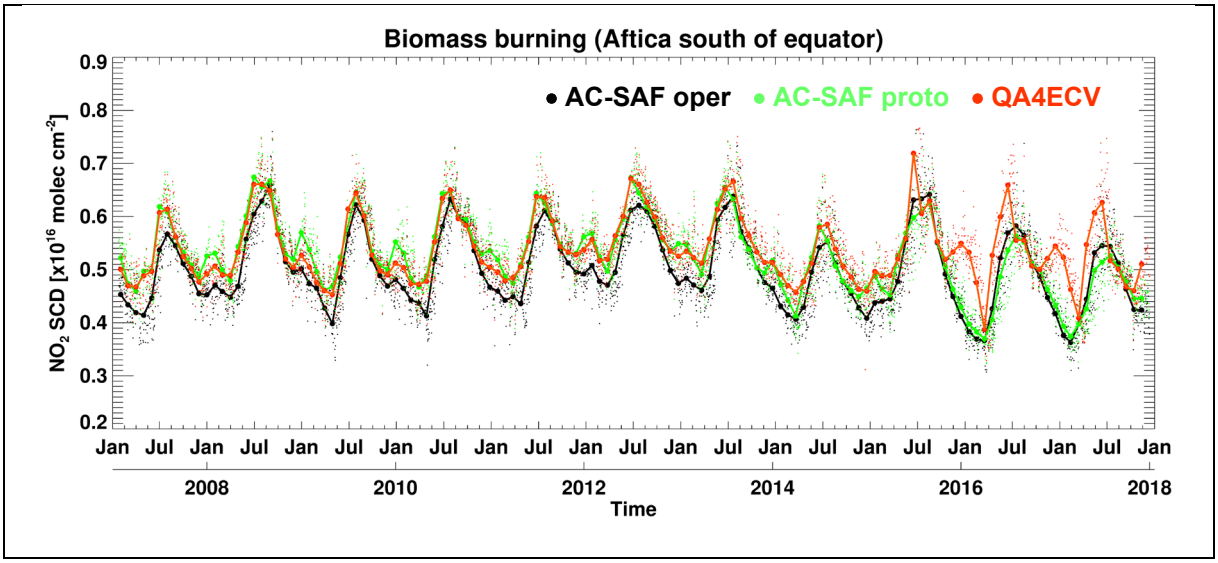


Figure 2.2: Same as Figure 2.1, but for sub-Saharan Africa (15 – 0°S, 10 – 35°E).

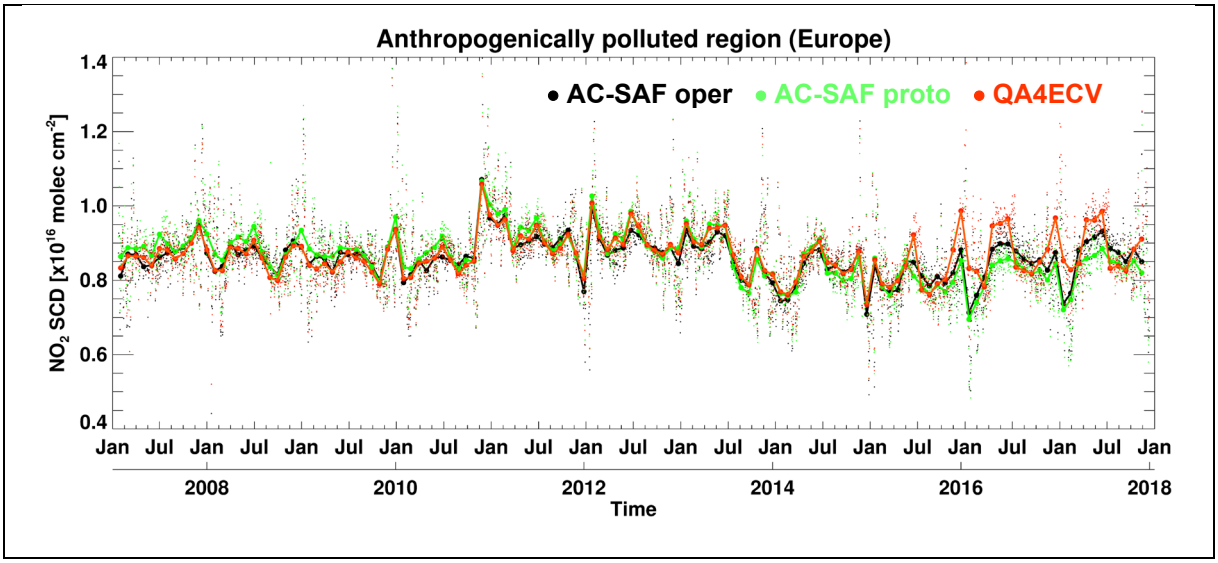


Figure 2.3: Same as Figure 2.1, but for anthropogenically polluted Europe (35 – 50°N, 10°W– 20°E).

To evaluate the NO_2 slant column uncertainties for the three different retrieval algorithms, a statistical analysis was performed using GOME-2A measurements over the clean Pacific region (60 °S - 60 °N; 150 - 180 °W). This area was divided into small boxes (2°x2°), and based on the spatial variability of NO_2 slant columns within each box, an estimate of the slant column uncertainties was obtained. This statistical approach can be used to quantify the NO_2 slant column uncertainties related to the instrument level-1 signal-to-noise and the spectral fit quality by limiting the contributions from other factors such as anthropogenic pollution sources and geophysical variability (Boersma et al., 2007; Richter et al., 2011, Zara et al., 2018).

Figure 2.4 illustrates the distribution of deviations of GOME-2A NO_2 slant columns from the box-averaged slant columns for the three different algorithms. The histogram contains all boxes within the reference Pacific region from the 15th day of each month spanning 2007 to 2017. These histograms follow a Gaussian distribution, reflecting the assumption that the variability of slant columns over the clean region is predominantly attributed to random errors originating from measurement noise and imperfections in the spectral fitting. Therefore, the width (1σ) of the Gaussian function fitted to the observed distributions serves as an indicator of the statistical uncertainty in the NO_2 slant columns for the different algorithms. The estimated NO_2 slant column uncertainties for each algorithm are 0.59 x 10^{15} for QA4ECV, 0.65 x 10^{15} for AC-SAF operational, and 0.47 x 10^{15} molec cm^{-2} for AC-SAF prototype algorithm, respectively. Mainly driven by the use of a wider fitting window of 425-497 nm, which improves the signal-to-noise ratio by including more NO_2 absorption structures, the AC-SAF prototype shows the smallest slant column uncertainties and the best spectral fit quality compared to the QA4ECV and AC-SAF operational algorithm, in line with earlier findings on the choice of fitting window (Richter et al., 2011).

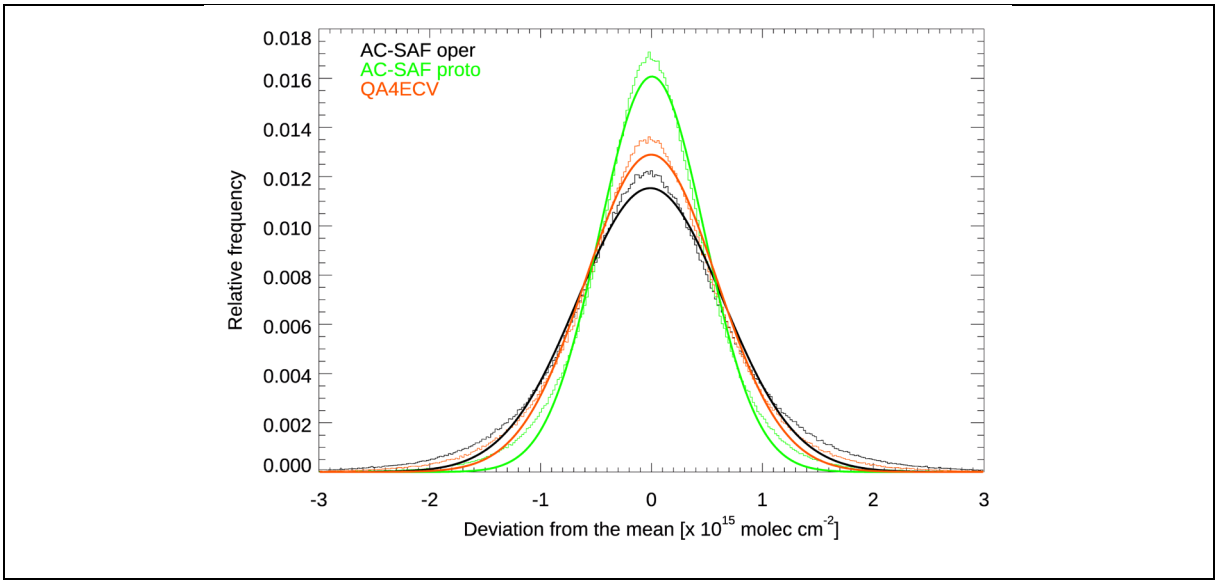


Figure 2.4: Distribution of the deviation the GOME-2A NO_2 slant columns from the mean slant column within each small-box in a histogram for the AC-SAF opertational (black), AC-SAF prototype (green), and QA4ECV (orange) algorithms. The histogram includes contributions from all boxes within the Pacific reference sector (60°S – 60°N, 150 – 180°E) from the 15th day of each month in the period 2007-2017. The fitted Gaussian function is also plotted as solid line.

We also investigated the temporal evolution of statistical NO_2 slant column uncertainties for the three GOME-2A products. Figure 2.5 depicts how NO_2 column uncertainties develop over time, and how instrument degradation affects the stability of long-term climate data records for each algorithm. All three algorithms show a positive trend due to the increase in measurement noise over time. A relatively fast increase rate of the GOME-2A NO_2 slant column uncertainties is found for all three products until the second throughput test in September 2009. The decline in spectral fitting quality appears to proceed at a notable annual increase rate of 5.2 %, 7.1 % and 7.8 % per year for the AC-SAF prototype, QA4ECV, and AC-SAF operational algorithm, respectively, relative to the mission start. However, the second throughput test, which initially resulted in an additional loss of signal-to-noise, has positive effects on the quality of the level-1 data by weakening the long-term drift of the instrument's slit function (EUMETSAT, 2012; Munro et al, 2016). Consequently, after the second throughput test, GOME-2A NO_2 slant column uncertainties increase at a slower pace until 2019, with a trend of 0.6 %, 1.5 % and 2.0 % per year for the AC-SAF operational, AC-SAF prototype, and QA4ECV algorithm, respectively. From 2020 until the end of its lifespan (not shown in Figure 2.5) the quality of GOME-2A NO_2 fits deteriorated rapidly due to a strong degradation of the level-1 data, resulting in a sharp increase in NO_2 slant column uncertainties at a rate of 17.5 – 17.8 % year $^{-1}$. This suggests that the GOME-2A NO_2 retrievals after 2019 should be excluded from the NO_2 climate data record development.

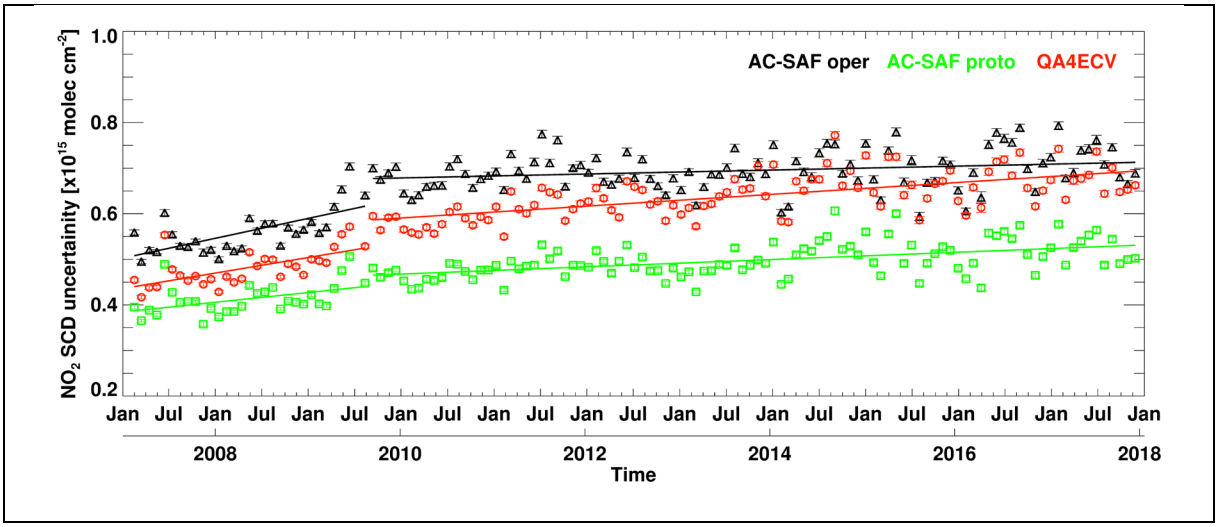


Figure 2.5: Temporal evolution of GOME-2A NO_2 slant column uncertainties from the AC-SAF operational (black triangle), AC-SAF prototype (green square), and QA4ECV (orange circle) algorithms using deviations of NO_2 slant columns from box mean values over the Pacific reference sector. The solid lines are linear regressions fitted to the data.

Table 2-5: Trends in GOME-2A NO_2 slant column uncertainty estimates for the sub-periods before and after the second throughput test (September 2009) for the three algorithms.

	QA4ECV	AC-SAF operational	AC-SAF prototype
Before the second throughput test	7.1 % $year^{-1}$	7.8 % $year^{-1}$	5.2 % $year^{-1}$
After the second throughput test	2.0 % $year^{-1}$	0.6 % $year^{-1}$	1.5 % $year^{-1}$

Lastly, we investigated the impact of different level 1b data versions on the GOME-2 NO_2 slant column retrievals for the AC-SAF prototype algorithm. Figure 2.6 presents the time-series of geometric NO_2 vertical columns, calculated as NO_2 slant columns divided by geometric air mass factors (AMFs), retrieved using the offline level 1b (v5.3-v6.2) and the reprocessed level 1b v6.3 data over the clean Pacific region from 2007 to 2017. Here, the GOME-2A NO_2 product retrieved using the reprocessed level 1b v6.3 is denoted as 'RPRO', while the one using the offline level 1b data is labelled as 'OFFL'. The variations in GOME-2 NO_2 slant columns are primarily driven by the azimuthal dependence of solar irradiance data from the solar diffuser on board the instrument. As the newly derived diffuser key-data is applied in the reprocessed level 1b v6.3 data, GOME-2A RPRO NO_2 columns shows more consistent seasonal variations compared to GOME-2A OFFL NO_2 columns. As the offline level 1b version approaches version 6.3, the differences between the RPRO and OFFL becomes smaller, approaching 0. The use of reprocessed level 1b data not only improves the consistency in the seasonal and annual variability within an individual instrument's NO_2 data record, but also the inter-instrument consistency between the NO_2 column retrieved from GOME-2A and GOME-2B (not shown).

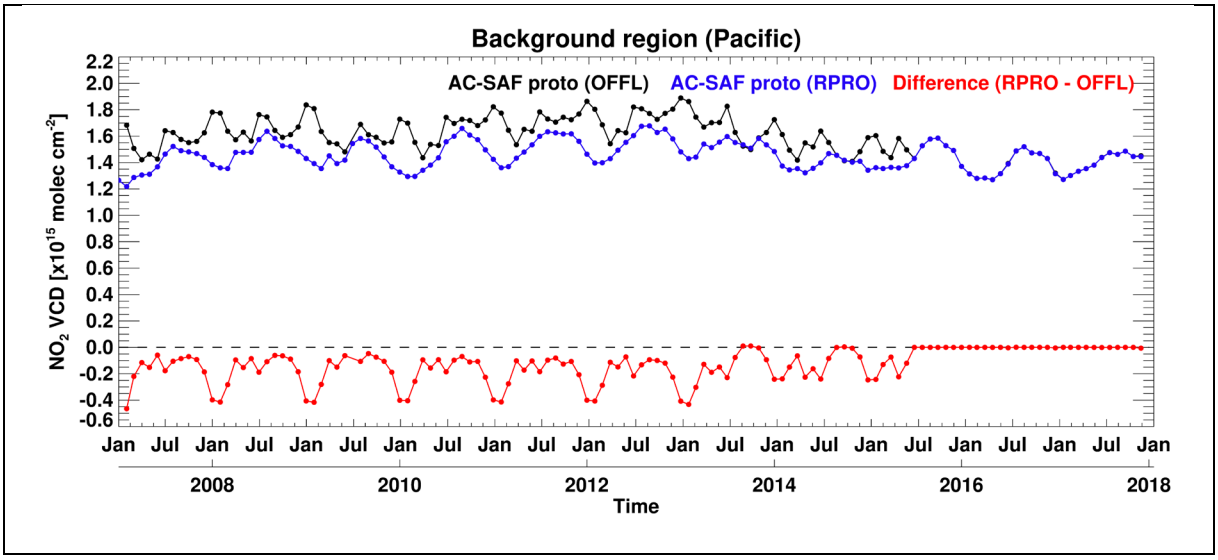


Figure 2.6: Time-series of GOME-2A geometric NO_2 columns from the AC-SAF prototype algorithm with offline (black) and reprocessed (blue) level 1b v6.3 data, along with the differences in geometric NO_2 columns between the two products (red).

Overall, the Round Robin exercise confirms that the AC-SAF prototype algorithm yields relatively small uncertainties and a high level of robustness in GOME-2 NO_2 slant column data. This retrieval algorithm, in conjunction with the level 1b v6.3 data, also significantly reduces the inter-instrument biases between GOME-2 sensors. Therefore, it is recommended for the development of integrated long-term NO_2 climate data records in the ESA CCI+ ECV Precursor project.

2.3. Stratosphere Troposphere Separation

Accurate correction of stratospheric nitrogen dioxide is important for long-term satellite NO_2 data records for several reasons. NO_2 is situated in the troposphere and in the stratosphere, and both contribute to satellite-retrieved signals. A primary interest of satellite NO_2 records is their use for atmospheric chemistry research, as well as for environmental and public health purposes, and accurate separation of stratospheric NO_2 from tropospheric NO_2 helps to obtain more precise retrievals of the tropospheric NO_2 column. The correction is also important to allow proper validation with ground-based measurements, which have different sensitivities to tropospheric and stratospheric NO_2 column amounts (e.g., Verhoelst et al., 2021).

Here we compare different stratospheric correction methods to assess the impact of algorithm settings and inputs as well as methods on the stratospheric NO_2 levels. We evaluate the differences (1) for a small number of specific days, in order to obtain a good understanding of the differences at small (daily) spatio-temporal scales, and (2) for the entire data record to assess aspects of robustness and sensor degradation in the context of trend analysis.

2.3.1 Comparison of stratospheric corrections for individual days

Our evaluation focused first on day-by-day comparisons of stratospheric NO_2 column fields from 3 different products available to us:

(1) QA4ECV data assimilation – this is the approach for TEMIS GOME-2 and OMI retrievals as well as the operational TROPOMI NO_2 retrieval as described in Boersma et al. (2018), and van Geffen et al. (2022).
(2) QA4ECV STREAM – this is an alternative to data assimilation also included in the QA4ECV data products and based on identical NO_2 slant columns as (1). Based on Beirle et al. (2016), and reported on for OMI in Boersma et al. (2018).
(3) AC-SAF STREAM – this is the recently improved prototype product from DLR, using updated NO_2 slant columns (as described in section 2.2, and an improved STREAM method described in Liu et al. (2021).

Figure 2.7 shows a polar-view panel with GOME-2 NO_2 columns on 5 November 2008. The upper left panel shows the total NO_2 column from the QA4ECV retrieval, with clear contributions from the troposphere visible over North America, western Europe, and eastern Asia. The other 3 panels in the upper row show the stratospheric separation from QA4ECV data assimilation (product (1) above), QA4ECV STREAM (product (2)), and AC-SAF STREAM (product (3)) orbits. All figures show the lowest NO_2 columns close to the terminator. The data assimilation stratospheric NO_2 field is lower than STREAM NO_2, despite the same NO_2 slant columns being used as input. This points at methodological differences between the two approaches, with STREAM attributing a higher portion of the slant column signal to the stratosphere than data assimilation. The AC-SAF (prototype) STREAM stratospheric NO_2 column is somewhat lower that (QA4ECV) STREAM, reflecting lower NO_2 slant column inputs.

The lower three panels show the tropospheric NO_2 columns, defined here as the tropospheric slant column divided by the geometrical air mass factor (AMF) to avoid AMF differences between QA4ECV and AC-SAF further influencing the comparison. We see that tropospheric enhancements occur in the same regions with all three separation methods, which indicates that both stratospheric corrections function as designed. However, tropospheric columns retrieved with STREAM more often show negative values, over the oceans as well as over the continents. To investigate this further, we analyzed the zonal variability in column NO_2 at mid-latitudes (45°N) on the same day.

Figure 2.8 shows substantial zonal variability in the total NO_2 column at 45°N on 5 November 2008. The stratospheric NO_2 columns from data assimilation and notably AC-SAF STREAM generally show less variability with longitude because they effectively screen out the total column enhancements coming from the troposphere over the polluted regions around the world. Outside eastern Asia, AC-SAF STREAM stratospheric NO_2 is 0.1-0.2 10^{15} molec. cm^{-2} higher than QA4ECV data assimilation. The higher stratospheric columns therefore lead to lower AC-SAF tropospheric columns over the polluted areas, and over the regions dominated by background NO_2 including the Pacific and Atlantic Ocean, and much of Asia. The other days investigated in 2008 and in 2012 show a similar pattern, but the conclusions from the 4 days investigated at the end of the GOME-2A mission deviate. We will next evaluate the reasons for this.

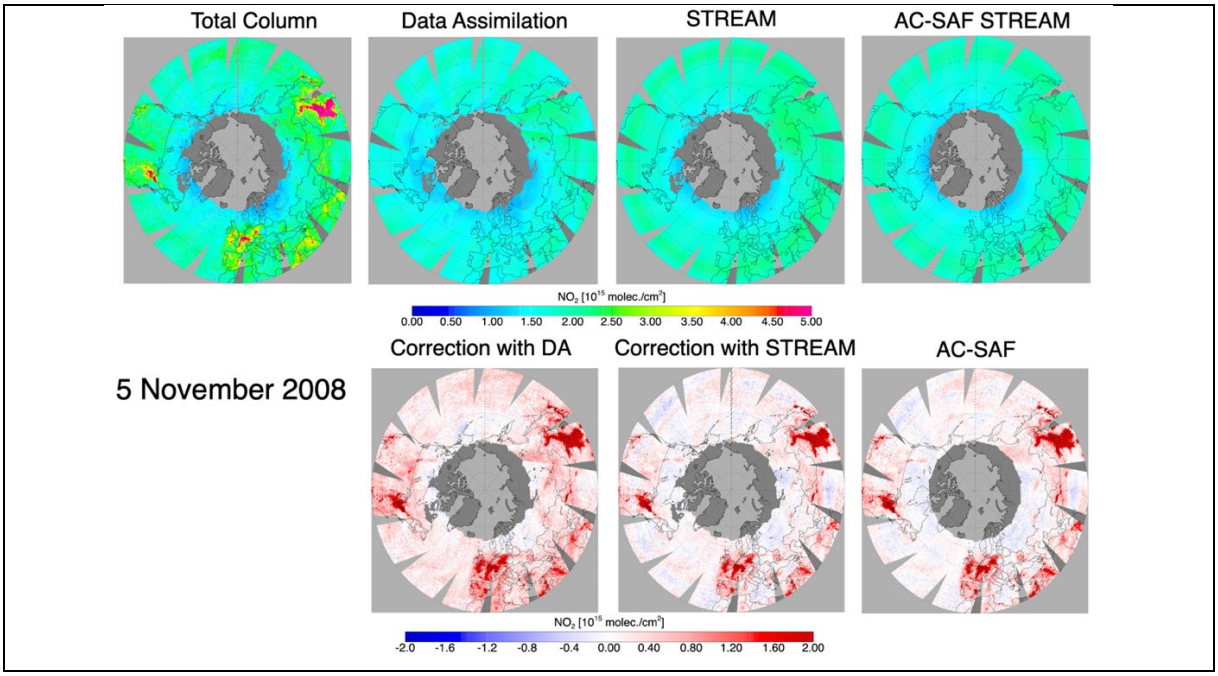


Figure 2.7: Comparison of GOME-2A stratospheric corrections. Upper left: total NO_2 column from QA4ECV. Upper middle: stratospheric NO_2 column from data assimilation and STREAM in QA4ECV data product. Upper right: stratospheric NO_2 from STREAM in AC-SAF product. Lower panels: tropospheric NO_2 column in QA4ECV product (lower left), in QA4ECV when using STREAM (lower center), and AC-SAF when using STREAM (lower right).

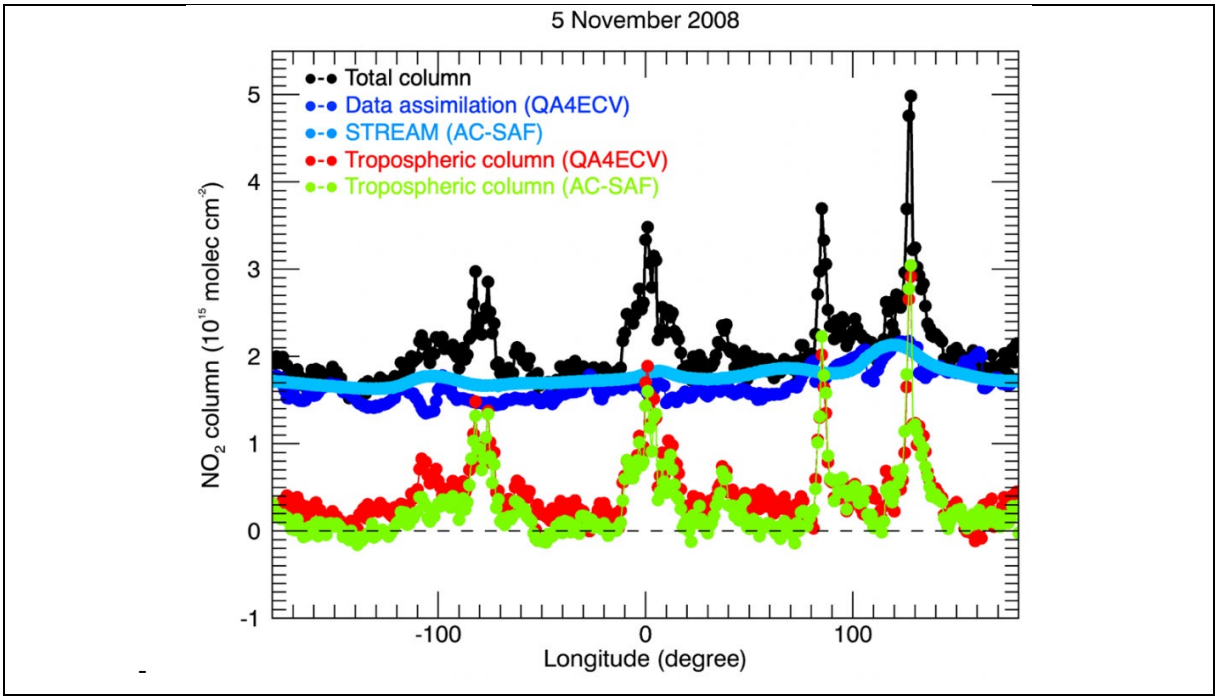


Figure 2.8: Zonal average NO_2 columns on 5 November 2008 between 44-46°N from GOME-2A. The tropospheric columns shown are from the QA4ECV and AC-SAF data products, calculated as the tropospheric slant column divided by the geometrical air mass factor.

2.3.2 Comparison of stratospheric corrections in the long-term

Figure 2.9 below shows a comparison of stratospheric and tropospheric NO_2 columns from the QA4ECV and AC-SAF prototype products over the Pacific sector for the period 2007-2017

(11 full years). We observe close agreement between the QA4ECV and AC-SAF stratospheric NO_2 columns up until July 2013, the month when the swath width of the GOME-2A sensor was reduced from 1920 to 960 km. After July 2013, stratospheric NO_2 from QA4ECV is higher by 0.0-0.3 10^{15} molec. cm^{-2}, caused by the somewhat higher QA4ECV NO_2 slant column densities also observed in that period (see section 2.2). After June 2015, when the level-1b data of GOME-2A was updated from v6.0 to 6.1, QA4ECV stratospheric NO_2 is higher by 0.3-0.7 10^{15} molec. cm^{-2}, which is explained by the 0.5-1.5 10^{15} molec. cm^{-2} higher QA4ECV NO_2 slant columns found after June 2015 compared to the AC-SAF slant columns. It is obvious that the stratospheric NO_2 estimates with both approaches closely follow changes in the NO_2 slant columns input to the stratospheric correction methods.

Figure 2.9 also shows the impact of the stratospheric corrections on QA4ECV (red) and AC-SAF prototype tropospheric columns (green) over the remote tropical Pacific. Despite lv1-calibration related changes in the NO_2 slant columns, the QA4ECV tropospheric columns remain relatively stable at values of 0.1-0.3 10^{15} molec. cm^{-2} throughout the 2007-2017 period. The AC-SAF prototype tropospheric columns drop to small but persistent negative values after the swath was reduced in July 2013. This is caused by a negative trend in stratospheric NO_2 columns in the AC-SAF product over the period 2013-2017: while QA4ECV stratospheric NO_2 have an insignificant positive trend (+0.01 10^{15} molec. cm^{-2} yr^{-1}), the trend in AC-SAF stratospheric columns is distinctly negative (-0.06 10^{15} molec. cm^{-2} yr^{-1}) in the same period. The absence of a trend in QA4ECV tropospheric columns over the background Pacific confirms the finding of Zara et al. (2018) (who analyzed GOME-2A trends for the period 2007-2016) that the combined uncertainties from spectral fitting and stratospheric corrections allow tropospheric NO_2 trend detection with precision of better than 0.1 10^{15} molec. cm^{-2} dec^{-1} over regions where AMFs are robust.

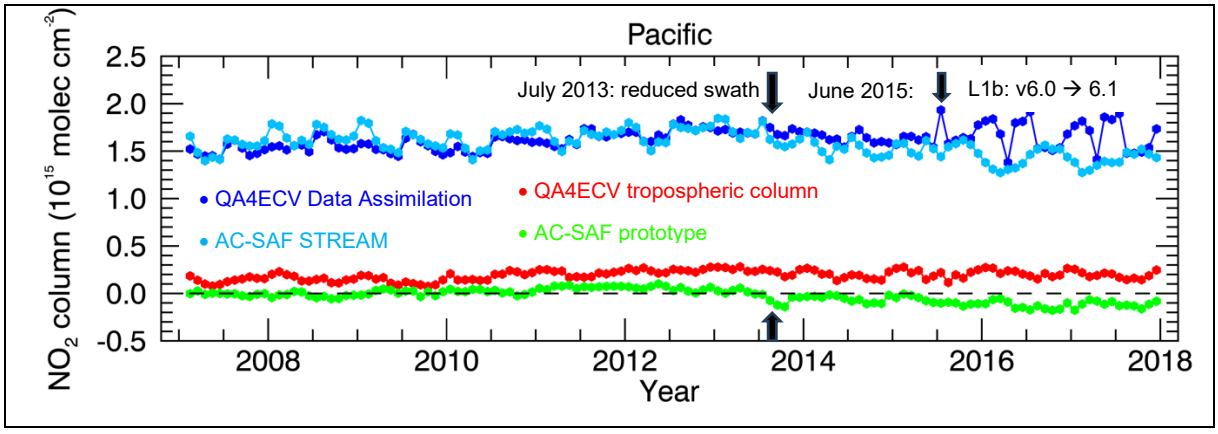


Figure 2.9: Timeseries of GOME-2A stratospheric and tropospheric NO_2 over the Pacific sector.

Figure 2.10 shows a similar comparison between QA4ECV and AC-SAF NO_2 columns over sub-Saharan Africa. Stratospheric NO_2 column values are similar as over the Pacific region, but the African region is characterized by episodic biomass burning, as indicated by tropospheric NO_2 columns peaking around 1.5 10^{15} molec. cm^{-2}. We observe again that QA4ECV and AC-SAF stratospheric columns are similar until July 2013, and that AC-SAF stratospheric NO_2 columns decrease after July 2013. Tropospheric NO_2 columns from QA4ECV and AC-SAF also agree well

until July 2013, but after that AC-SAF tropospheric NO_2 decrease relative to the values in the 2007-2013 period, and relative to QA4ECV columns.

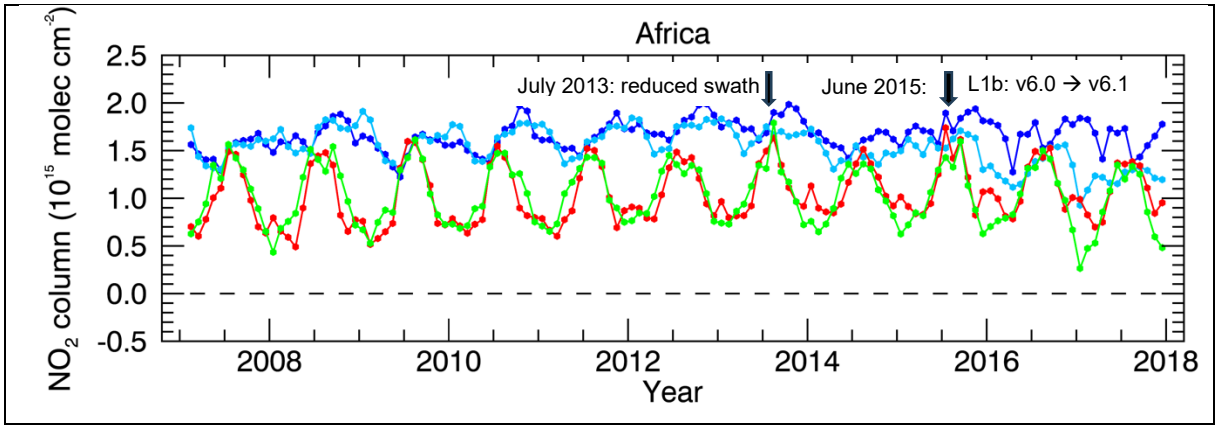


Figure 2.10: As Figure 2.9 but for sub-Saharan Africa.

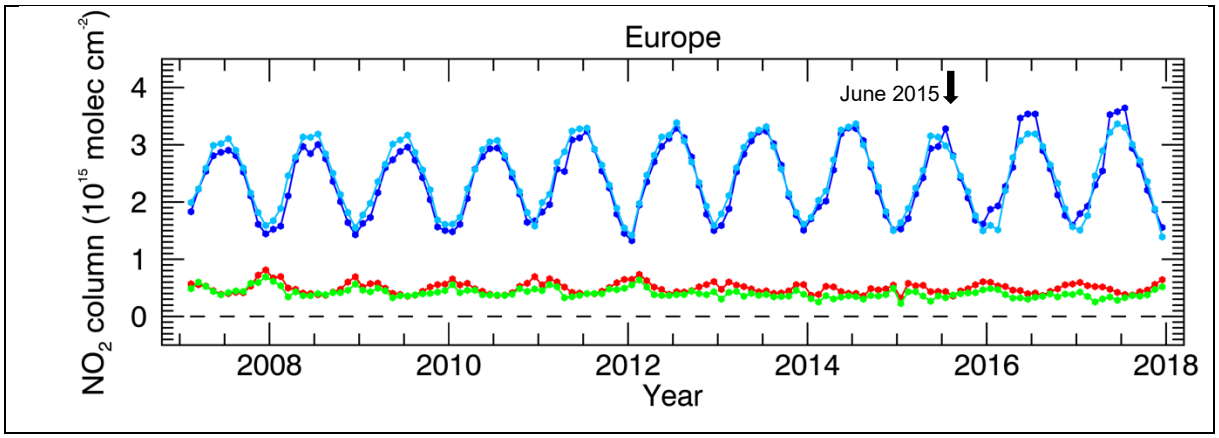


Figure 2.11: As Figure 2.9, but for Europe. Tropospheric NO_2 column values (indicated in red and green) have been divided by a factor of 4.

The consistency and stability of the stratospheric and tropospheric NO_2 columns was also evaluated over Europe and China, two regions with high, year-round anthropogenic NO_x emissions. In Figure 2.11 we observe that the consistency of stratospheric NO_2 over Europe is high until June 2015, with slightly lower QA4ECV than AC-SAF stratospheric NO_2 columns. After June 2015, the discrepancies between QA4ECV and AC-SAF increase, especially in summer, as also observed over the Pacific and over Africa. Tropospheric NO_2 columns (here divided by a factor of 4 for visualization purposes) retrieved with the data assimilation approach (QA4ECV: red) are generally higher and show a stronger seasonal cycle than those from AC-SAF (green).

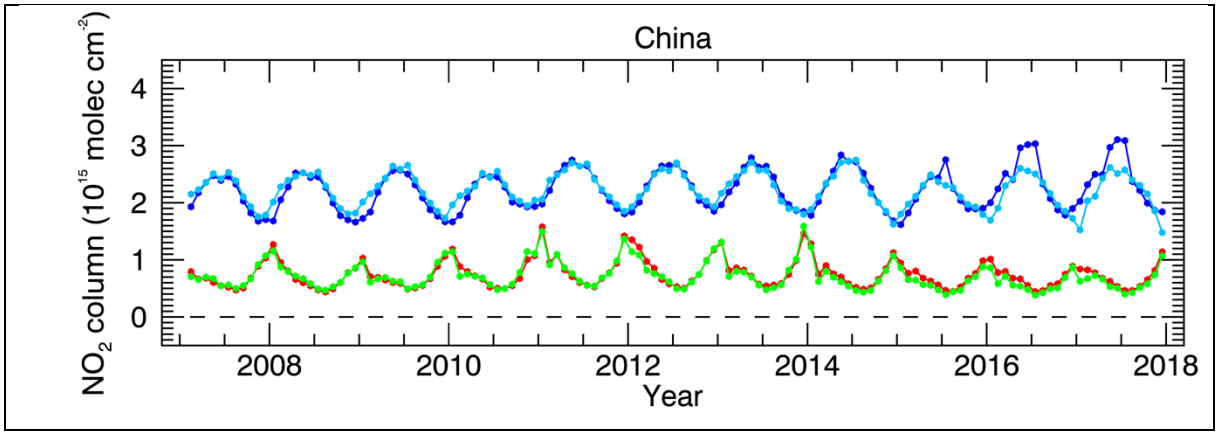


Figure 2.12: As Figure 2.9, but for eastern China. Tropospheric NO_2 column values (indicated in red and green) have been divided by a factor of 4.

Figure 2.12 shows a similar picture over eastern China as over Europe. Stratospheric NO_2 columns start to deviate after June 2015, in line with the higher NO_2 slant columns feeding the QA4ECV retrieval in comparison to the AC-SAF prototype retrieval, which uses more recent NO_2 slant columns from recalibrated level-1 data. Again, tropospheric NO_2 columns retrieved with the data assimilation approach (QA4ECV: red) are slightly higher and show a stronger seasonal cycle than those from AC-SAF (green). But in relative terms, the QA4ECV and STREAM tropospheric NO_2 columns are more consistent between QA4ECV and STREAM than over other regions of interest. This might (partly) be related to the larger tropospheric NO_2 contributions to the total column over eastern China (with values ranging from 2 to 6 10^{15} molec. cm^{-2}). When total columns are dominated by the stratospheric signal, errors on the order of 0.2 10^{15} molec. cm^{-2} in the separation method can easily lead to proportionally large errors in the tropospheric residual, because the (background) tropospheric NO_2 columns are of similar magnitude. However, when the tropospheric NO_2 columns are much larger, errors of 0.2 10^{15} molec. cm^{-2} in the separation method proportionally have less impact.

2.4. AMF Impacts

2.4.1. LER

One parameter which is of importance for the calculation of air mass factors is the assumed surface reflectance. In most retrievals, a Lambertian Equivalent Reflectivity (LER) is used, which neglects the directional dependency of surface reflectance as described by the full BRDF. More recently, Directional LERs have been produced for the GOME-2 and TROPOMI instruments by computing LERs as a function of the viewing zenith angle of the satellite (Tilstra et al., 2017, Tilstra et al., 2023). As the instruments are in sun-synchronous orbits, the viewing geometry changes only slowly with season, and, therefore, DLER includes a large part of the directional effect if retrieved from measurements of the same instrument.

Four different reflectivity data sets have been evaluated for the NO_2 Round Robin at the wavelength relevant for NO_2 (close to 440 nm):

- The GOME-2 minimum reflectivity v4.0 (Tilstra et al., 2017)
- The GOME-2 minimum directional reflectivity v4.0 (Tilstra et al., 2017)

- The OMI minimum reflectivity (Kleipool et al., 2008)
- The TROPOMI minimum surface reflectivity v1.1 (Tilstra et al., 2023).

In order to evaluate the effect of using different surface reflectances in the NO_2 retrieval, four days of GOME-2A data, distributed over the year 2008 have been selected. Tropospheric AMFs have been computed and compared, using the GOME-2 minimum reflectivity as reference. The relative changes in NO_2 AMF were computed as $(M_{test} - M_{G2min})/M_{G2min}$ where M is the air mass factor.

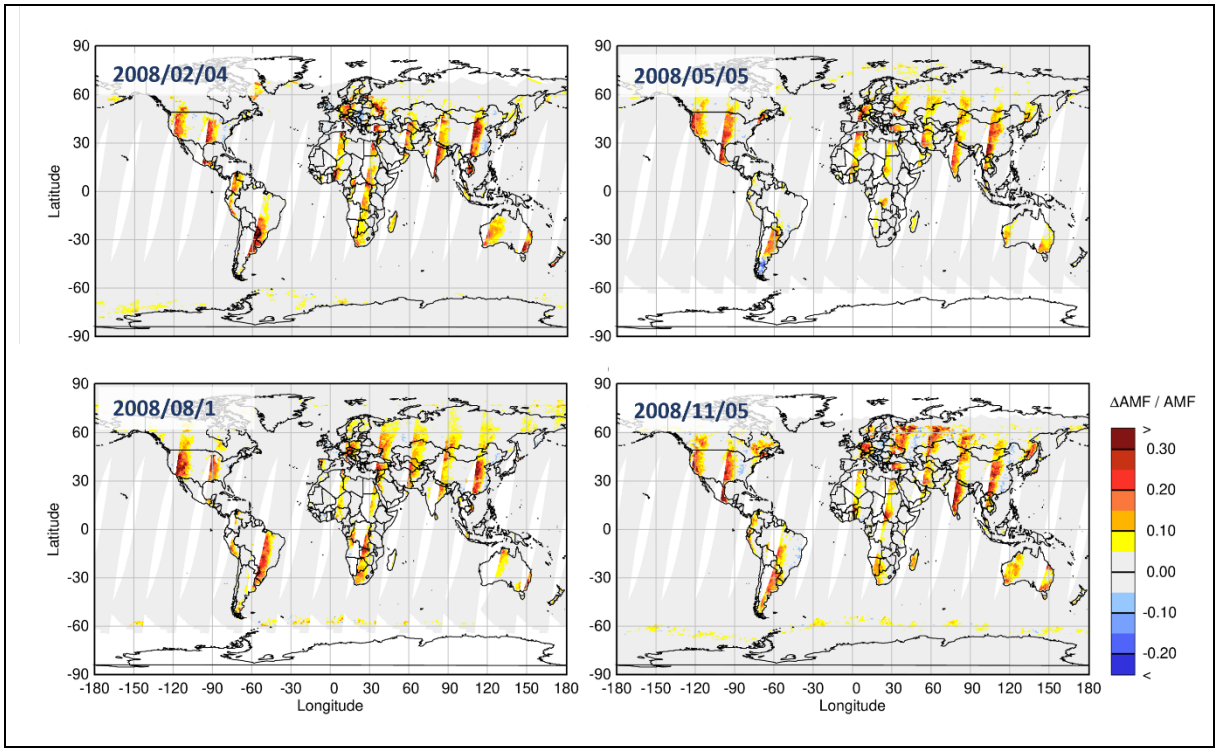


Figure 2.13: Relative change in the tropospheric NO_2 clear-sky AMF when using the GOME2 DLER min instead of the GOME2 LER min surface reflectance on GOME-2A data for 4 selected days.

In Figure 2.13 the relative change in tropospheric NO_2 AMF is shown when replacing the GOME-2 minimum reflectivity by its DLER version. The result is a systematic increase in AMF in the western part of the orbits over land, with little seasonal variability. While the magnitude of the effect can be up to 20% for individual scenes, it will be smaller on average because of the moving scan pattern and because of clouds.

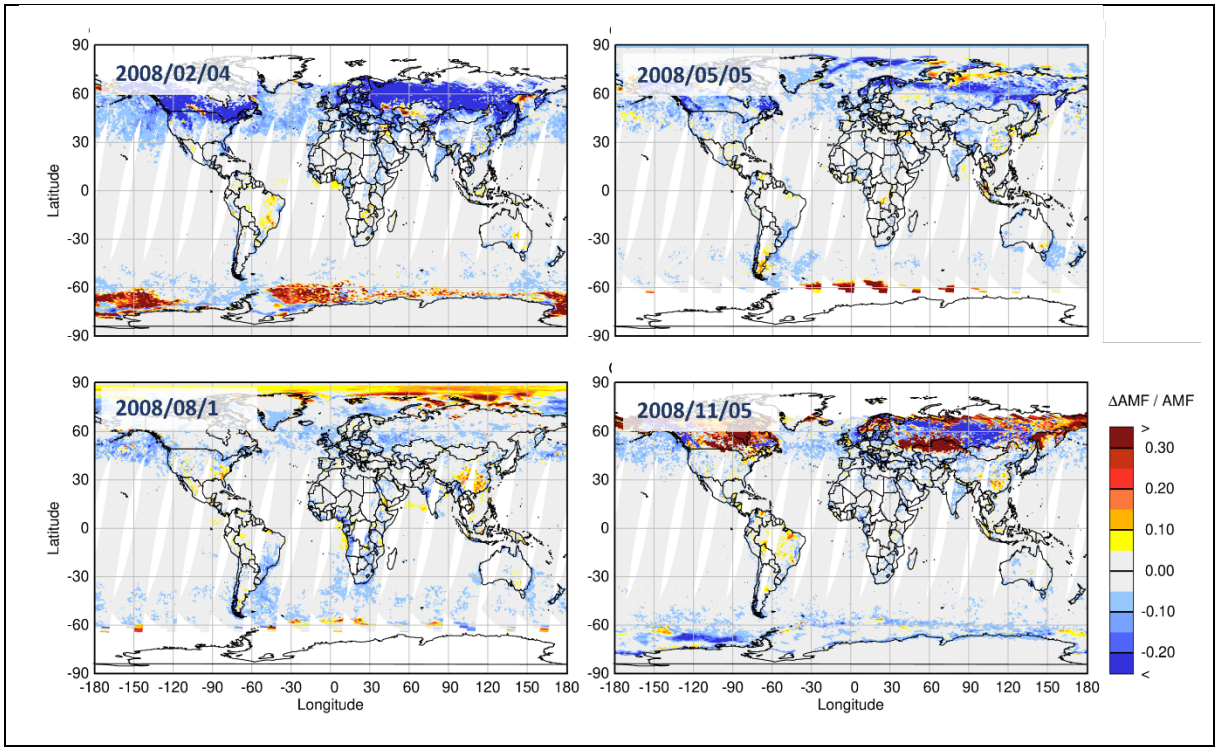


Figure 2.14: Relative change in the tropospheric NO_2 clear-sky AMF when using the OMI min instead of the GOME2 LER min surface reflectance on GOME-2A data for 4 selected days.

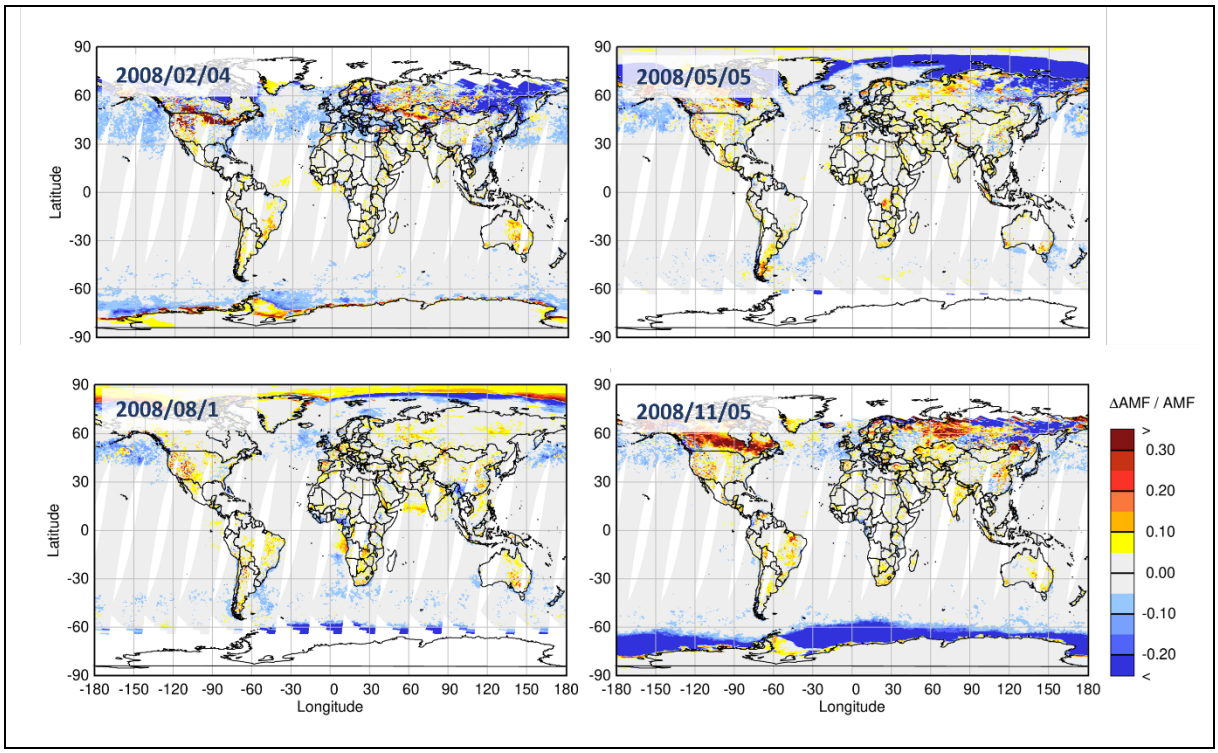


Figure 2.15: Relative change in the tropospheric NO_2 clear-sky AMF when using the S5P LER min instead of the GOME2 LER min surface reflectance on GOME-2A data for 4 selected days.

When evaluating the relative changes for a switch to the OMI minimum LER (see Figure 2.14), larger variations are found, sometimes up to and in excess of 30%. Both increases and decreases are visible. From the spatial pattern it is apparent, that most of the large differences

are linked to snow- and ice-covered regions. Here, the differences in overpass time, spatial resolution, and the time period of measurements used for the creation of the data base can have a large impact. Differences in the approach used to derive the minimum reflectivity could also play a role. For most of the tropospheric NO_2 observations, these regions are of less relevance, and often not considered at all.

Similar conclusions can be drawn when considering the change in AMF for a switch to the TROPOMI minimum LER (Figure 2.15). Relatively large changes are observed over snow- and ice-covered regions, often with similar sign and spatial pattern as already found for a switch to the OMI LER. This points at a systematic difference between these two surface reflectances and the GOME2 LER data set. In contrast to the results found for the OMI LER, the TROPOMI LER v1.1 leads to overall larger NO_2 AMFs by up to 10%.

Summarising the findings for the LER dependence of the tropospheric NO_2 air mass factor, large differences are found of snow- and ice-covered regions, but small differences over most of the polluted areas (< 10%). Some of the differences are linked to geophysical parameters (for example the shipping lanes) and are therefore systematic. This will have to be considered in the uncertainty budget. The directional effect as shown in the GOME-2 DLER data is considerable (up to 20%) and should be included wherever possible. As all calculations shown here were made for cloud-free scenes, the effect on real data will be smaller. Also, aerosols in the atmosphere tend to reduce the effect. Future work will have to cover several aspects, including

- Consistent treatment of snow / ice covered regions in the NO_2 products from all satellite instruments, possibly following the TROPOMI example, which is the most developed
- Consistent treatment of the directional dependence of surface reflectance in cloud products and AMF calculations, again for all instruments

Evaluation of the effect of realistic aerosol loading on the importance of the directional dependence of surface reflectance.

2.4.2. **A priori profiles**

UV/vis satellite observations do not provide information on the vertical distribution of weak absorbers such as NO_2. As the measurement sensitivity depends on the altitude at which an absorber is located in the atmosphere, a priori information on the vertical NO_2 distribution is needed to compute appropriate air mass factors. This information is usually taken from 3d-CTM model runs, which provide a good representation of the spatial and temporal distribution of tropospheric NO_2. However, depending on the emission data used, the spatial resolution of the model, and the chemical and dynamical schemes employed, modelled NO_2 distributions can differ significantly between models.

Therefore, three different sources of NO_2 a priori data were evaluated in the NO_2 Round Robin:

- The CAMS o-suite forecast (Flemming et al., 2015)
- The CAMS reanalysis (Iness et al., 2019).
- The TM5-MP model run distributed with the TROPOMI data (Huijnen et al 2010; Williams et al, 2017)

As a first test, the tropospheric NO_2 columns simulated by the three models are compared. Here, we take the CAMS o-suite as the arbitrary reference. In Figure 2.16, differences between TM5-MP and CAMS o-suite as well as the CAMS reanalysis and the CAMS o-suite are shown for Janaury 2022. Large differences are observed over polluted regions, with both CAMS reanalysis and TM5-MP showing smaller NO_2 columns over the Arabian countries than the CAMS o-suite. Over China, TM5-MP sees less, and CAMS reanalysis more NO_2 than the CAMS o-suite. There are also other regions where large differences (> $5x10^{15}$ molec cm^{-2}) are found. Differences in summer months (not shown) are smaller, in line with overall smaller NO_2 columns.

As shown in Figure 2.17, differences of a similar magnitude are found between GOME2-C retrievals and the CAMS o-suite and TM5-MP, respectively. GOME2-C reports less NO_2 over China and the Arabian countries than the CMS o-suite, but more than TM5-MP. There are also other differences, which could in part be linked to the stratospheric correction applied (STREAM-B, an IUP-Bremen variant of the STREAM algorithm by Beirle et al., 2016).

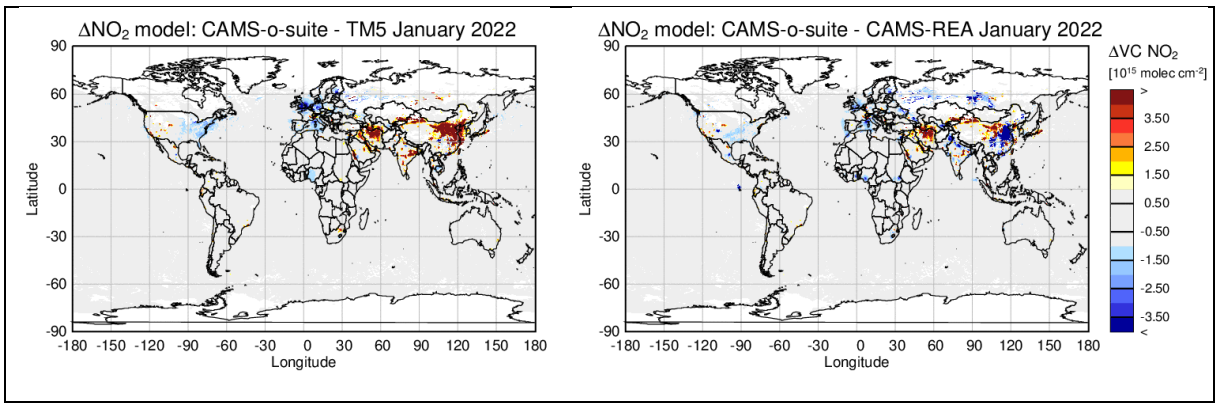


Figure 2.16: Differences in tropospheric NO_2 columns between different models. Shown are the differences between the CAMS o-suite and the TM5-MP (left) and the CAMS-reanalysis (right), both for Janaury 2022

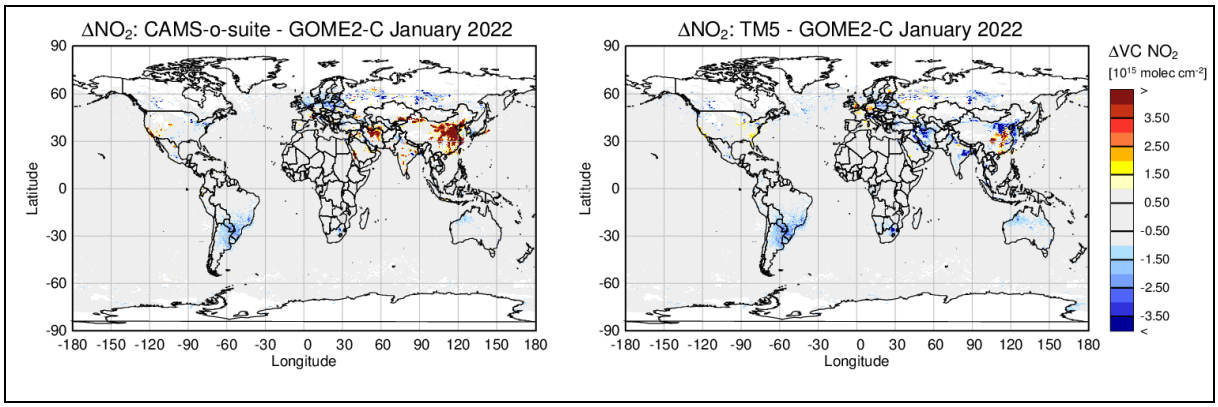


Figure 2.17: Differences in tropospheric NO_2 columns between GOME-2C retrievals and CAMS o-suite (left) and TM5-MP (right) for January 2022.

For the retrieval of tropospheric NO_2, it is not relevant how large the NO_2 columns of the models are, but rather what the AMFs are which result from the vertical shape of the NO_2 profiles simulated by the models. In Figure 2.18, ratios of tropospheric NO_2 AMFs are shown for CAMS o-suite and TM5 respectively CAMS reanalysis. Large changes in excess of 30% are

apparent over many areas, but much smaller changes over pollution hotspots. This indicates that the models predict a different vertical distribution of NO_2 under background conditions, presumably because of differences in spatial resolution and vertical mixing schemes. An exception may be the differences over shipping lanes, which could be linked to differences in the emission inventories used. As before, this points at systematic differences to be considered in an advanced uncertainty estimation.

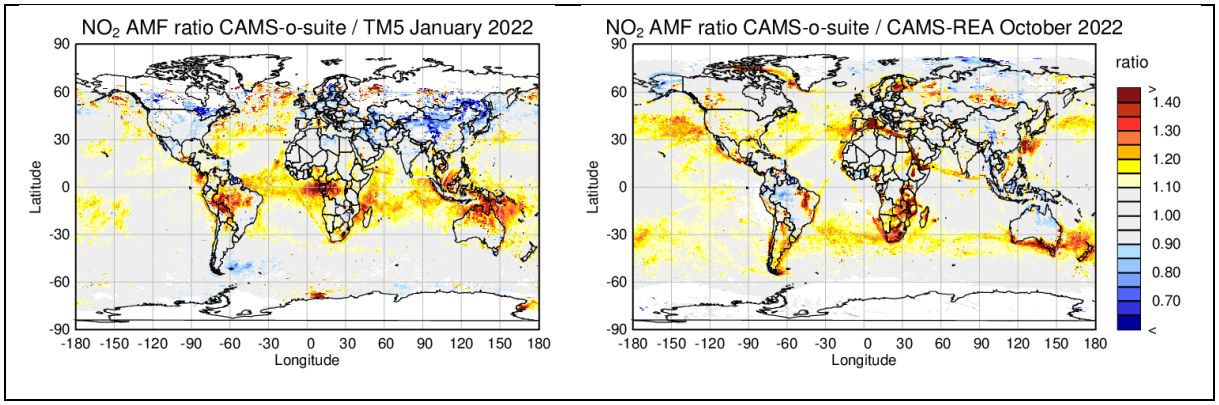


Figure 2.18: Ratio of tropospheric NO_2 AMFs when replacing the CAMS o-suite a priori by TM5-MP (left) and CAMS reanalysis (right) for Janaury 2022.

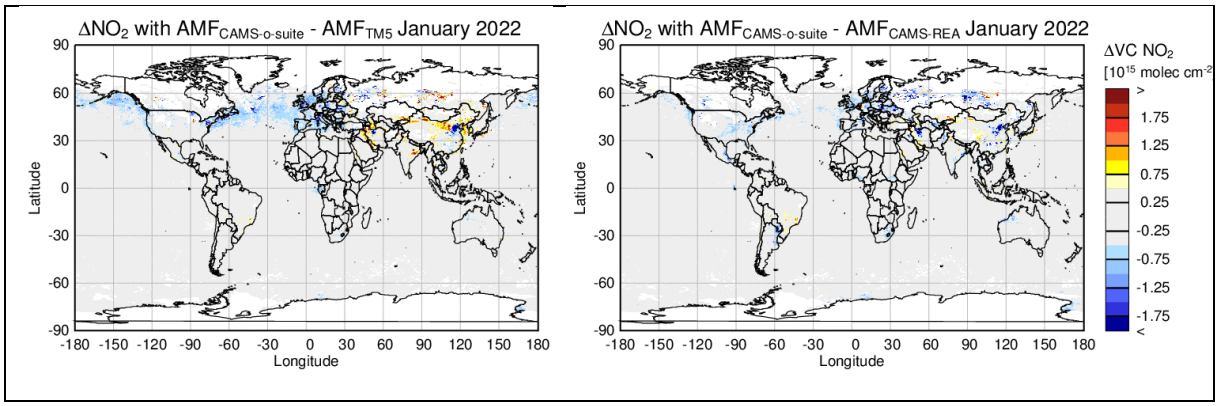


Figure 2.19: Difference of tropospheric NO_2 columns when replacing the CAMS o-suite a priori by TM5-MP (left) and CAMS reanalysis (right) for Janaury 2022.

As most of the large changes in air mass factors are located over regions having small NO_2 columns, the changes in tropospheric NO_2 column when switching NO_2 a priori is small as shown in Figure 2.19. For most of the globe, changes are below $0.5x10^{15}$ molec cm^{-2} in winter, and even smaller in summer. The conclusion of this sensitivity study on a priori profiles for NO_2 is, that all three state of the art atmospheric models used lead to comparable NO_2 retrievals. The larger relative differences over remote regions can be relevant for some studies focusing on free tropospheric NO_2, but are not the main point of interest for a climate data record.

2.4.3. **Clouds**

Two algorithms, OCRA and ROCINN (Loyola, 2007), (Lutz et al., 2016), are synergistically used for retrieving the cloud information from the GOME-2 sensor series. OCRA combines broadband spectral windows from UV (322–384 nm), vis (400–557 nm) and NIR (569–804 nm)

and retrieves the radiometric cloud fraction by examining separations between measured reflectances and their cloud-free composite values. ROCINN CRB (Cloud-as-Reflecting-Boundaries) is the follow-up algorithm which accepts as an a priori input the radiometric cloud fraction and retrieves within the O_2 A-band window two additional cloud properties; the cloud height and cloud albedo (Loyola, 2004). The CRB model assumes that the cloud behaves as a Lambertian reflector. Line-by-line transmittances are calculated using spectroscopic line information for the O_2 A-band from the HITRAN (high-resolution transmission molecular absorption) database. The information on the surface albedo quantity is taken from the MERIS black-sky albedo climatology (Popp et al., 2011). At first, the OCRA/ROCINN CRB cloud algorithm has been compared to the FRESCO+ v7 algorithm for the full time series (2007–2018). FRESCO+ is a cloud retrieval algorithm in the O_2 A-band. Three wavelength windows are used: (1) 758–759 nm (no absorption), (2) 760–761 nm (strong absorption), and 765–766 nm (moderate absorption). Each window comprises five GOME-2 wavelengths. FRESCO+ involves a forward RTM to build the Look-Up Tables (LUT) for the transmission database due to oxygen absorption, including the single Rayleigh scattering computations in a separate LUT (Wang et al., 2008). One remark is that OCRA produces a radiometric cloud fraction without making any assumptions on the cloud albedo, whereas FRESCO+ produces an effective cloud fraction assuming a fixed cloud albedo of 0.8.

Table 2-6: Regions used for the NO_2 cloud analysis

Region	**Region Type**	**Geophysical coordinates Latitude, Longitude**
Africa South of Equator	Biomass Burning	15°S – 0, 10 – 35°E
Europe	Anthropogenically polluted	35 – 50°N, 10°W – 20°E
Eastern China	Anthropogenically polluted	20 – 40°N, 105 – 125°E
Pacific	Clean Remote	20°S – 20°N, 110 – 150°W

For the OCRA/ROCINN algorithm, the operational GOME-2A product from EUMETSAT AC SAF was used (Valks, 2019). For FRESCO+ cloud properties, we used the dataset produced in the framework of the QA4ECV project. At a later step, FRESCO+ v8 became available, which is using the GOME-2 DLER surface albedo data. Therefore, the initial comparison has been done using ASCAF_OFF Vs FRESCO+ v7, which are based on the old L1 data, whereas the latest comparison was completed using ACSAF_RPRO Vs FRESCO+ v8, which are based on the new reprocessed GOME-2 L1 v6.3 data.

Monthly means have been calculated for the full time series (2007–2018) for AC SAF and QA4ECV (FRESCO+ V7) cloud parameters. During the last years (2017–2018), there is an artificial trend in the OCRA cloud fraction, resulting from the Level-1 data, because GOME-2 degradation correction has not been applied to the Level-1 product. To compensate for this instrument degradation, a soft-correction as a function of time and viewing geometry is computed based on the former years of the sensor's lifetime. A fourth-order polynomial is fitted to the PMDs mean reflectances. For the years after 2013, the degradation effect is partially removed from the PMD reflectances using extrapolated soft-correction factors. As

the global cloud-free composites for OCRA are based on GOME-2A data from April 2008 until June 2013, the data after 2013 do not have the optimal soft-correction factors. Therefore, the best period for drawing conclusions is between April 2008 and June 2013 due to the accurate degradation correction in OCRA colors and temporal match with the clear-sky composite maps (Argyrouli et al., 2023).

Differences between OCRA and FRESCO+ cloud fraction may originate from the different spectral fitting windows, since OCRA uses broadband wavelength bands covering the whole UV-VIS-NIR spectrum. The cloud height differences originate primarily from the different line-by-line models and forward models used in ROCINN and FRESCO+.

The comparison of cloud fraction and cloud height are performed over selected regions with diverse NO_2 conditions, summarized in Table 2-6. In general, the mean cloud heights and cloud fractions over Europe show the same seasonal variations for OCRA/ROCINN and FRESCO+. The cloud fractions for FRESCO+ v7 are systematically higher (by ±0.05) than the ones from ACSAF_OFF (see Figure 2.20). They also show smaller seasonal variations compared to OCRA/ROCINN. The use of FRESCO+ v8 brings better agreement with OCRA/ROCINN cloud fraction variabilities as the cloud fractions from FRESCO+ drop significantly for the summer months (see Figure 2.21) , and cloud fraction discrepancies are now mostly within 0.02. Similar patterns are identified for the cloud height time series with the difference that the FRESCO cloud heights are smaller than the OCRA/ROCINN ones. With the FRESCO version upgrade to v8, the differences with ACSAF_RPRO reduce for the cloud height in summer months but not in the winter months (see Figure 2.23). The cloud correction applying OCRA/ROCINN is expected to increase the tropospheric NO_2 columns more than when using FRESCO+ for the cloud correction.

For Eastern China, we systematically find the cloud fractions with FRESCO+v7 to be larger than ACSAF_OFF (see Figure 2.24), but we do not observe very strong seasonal variations in both datasets. For the cloud heights, the seasonal variations are strong over this region. Seasonal variations in FRESCO cloud heights are on average ~1km smaller than those in the OCRA/ROCINN cloud heights (see Figure 2.26). Overall, applying the cloud correction with FRESCO product will result in lower tropospheric NO_2 columns. As for the other regions, the discrepancies disappear when using the upgraded FRESCO+ v8 cloud heights and the only differences which remain are in the winter months December-January-February (see Figure 2.27).

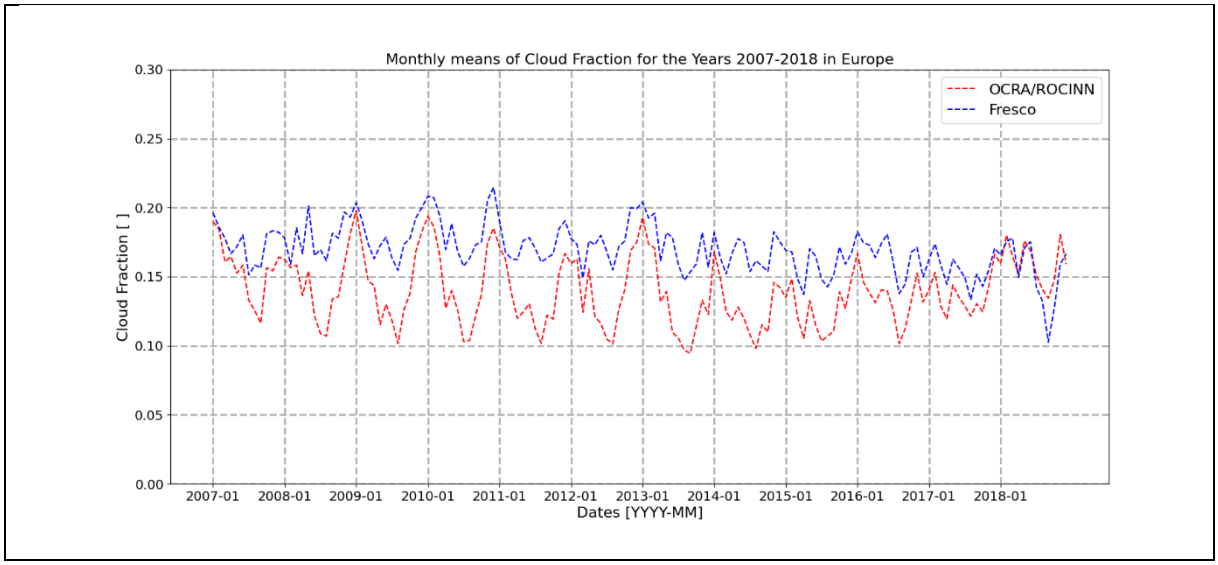


Figure 2.20: Time series of the monthly means of Cloud Fraction over Europe: ACSAF_OFF (red line) versus FRESCO+ v7 (blue line)

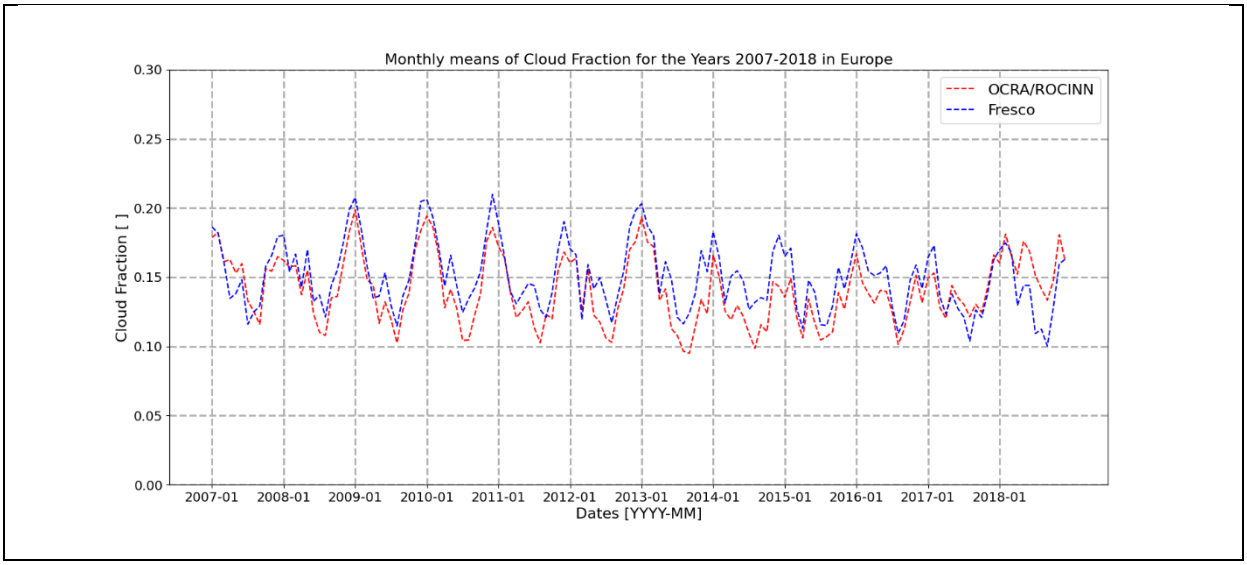


Figure 2.21: Time series of the monthly means of Cloud Fraction over Europe: ACSAF_RPRO (red line) versus FRESCO+ v8 (blue line)

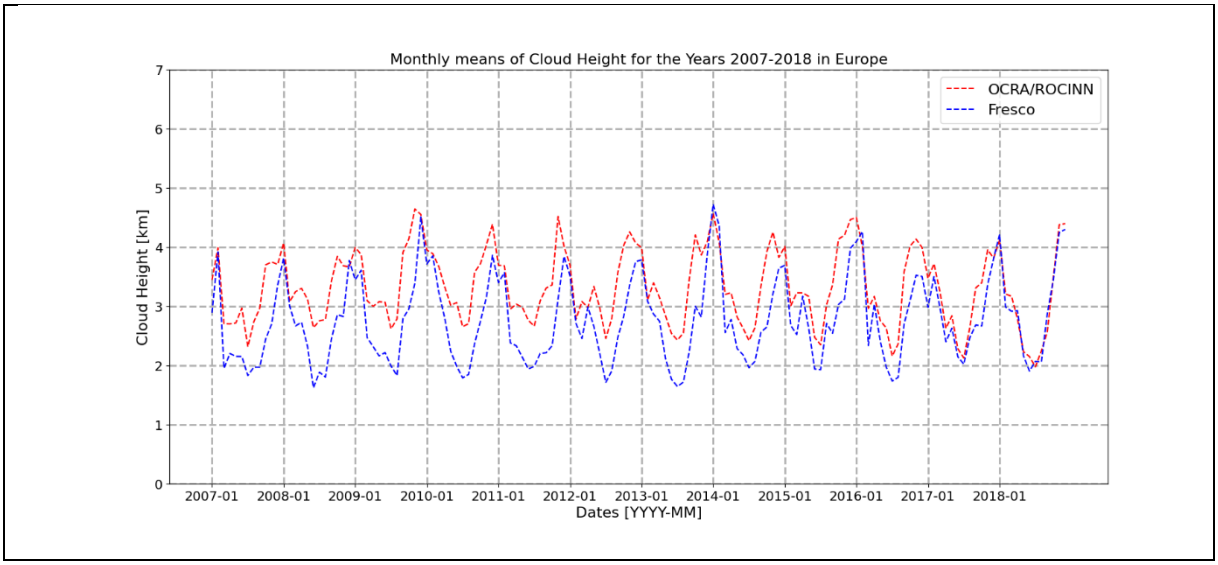


Figure 2.22: Time series of the monthly means of Cloud Height over Europe: ACSAF_OFF (red line) versus FRESCO+ v7 (blue line)

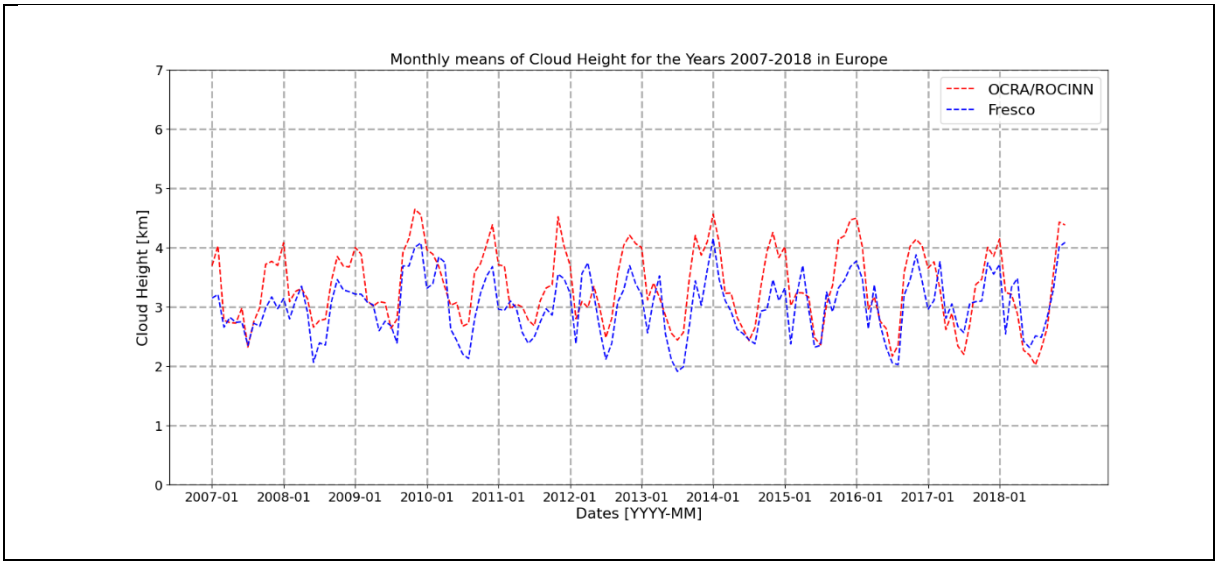


Figure 2.23: Time series of the monthly means of Cloud Height over Europe: ACSAF_RPRO (red line) versus FRESCO+ v8 (blue line)

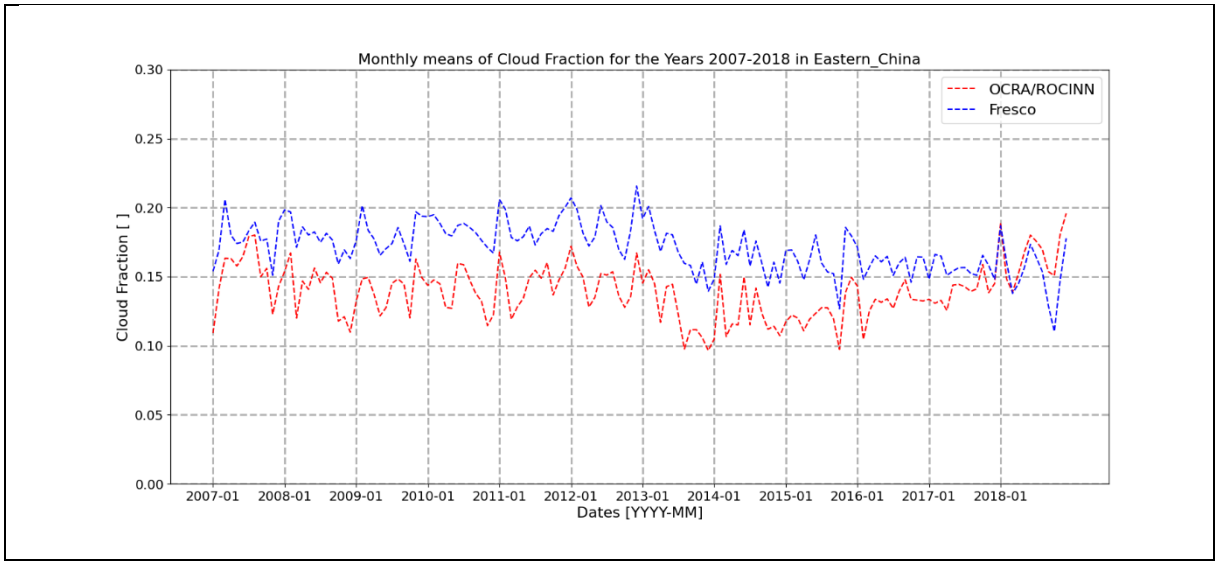


Figure 2.24: Time series of the monthly means of Cloud Fraction over Eastern China: ACSAF_OFF (red line) versus FRESCO+ v7 (blue line)

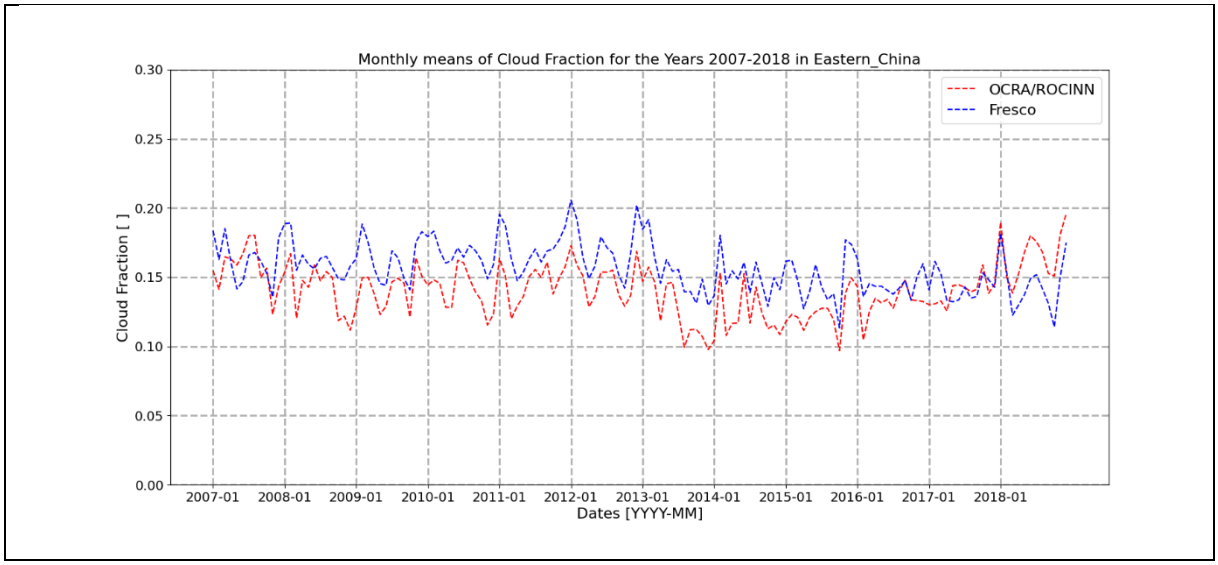


Figure 2.25: Time series of the monthly means of Cloud Fraction over Eastern China: ACSAF_RPRO (red line) versus FRESCO+ v8 (blue line)

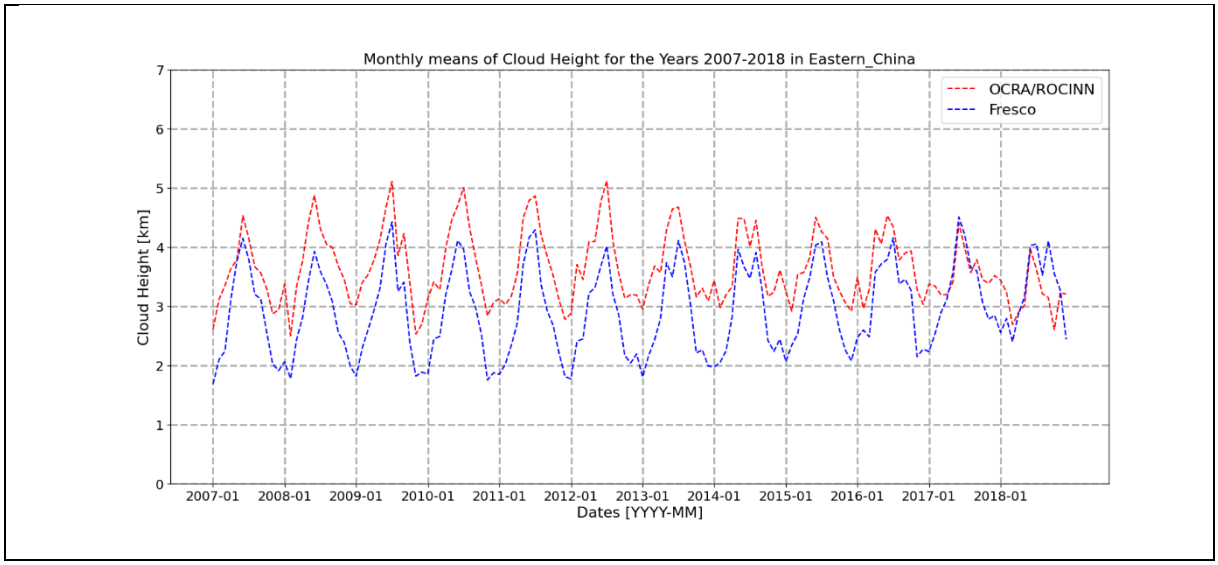


Figure 2.26: Time series of the monthly means of Cloud Height over Eastern China: ACSAF_OFF (red line) versus FRESCO+ v7 (blue line)

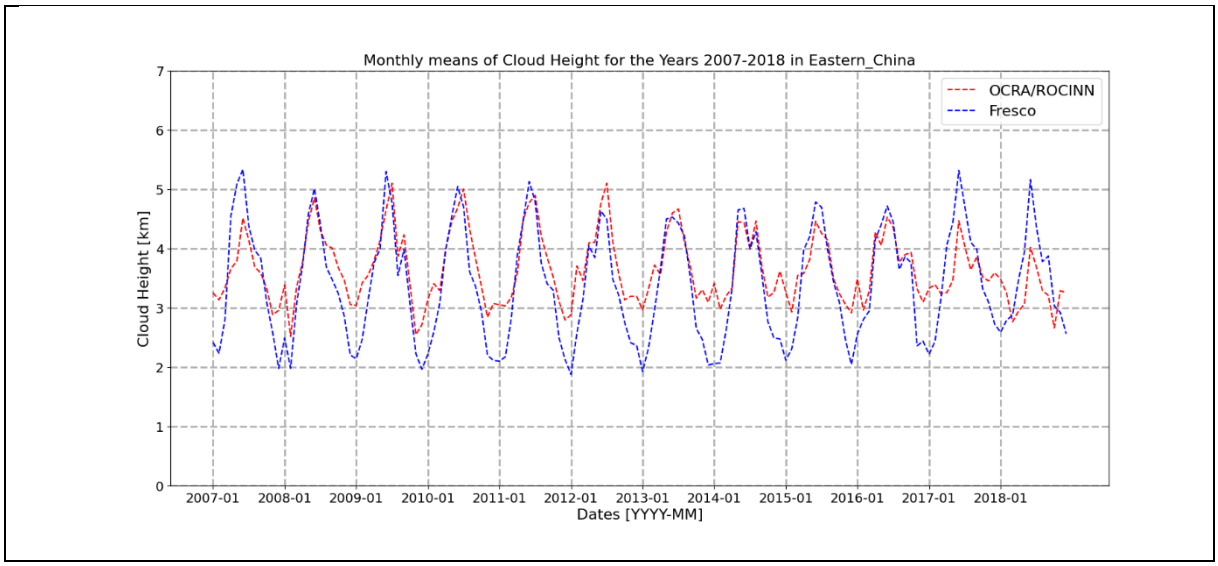


Figure 2.27: Time series of the monthly means of Cloud Height over Eastern China: ACSAF_RPRO (red line) versus FRESCO+ v8 (blue line)

The impacts of two different cloud products (OCRA/ROCINN used in the AC-SAF product and FRESCO+ v7 used in the QA4ECV product) on the AMF calculation are analysed. To assess the influence of cloud corrections on tropospheric NO_2 column retrievals, we compute the ratio of total tropospheric NO_2 vertical columns to clear-sky tropospheric NO_2 vertical columns. As shown in Figure 2.28 and Figure 2.29, the temporal variabilities of the cloud corrections from OCRA/ROCINN and FRESCO+ v7 are generally similar. In the clean Pacific region, the influence of cloud correction on the GOME-2 NO_2 retrieval is small compared to other source regions.

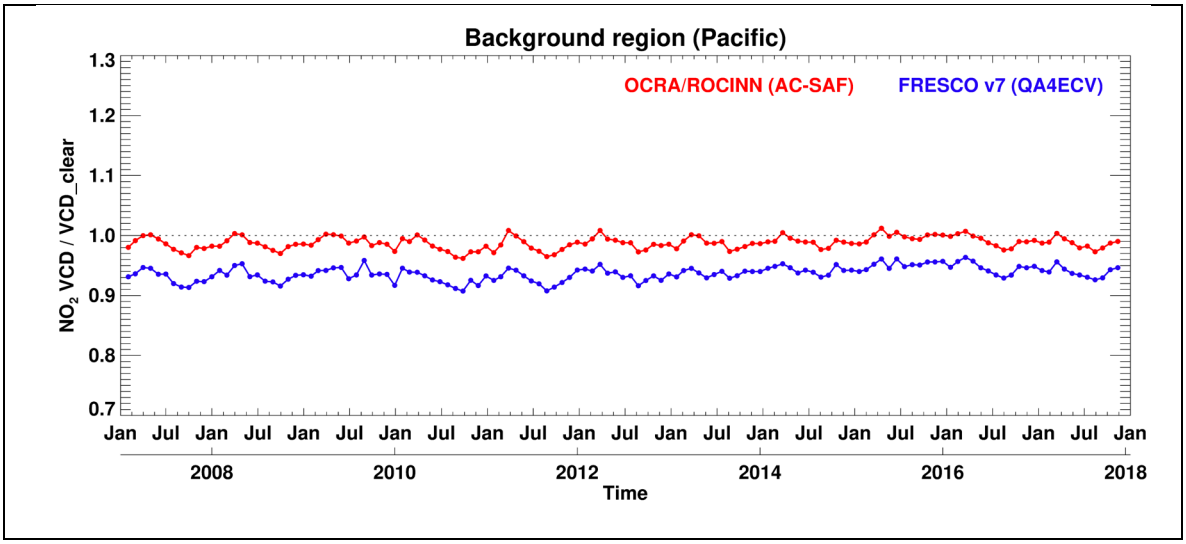


Figure 2.28: Time-series of the ratio of total tropospheric NO_2 vertical columns to clear-sky tropospheric NO_2 vertical columns applying the OCRA/ROCINN (used in AC-SAF OFFL) and the FRESCO+ v7 (used in QA4ECV) cloud products for clean Pacific sector.

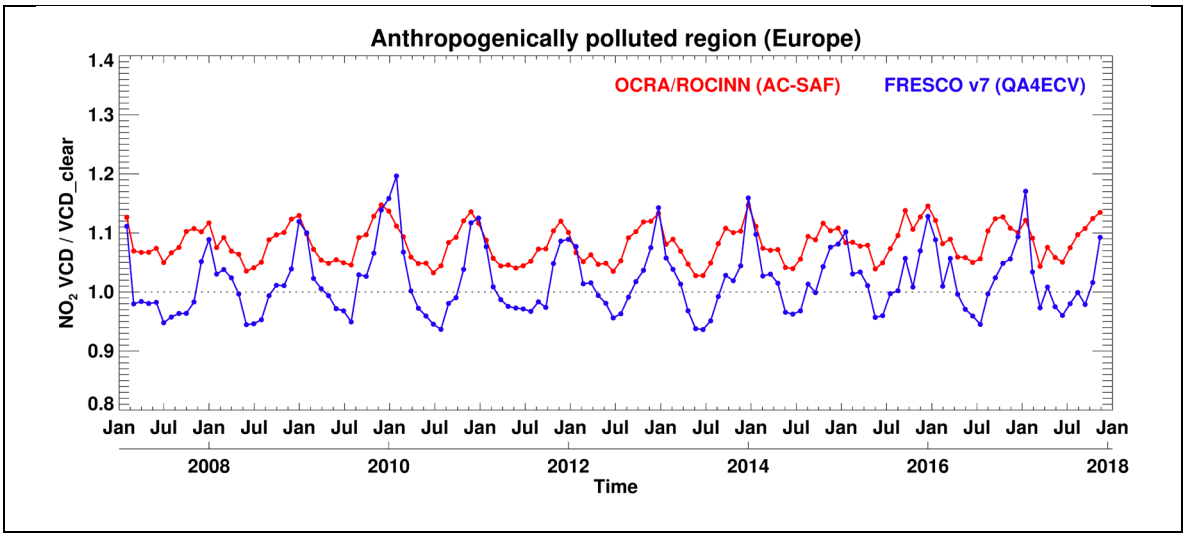


Figure 2.29: Same as Figure 2.28, but for Europe.

For anthropogenically polluted regions, cloud corrections using both OCRA/ROCINN and FRESCO+ v7 lead to changes in tropospheric NO_2 columns by 5-15 % and -5 – 15 %, respectively, showing seasonal variations. Cloud corrections using OCRA/ROCINN result in up to ~ 20 % higher tropospheric NO_2 columns over Europe (Figure 2.29) compared to those from FRESCO+ v7, reflecting higher OCRA/ROCINN cloud heights (Figure 2.23) leading to a stronger 'screening effect' in the AMF calculations.

The Round Robin activities reported here demonstrate that both cloud algorithms are fit-for-purpose to generate level 2 NO_2 data for the GOME-2 sensors. Additionally, the improved consistency between the reprocessed OCRA/ROCINN (AC-SAF RRPO) and updated FRESCO+ v8 is expected to yield better agreement between cloud corrections in morning sensor NO_2 retrievals.

2.5. Validation

To assess the accuracy of the different NO_2 retrieval approaches and settings, a ground-based validation of the stratospheric NO_2 vertical column densities was performed by comparison of various satellite NO_2 data sets to NDACC zenith-sky scattered light DOAS (ZSL-DOAS) measurements (see Solomon et al., 1987 for the measurement principles). These comparisons test both the total column retrievals in very clean conditions, i.e., those with a negligible tropospheric component, and the stratosphere-troposphere separation (cf. Section 2.3) in moderately polluted regimes. Results in highly polluted conditions are not representative as the ZSL-DOAS measurements may have some sensitivity to very elevated tropospheric columns. For various reasons, including network homogeneity and computational effort, the analysis presented here relies solely on the LATMOSv3-processed SAOZ network data (Pommereau & Goutail, 1988). More extended comparisons covering all NDACC ZSL-DOAS instruments will be presented in the 2nd iteration of the Product Validation and Intercomparison Report (PVIR). Details of the validation methods used are summarised in Appendix A1.

More specifically, the analysis was designed to detect differences between three stratospheric NO_2 products obtained from GOME-2 on MetOp-A observations:

- QA4ECV data assimilation (QA4ECV DA hereafter)
- QA4ECV STREAM
- AC-SAF STREAM

To ensure minimal impact of differences due to mismatch in space and time between satellite and ground-based measurements, only sunrise ZSL-DOAS measurements were used, and these were adjusted to the GOME-2 overpass time using a look-up table based on a photochemical box model. Satellite pixels were averaged over the area of sensitivity (towards the rising sun) of the ZSL-DOAS measurements.

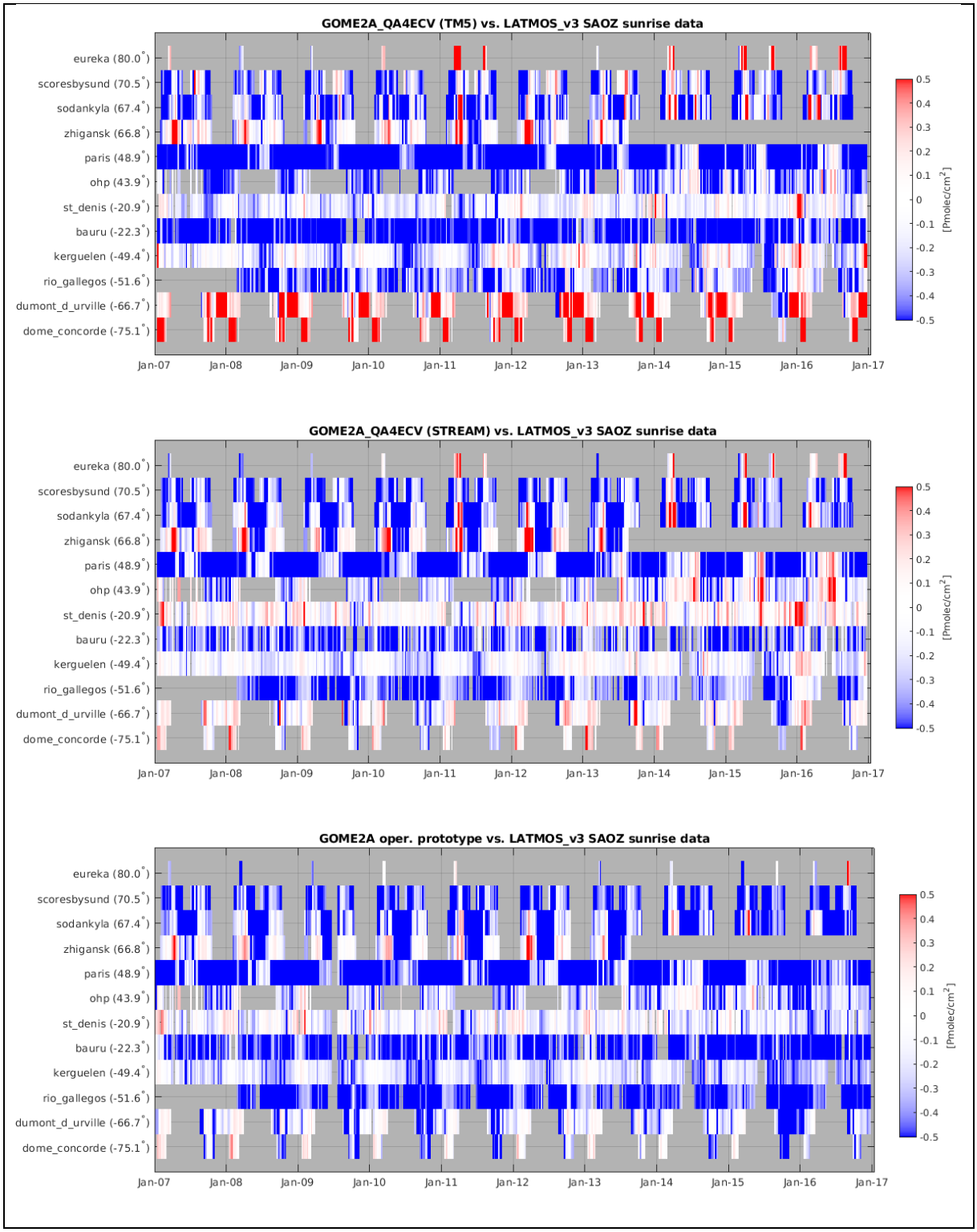


Figure 2.30: Time-series of the absolute differences between the satellite-based stratospheric NO_2 VCDs and those measured with the ground-based SAOZ ZSL-DOAS instruments (LATMOSv3 processing). Top to bottom: QA4ECV DA, QA4ECV STREAM, AC SAF prototype.

Figure 2.30 contains time series of differences between satellite and ground-based measurements for the 3 GOME-2A products. The derived per-station statistics are presented in Figure 2.31 as box-whisker graphs. Several observations can be made:

- Overall, satellite stratospheric columns underestimate the ground-based measurements by about 0.1 – 0.3x10^{15}molec/cm^2. The smallest network-wide median difference is observed for QA4ECV STREAM. The largest underestimation is seen for the AC-SAF prototype product, but the differences between the products are hardly significant w.r.t. the total uncertainties in the validation exercise.
- Comparisons at Paris and Bauru show persistent negative biases for all three data sets, but these are related to tropospheric contamination in the ZSL-DOAS data and therefore not a reason for concern.Data assimilation leads to positive biases at the highest latitudes not seen in the STREAM data.
- The dispersion of the differences, as a proxy for more random errors, ranges from about 0.1 to 0.3x10^{15} molec/cm^2, again within the total uncertainty of the validation exercise. However, data assimilation leads to larger dispersion at high latitudes (see further analysis below).
  Correlation coefficients (not shown) are typically of the order of 0.8-0.9, indicating that the evolution of stratospheric NO_2 is well captured.

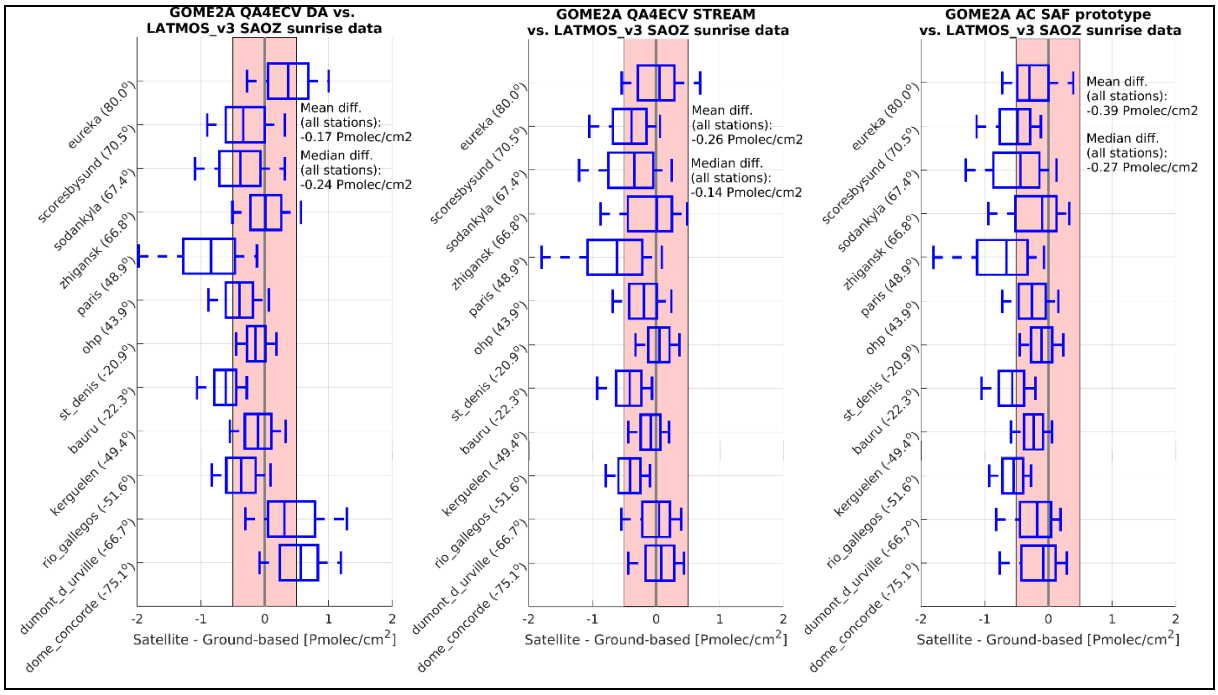


Figure 2.31: Box-whisker plots corresponding to the comparisons shown in Figure 2.5-1. Comparisons at Paris and Bauru are affected by tropospheric pollution in the ZSL-DOAS signal and the large negative deviations are therefore not a reason for concern. Most noteworthy are the different signals at the highest latitudes: data assimilation (left-most panel) leads to slight positive biases not seen in the STREAM data.

As illustrated above, the main difference between the stratospheric products is found at high latitudes. Further analysis reveals that this occurs for GOME-2A and SCIAMACHY, but not for GOME and TROPOMI. The most interesting station in this regard is Dumont d'Urville, on the East coast of Antarctica. For QA4ECV DA, the histogram of the differences shows a bimodal distribution, which is not seen for QA4ECV STREAM (see Figure 2.32).

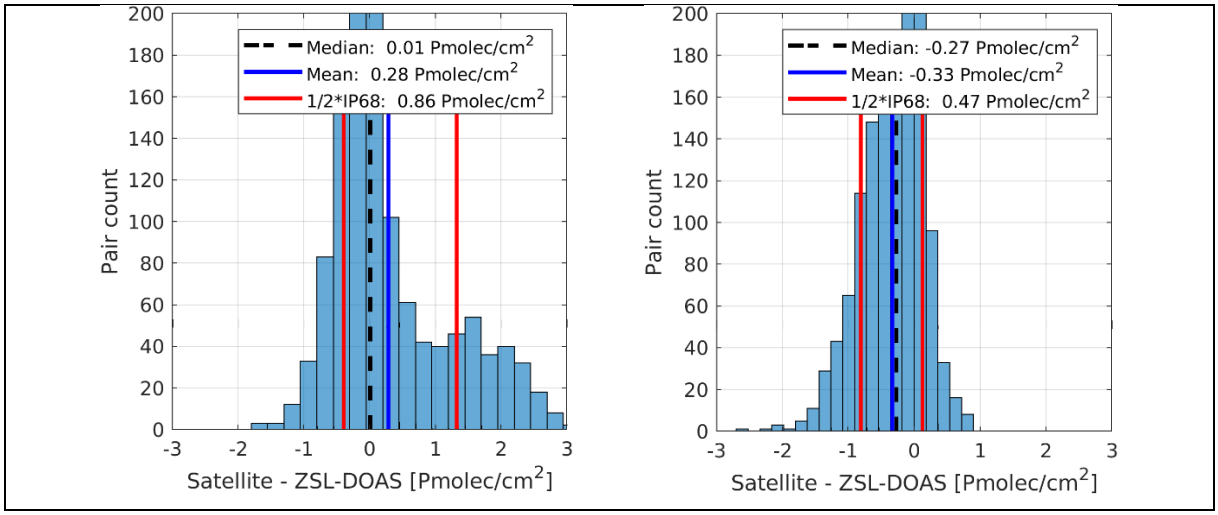


Figure 2.32: Histograms of the differences between satellite and ground-based stratospheric NO_2 column measurements at Dumont d'Urville (East coast of Antarctica), in this case for SCIAMACHY with either QA4ECV DA (left-hand panel) or QA4ECV STREAM (right-hand panel) processing. Note the bimodal distribution for QA4ECV DA.

The additional mode at significant positive differences causes the observed positive bias and the large dispersion. It is also clearly distinguishable in the scatter plot in Figure 2.33. This figure also demonstrates the large negative tropospheric columns that result from the overestimated stratospheric column. In other words, the positive bias in the stratospheric columns is not due to an overestimated total column, but due to an erroneous estimate of the stratospheric column.

A stricter filtering, using the snow-ice flag to remove snow and ice scenes as recommended for the tropospheric column product, removes these erroneous data, but at the expense of significant data loss. Applying this filter to the southernmost station, Dome Concordia, removes all observations there. A potential cause of this issue is the assimilation of data obtained in de descending part of the orbit at early solar times, which has been found to result in substantial deviations in the TROPOMI processor. This is under investigation and will likely be revised for future processing.

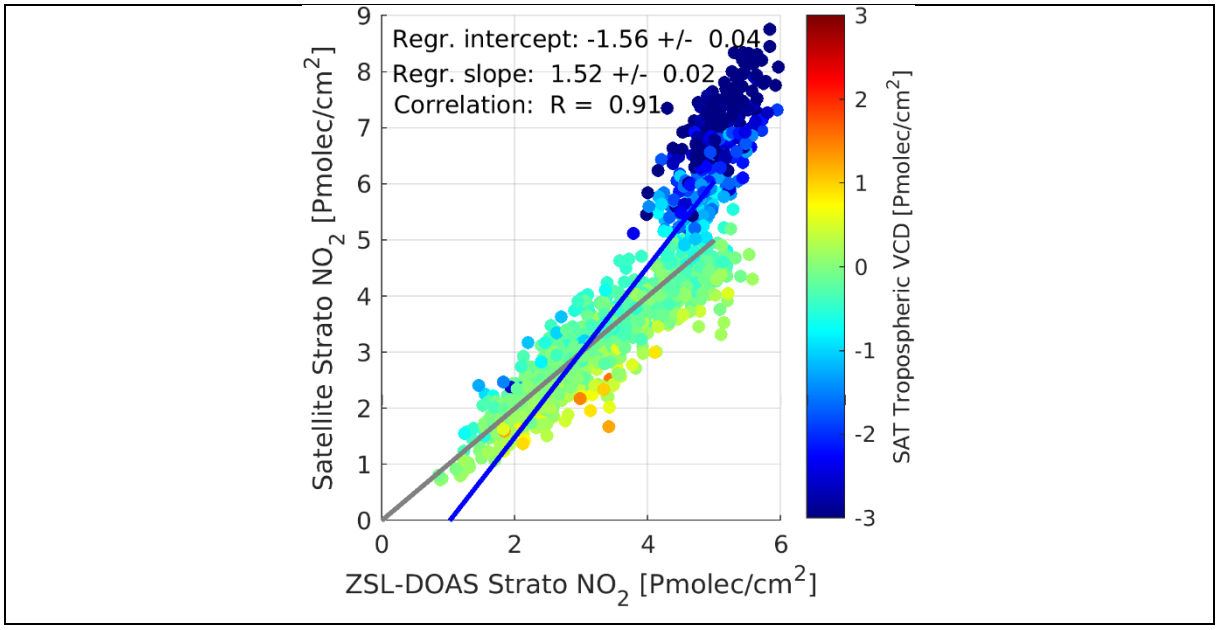


Figure 2.33: Scatter plot for the SCIAMACHY QA4ECV DA comparisons at Dumont d'Urville (cf. the left-hand panel of Figure 2.32). Markers are coloured by tropospheric VCD.

In conclusion, the ground-based validation does not reveal one product/approach to be significantly better than another at low and mid-latitudes. At high latitudes, the QA4ECV DA product has issues potentially related to an assimilation of the descending part of the orbit for GOME-2(A) and SCIAMACHY which is investigated further and may be revised for future processing.

2.6. Algorithm Selection

Based on the comparisons between various algorithm settings in the NO_2 Round Robin, and subsequent evaluation and validation efforts, we select the algorithms for the NO_2 climate data record outlined in Table 2-7 below. The algorithms for GOME, SCIAMACHY, and OMI (and in practice also TROPOMI) have been developed and applied in the QA4ECV-project, and the algorithms for the GOME-2 sensors have been developed in the AC-SAF project. For TROPOMI, cloud characteristics and surface pressure are obtained with the approach developed for the operational S5P product, i.e. cloud fraction is obtained from reflectances in the NO_2 fitting window, whereas cloud pressure is obtained with the FRESCO+ approach. These settings have been selected for their consistency with the QA4ECV-approach. More details about specific settings can be found in the ATBD for the CCI+ ECV Precursors project.

Table 2-7: Algorithm choices for the tropospheric NO_2 CDR

	GOME	SCIAMACHY	GOME-2A	GOME-2B	GOME-2C	OMI	TROPOMI
Spectral fitting							
Level-1 input	V5.1	V8	V6.3	V6.3	V6.3	Coll3	Coll2
Fitting window	425-465 nm	425-465 nm	425-497 nm	425-497 nm	430-465 nm	405-465 nm	405-465 nm
Stratospheric correction							
Method	Data assimilation	Data assimilation	STREAM	STREAM	STREAM	Data assimilation	Data assimilation
Air Mass Factor calculation							
Surface albedo	GOME-2 LER v2 MIN	GOME-2 LER v2 MIN	GOME-2 DLER v4 MIN	GOME-2 DLER v4 MIN	GOME-2 DLER v4 MIN	OMI LER	OMI LER MODE
Surface pressure	ECMWF/DEM	ECMWF/DEM	GTOPO30	GTOPO30	GTOPO30	ECMWF/ DEM	GMTED2010
Cloud fraction	FRESCO v7	FRESCO v7	OCRA	OCRA	OCRA	O2-O2	NO_2 window
Cloud pressure	FRESCO v7	FRESCO v7	ROCINN	ROCINN	ROCINN	O2-O2	FRESCO+ wide
Cloud albedo	0.8	0.8	ROCINN	ROCINN	ROCINN	0.8	0.8
A priori NO_2 profile	TM5-MP	TM5-MP	TM5-MP	TM5-MP	TM5-MP	TM5-MP	TM5-MP

The main lessons learned from the NO_2 Round Robin have driven this selection of instrument-specific algorithms:

- While the spectral fitting procedures for the GOME-2 AC-SAF retrieval are as much as possible aligned with the choices for the QA4ECV algorithms for GOME, SCIAMACHY, and OMI, it was found that using the 425-497 nm window for GOME-2A and GOME-2B, and the 430-465 nm window for GOME-2C gives the best results when applied on the recently reprocessed EUMETSAT (v6.3) level-1 data.
- Because data assimilation and STREAM agree to within 0.2 10^{15} molec. cm^{-2}, and both are validated and found to be fit-for-purpose, the STREAM algorithm is selected for the GOME-2 algorithms, and data assimilation for GOME, SCIAMACHY, OMI and TROPOMI. For the QA4ECV algorithms, STREAM estimates of the stratospheric correction are included as alternatives to the data assimilation estimate.

- The Round Robin pointed out that agreement between the AC-SAF and QA4ECV cloud parameters has improved considerably, so that selection of OCRA/ROCINN versus FRESCO-type cloud retrievals is not introducing large uncertainties in the air mass factor calculations. With the implementation of the FRESCO v8 cloud product, we anticipate somewhat smaller cloud fractions, which may lead to lower tropospheric NO_2 columns (less screening) for the GOME and SCIAMACHY retrievals.

One remaining discrepancy in the selection of the algorithms is the use of a directional surface reflectivity database (DLER) for the AC-SAF algorithms, whereas the QA4ECV algorithms use a non-directional database (LER). It is foreseen that during the extension and upon granting of the requested optional work packages in the CCI+ Precursors project, QA4ECV algorithms improving to DLER instead of LER become available. These are anticipated to further improve the consistency between the AC-SAF and QA4ECV algorithms selected.
From the sensitivity tests and comparisons performed, uncertainties can be estimated, as well as their random or systematic behaviour in time. This is summarized in Table 2-8. We consider measurement noise to be the dominant contribution to slant column uncertainty (although there are systematic contributions from inaccurate cross section spectra, modelling errors, etc.). The error associated with selecting one stratospheric correction over the other is systematic, as it affects all tropospheric retrievals, but its impact is most relevant when tropospheric NO_2 columns are low. In regions with high tropospheric NO_2 columns, the uncertainty will not be dominated by the selection of a particular method, but rather by AMF errors. Differences in the choice of surface albedo ((D)LER) dataset have a local impact on the AMF, but this impact is complicated because of algorithm interactions with the cloud retrieval. Moreover, one should anticipate differences with considerable spatial correlation - on the regional scale because of the relatively low resolution of (D)LER databases and because albedo modifications (rainfall, droughts, …) are often spatially correlated. Similar arguments hold for the selection of a particular set of a priori NO_2 profiles. For clouds, differences in cloud fraction and cloud height will have a local, systematic impact on AMF calculations. Because there are differences in cloud correction schemes, and (residual) clouds are highly variable in time, we anticipate that at least part of the cloud-related errors will be random.

Table 2-8: Summary of NO_2 sensitivity analysis for each of the retrieval tests.

Test	Sensitivity	Behaviour in time
ΔSCD	±0.6 10^{15} molec./cm^2	Mostly random
Stratospheric correction	±0.2 10^{15} molec./cm^2	Mostly systematic
(D)LER	±20%	Mostly systematic
Profiles	±10%	Mostly systematic
Clouds	±20%	Random and systematic

3. HCHO

3.1. Data and Algorithms used

The HCHO algorithmic developments build on a solid heritage from the QA4ECV project, the TROPOMI Level-2 processor development and the AC SAF operational GOME-2 products. The "afternoon" OMI and TROPOMI algorithms have been developed in parallel. The "morning" QA4ECV datasets share mostly the same settings as OMI. The GOME-2 AC SAF HCHO products differ from the QA4ECV and TROPOMI column retrievals in several aspects that have been revisited for this Round Robin (sections 3.2 and 3.3). The different datasets and versions are listed below.

- AC SAF GOME-2A, GDP-4.8: https://acsaf.org/products/oto_hcho.php
- QA4ECV GOME-2A and OMI, v1.2: https://doi.org/10.18758/71021031, https://www.temis.nl/qa4ecv/hcho.html
- TROPOMI Reprocessed (RPRO) dataset, v02.03: https://doi.org/10.5270/S5P-vg1i7t0

The aim of the first part of the Round Robin activities for HCHO was the preparation of an improved prototype AC SAF algorithm, consistent with the other HCHO retrievals, and including findings from the Round Robin.

The second part of the Round Robin activities mainly focused on the air mass factor auxiliary data selection for all sensors: the surface albedo (section 3.4.1), the model a priori profiles (3.4.2) and the cloud products (3.4.3).

3.2. Slant Column Intercomparison

The QA4ECV and operational AC SAF GOME-2A slant column retrievals have been intercompared on 12 selected test days at the beginning, middle and end of the METOP-A lifetime and covering the four seasons.

Figure 3.1 and Figure 3.2 present comparisons of respectively HCHO background-corrected SCDs and fit residuals from the AC SAF GDP-4.8 and the QA4ECV algorithms. Results are presented over the complete time series over the reference sector (the Equatorial Pacific: [15°S to 15°N], [180°W to 120°W]), and globally for one test day at the beginning of GOME-2A measurements. The East/West bias visible in the GDP-4.8 HCHO ΔSCD are reduced in the QA4ECV HCHO retrievals. Residuals of the fits are globally lower in the QA4ECV retrievals, especially on the edges of the orbits, and over the Southern Atlantic Anomaly (SAA) region. For these reasons, the QA4ECV fit settings have been adopted in the AC SAF prototype algorithm.

Besides the operational (GDP-4.8, AC-SAF) and the QA4ECV products, we included a prototype algorithm that is currently being developed and is planned to be implemented for AC-SAF. The algorithm is similar to the QA4ECV in the SCD calculation, the background correction, and the AMF calculation. GDP-4.8 uses two retrieval windows and a solar reference, while the other two make use of an earth shine reference over the Pacific Ocean and use one fitting window to fit both BrO and HCHO simultaneously. Because of this systematic difference, a respective bias between the SCDs is to be expected, we hence compare the background corrected SCD (or ΔSCD, see section 3.3). Other differences between the fit settings include an iterative spike

removal approach in the QA4ECV and AC SAF protype algorithms, that improves the fits over the South Atlantic Anomaly region. Pseudo-cross sections are also added in the fits, in order to reduce East-West orbit bias. The details of the retrieval settings can be found in the ATBD of the CCI + precursors [Precursors_cci+_D2.1_ATBD_01_00].

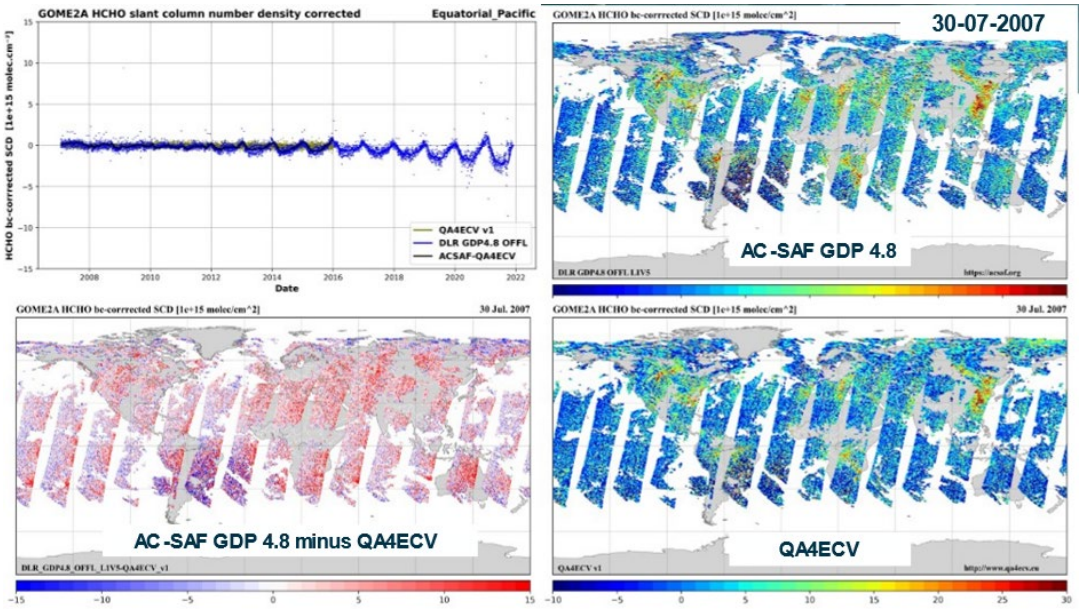


Figure 3.1: GOME-2A HCHO background-corrected slant columns retrieved with the QA4ECV and the GDP-4.8 algorithms. The first panel shows time series of the HCHO bc-SCDs averaged over the reference sector. The upper and lower right panels show global daily maps on 2007-07-30, while the lower left panel shows the corresponding SCD differences.

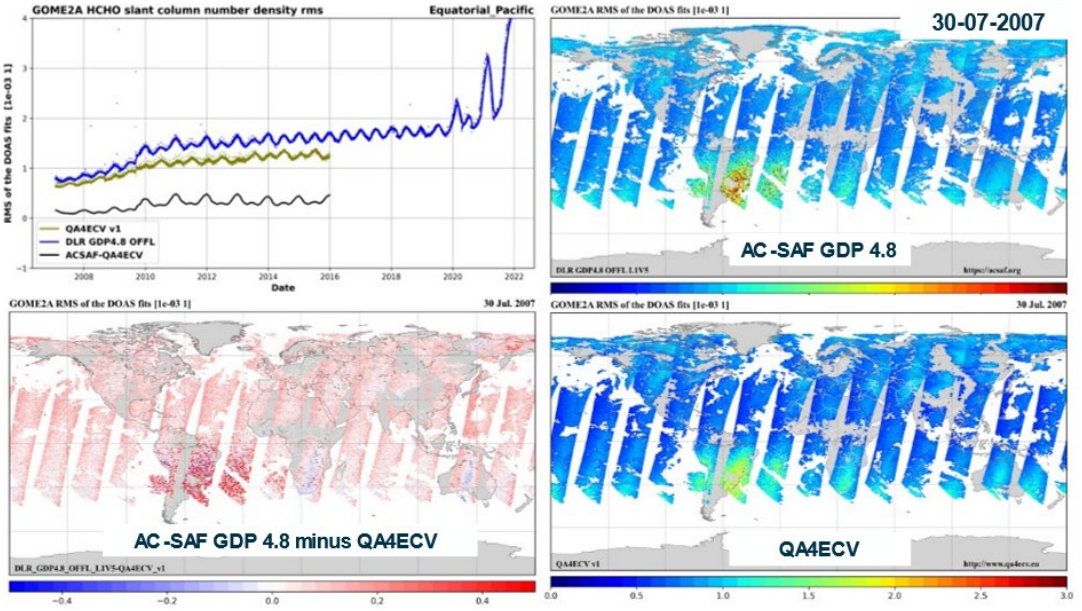


Figure 3.2: Same as Figure 3.1, but showing the residuals of the fits (RMS).

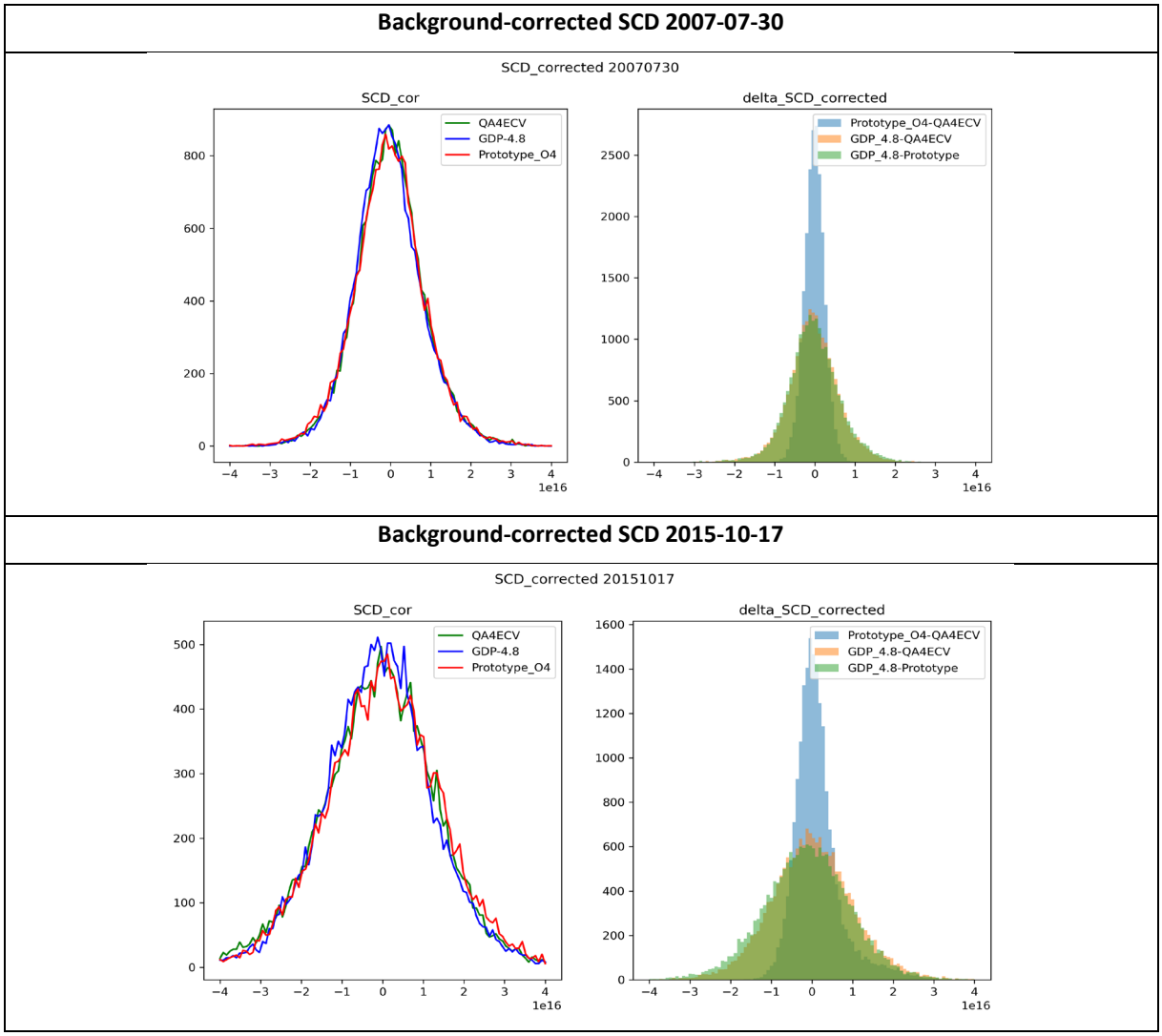


Figure 3.3: Histograms of the background-corrected HCHO SCDs over the Pacific Ocean for two selected days, 2007-07-30 and 2015-10-17. The right plots show the histograms of the differences between the three retrievals (QA4ECV, AC SAF GDP-4.8 and AC SAF Prototype).

Formaldehyde optical densities are more than 10 times smaller than those of NO_2, and the retrieved HCHO SCDs therefore suffer from larger noise. To compare the noise of the retrieved SCDs, we looked at two orbits over the Pacific Ocean. The standard deviation in this region is dominated by the retrieval noise and not affected by real atmospheric signals. The histograms in Figure 3.3 show hardly any difference in their width between the 3 considered algorithms, indicating that the noise in the retrievals is very similar. Comparing the top figures (2007) with the bottom (2015), the instrument degradation becomes obvious as the width of the distribution increases. On average, the difference between all background-corrected SCDs is close to zero, suggesting that the background correction yields similar SCDs over the Pacific for all three algorithms. The width of the peak reflects the level of agreement between the two algorithms compared. Notably, the histogram of the difference for Prototype-QA4ECV (blue) is both narrower and higher than that of GDP-4.8-QA4ECV (orange), indicating better agreement.

There is a small deviation in the differences between the prototype and QA4ECV towards higher values, which might also be caused by negative corrected SCD in the QA4ECV data in the Polar regions (Figure 3.4). We therefore attribute it to an albedo effect in the background correction of the QA4ECV data.

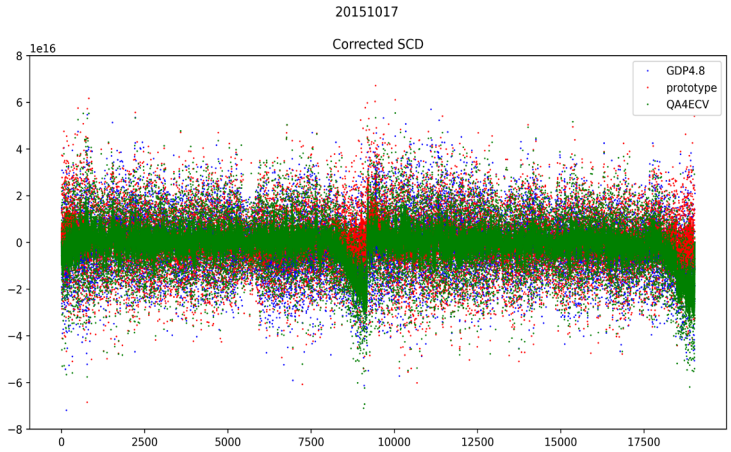


Figure 3.4: Time series of the corrected HCHO SCD of the two orbits over the Pacific for 2015-10-17. The line shows the running mean over 11 Pixels. Towards the end of the orbit, the green line (QA4ECV) deviates towards lower values.

3.3. Background correction

As previously mentioned, the low HCHO optical absorption result in larger noise in the retrieved HCHO slant columns, but also in systematic offsets. These offsets show dependencies mainly on viewing zenith angle (across-track), on latitude (along track), and on time (instrumental degradation).

To account for this, a background correction is generally performed in a two-step approach (see Table 3-1). First the data are corrected for any global offset and viewing angle (VZA) dependency, based on the retrieved slant columns selected in the central Equatorial Pacific. In a second step, a latitudinal correction is applied over the entire Pacific Ocean. The background VCDs are then approximated by modelled values.

In the GDP-4.8 product, no correction is applied for the VZA dependency. Another difference lies in the calculation of the background VCDs added to the final VCDs. In GDP-4.8, this correction is applied to the VCD. However, the offsets are probably present in the SCD instead of the VCD. Therefore, the QA4ECV algorithm and the current prototype correct the SCD by applying the air mass factors on the VCD offset.

Table 3-1: Summary of the 2 step HCHO background correction

GDP-4.8	QA4ECV and AC SAF prototype
ΔSCD: Global offset in the Equatorial Pacific + latitudinal correction in the full Pacific.	ΔSCD: Global offset + VZA correction in the Equatorial Pacific + latitudinal correction in the full Pacific.
$VCD_{cor} = \frac{\Delta SCD}{M} + VCD_{mod}$	$VCD_{cor} = \frac{\Delta SCD + VCD_{mod} \cdot M_{ref}}{M}$

The final background-corrected VCD (VCD_{cor}) are derived from the background-corrected SCDs (ΔSCD) divided by the air mass factors (M). Modelled HCHO columns (VCD_{mod}) in the reference sector are finally added. In the QA4ECV and AC SAF prototype algorithms, this modelled column is multiplied by the corresponding air mass factor in the reference sector (M_{ref}). In the previous section, only ΔSCD have been compared. In Figure 3.1, the histograms of the differences between the retrievals are all centred around zero, showing a good correction of the global offsets. However, a much thinner spread is achieved for prototype-QA4ECV than if GDP-4.8 is included. This shows that the East/West bias and the latitudinal corrections are now better aligned. This can also be seen in Figure 3.1, as well as the improved stability over time.

We have not looked at the final VCD_{cor} comparisons. The background correction ultimately depends on the model data used for the a priori profiles, through VCD_{mod}. Furthermore, in the QA4ECV and prototype approach, the background correction depends on the air mass factor ratio between the Pacific region and the other locations (M_{ref}/M) (and therefore on the selected albedo, profiles and cloud correction approach). We did not investigate in detail which of the mentioned parameters has the largest influence on the results. This will be revisited when the final AMF settings are implemented (see section 3.4).

3.4. AMF Impacts

3.4.1. LER

The impact of the surface albedo (LER) has been tested using 4 full months of TROPOMI data (Jan., Apr., Jul., and Oct. 2020), at the global scale. Three albedo climatologies have been compared:

- GOME-2 surface LER & DLER database v4.0 (Tilstra et al., 2017)
- OMI surface LER database (Kleipool et al., 2008): currently used in the QA4ECV HCHO products (GOME, SCIAMACHY, GOME-2, OMI) and in the operational TROPOMI HCHO product.
- TROPOMI surface LER & DLER database v2.0 (Tilstra et al., 2023).

As can been seen in Figure 3.5, the agreement between the different LER database (min LER) is significantly better in the VIS (440nm) than in the UV (340nm). In the UV, outside of the snow/ice regions, the TROPOMI min LER values are globally lower than the OMI min LER values by about 0.01 to 0.02. Differences are more important for the GOME-2 min LER values, which are significantly larger than the two other databases (by 0.02 up to 0.06).

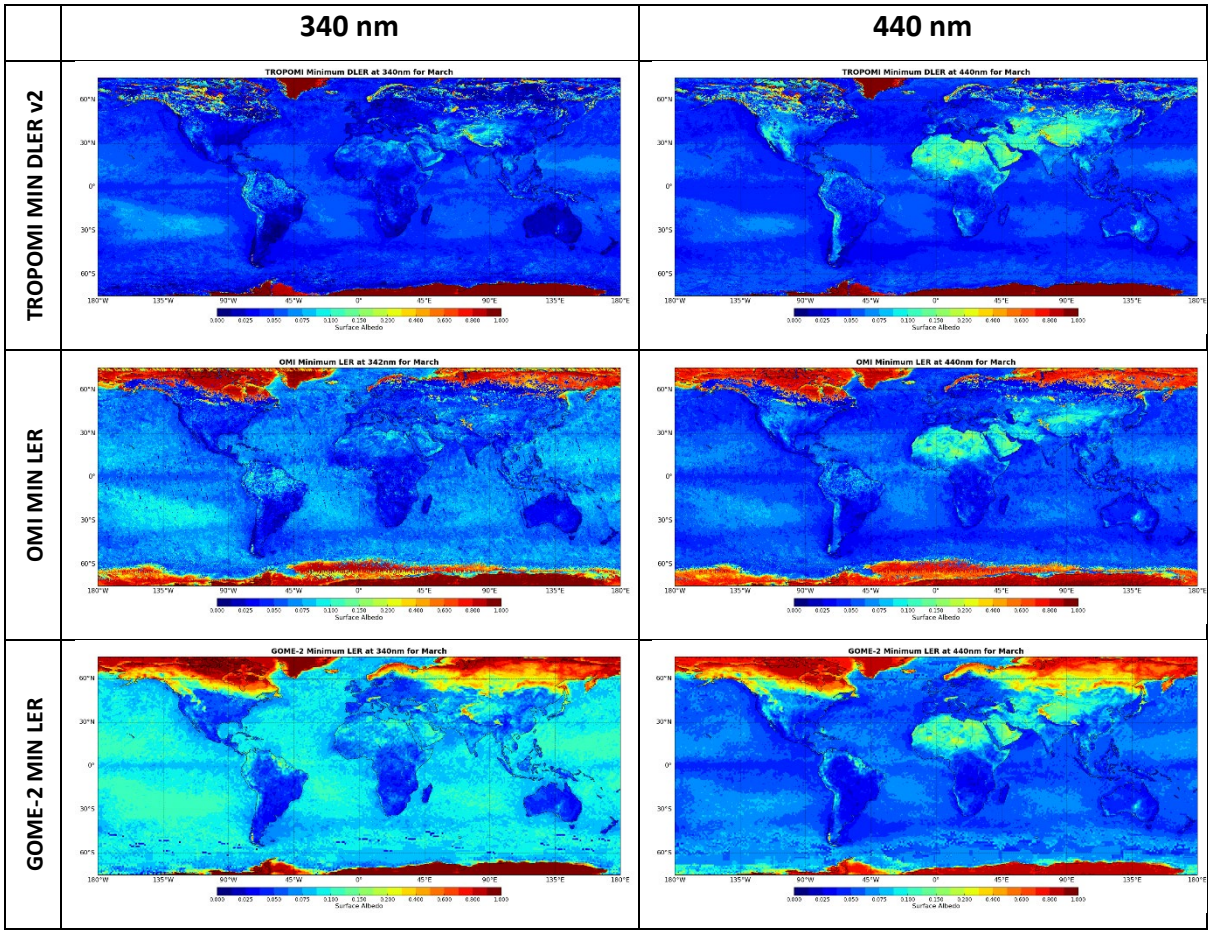


Figure 3.5: Comparison of March min LER values at 340 and 440 nm in the TROPOMI, OMI and GOME-2 databases.

Figure 3.6 presents the TROPOMI HCHO VCDs (no cloud correction) for July 2020 obtained using the OMI or TROPOMI min LER databases, and the differences between the two retrievals. TROPOMI HCHO VCDs are globally increased by 10 to 20% when using TROPOMI min LER instead of OMI min LER (Figure 3.6d). Localized decreases are also observed over bright surfaces (Figure 3.6e and f), thanks to the improved spatial resolution of the new albedo database, allowing for a reduction of known HCHO VCD biases over bright scenes (see examples in the Terrascope data viewer https://bit.ly/48K8wu6 or https://bit.ly/3St70qL).

The results obtained for the other months are equivalent to the values obtained in July, as is shown on Figure 3.7 (a) for January, pointing to a systematic effect of the climatology version. Another test comparing the use of the directional LER instead of the minimum LER, both provided in the TROPOMI albedo database, is illustrated in Figure 3.7 (b). It shows the small impact of the directional LER in the UV wavelength range.

Because of the suspiciously high values of the GOME-2 LER database v4 at 340 nm, which is attributed to a wrong radiometric calibration in GOME2 spectra (confirmed by Tilstra et al. (2017)), the TROPOMI albedo database at 340 nm is the preferred choice for all sensors. Updating the albedo however requires a reprocessing of all the L2 datasets, which is time and resources consuming. This will be addressed in this project by creating intermediate L2 products, on which the CCI+ L3 product will be based.

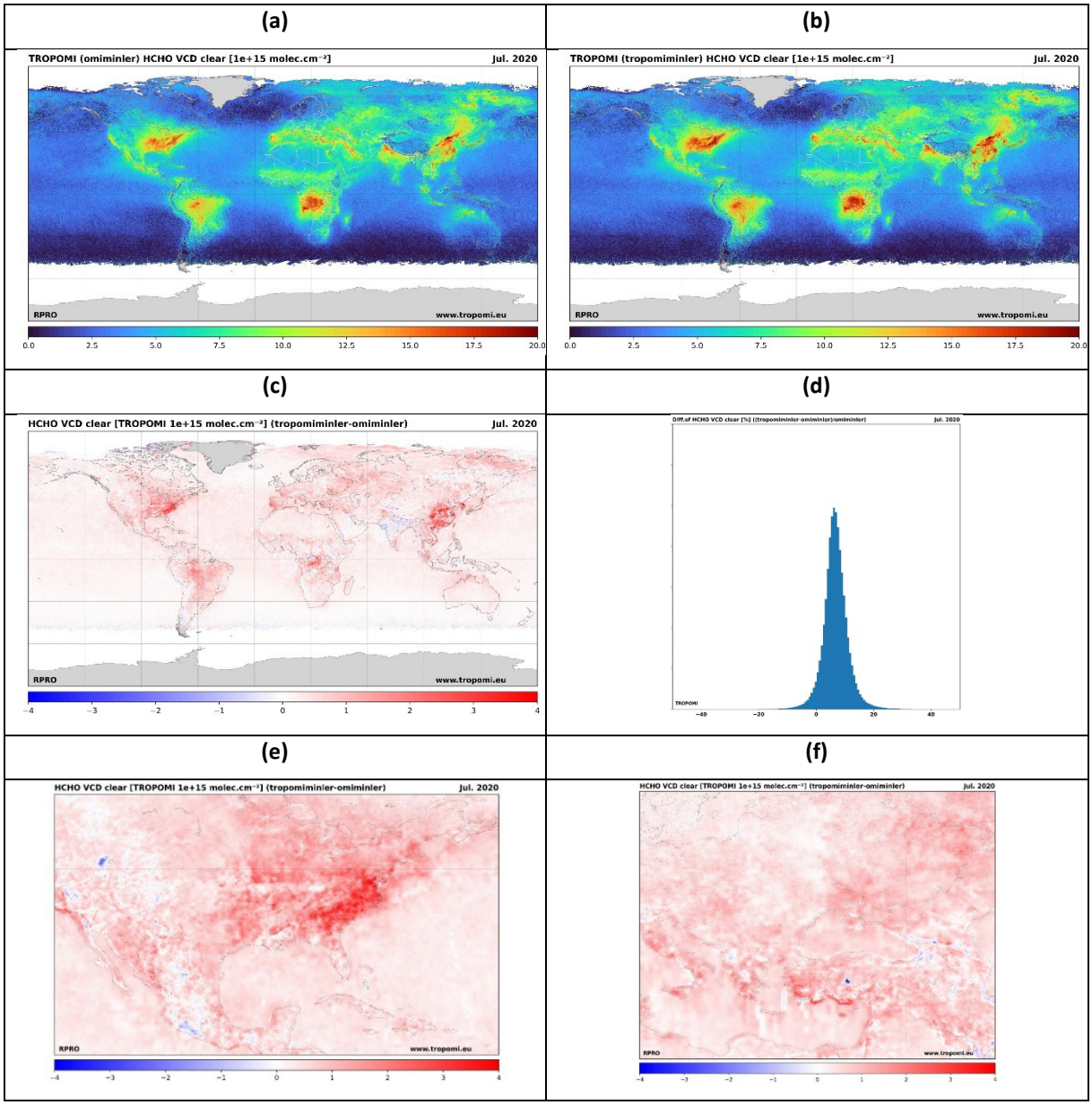


Figure 3.6: TROPOMI HCHO VCD clear in Jul. 2020 obtained using the OMI (a) or TROPOMI (b) min LER databases. Figures (c) and (d) show the absolute differences between (b) and (a) and the histogram of the relative differences, respectively. Figures (e) and (f) show the absolute differences zoomed over the US and Southeastern Europe.

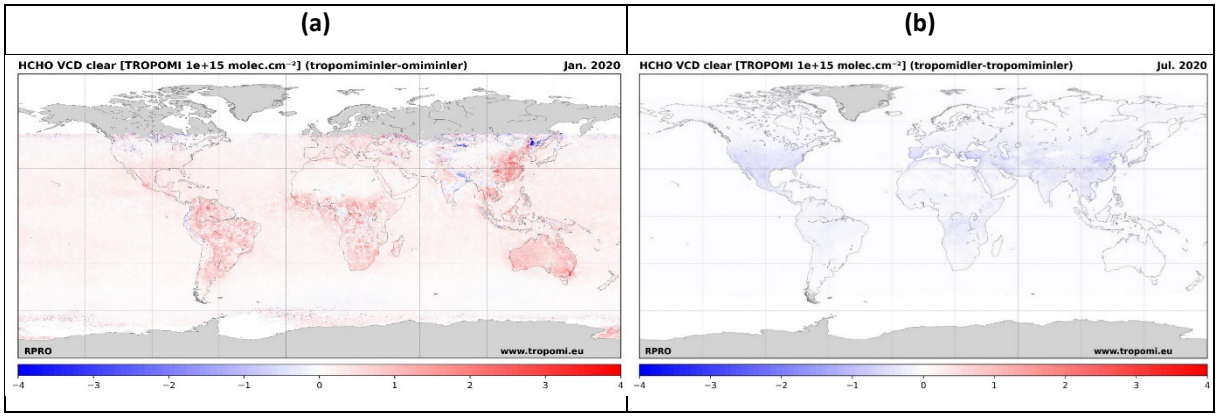


Figure 3.7: (a) Differences of TROPOMI HCHO VCD clear in Jan. 2020, as in Figure 3.6c. (b) Differences of TROPOMI HCHO VCD clear in Jul. 2020, obtained using TROPOMI DLER or min LER.

3.4.2. **A priori profiles**

The impact of the a priori profiles has been tested on one full year of GOME-2A (AC SAF and QA4ECV, 2012), OMI (2012) and TROPOMI (2021), at the global scale.

The use of three different CTMs has been evaluated:

- TM5(-MP) [1995-2017 (2018-now)], 1°, 30min, 34 levels (Huijnen et al 2010; Williams et al, 2017): currently used in the QA4ECV products from 1995 to 2017 (GOME, SCIAMACHY, GOME-2, OMI). Since 2018, the TROPOMI HCHO operational product uses profiles from the TM5-MP model in its forecast mode.
- MAGRITTE v1.1 (IMAGES) [2005-2021], 1°, 20min, 40 levels (Müller et al., 2019): A monthly climatology of IMAGES profiles is currently used in the AC SAF GOME-2 operational HCHO products.
- CAMS REA N123 [2003-2012], 80km, 3h, 60 levels (Iness et al., 2019).

Model simulations have been directly compared in terms of model HCHO columns, their seasonal and diurnal variations, and in terms of their vertical distribution. We have then quantified the sensitivity of the satellite HCHO column retrievals when replacing the TM5-MP (IMAGES) a priori profiles by either CAMS or MAGRITTE profiles.

Figure 3.8 shows an example of HCHO annual averages in 2021 simulated by the three models and measured by TROPOMI. Note that the TROPOMI VCDs are calculated using the TM5-MP a priori profiles. The differences between the model columns can be as large as 60% on annual average. TM5 columns are significantly larger than MAGRITTE and CAMS over the Amazon and African tropical forests. Figure 3.9 presents two examples of model profiles over Porto Velho (Brazil) and over Xianghe (China). The values of the model integrated columns, the profile height (pressure under which 70% of the model column lies), and the corresponding calculated satellite AMF are provided in the insets.

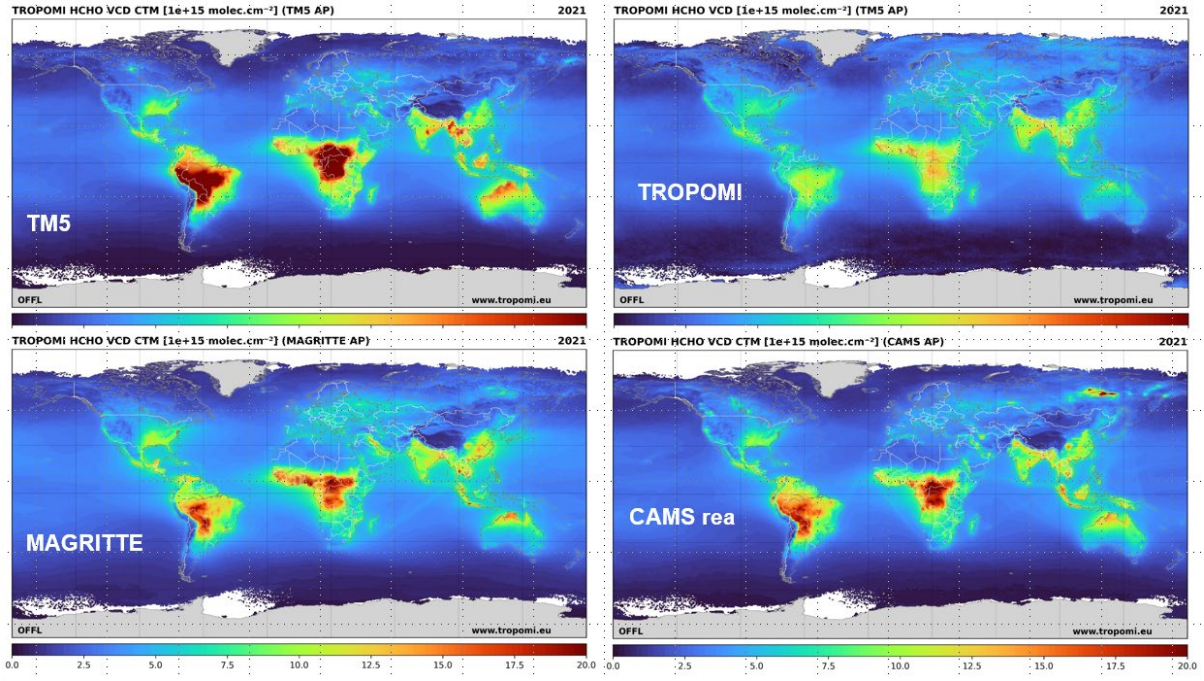


Figure 3.8: Annual mean of model HCHO columns in 2021 in the afternoon: TM5, MAGRITTE and CAMS REA, and comparison with TROPOMI HCHO columns.

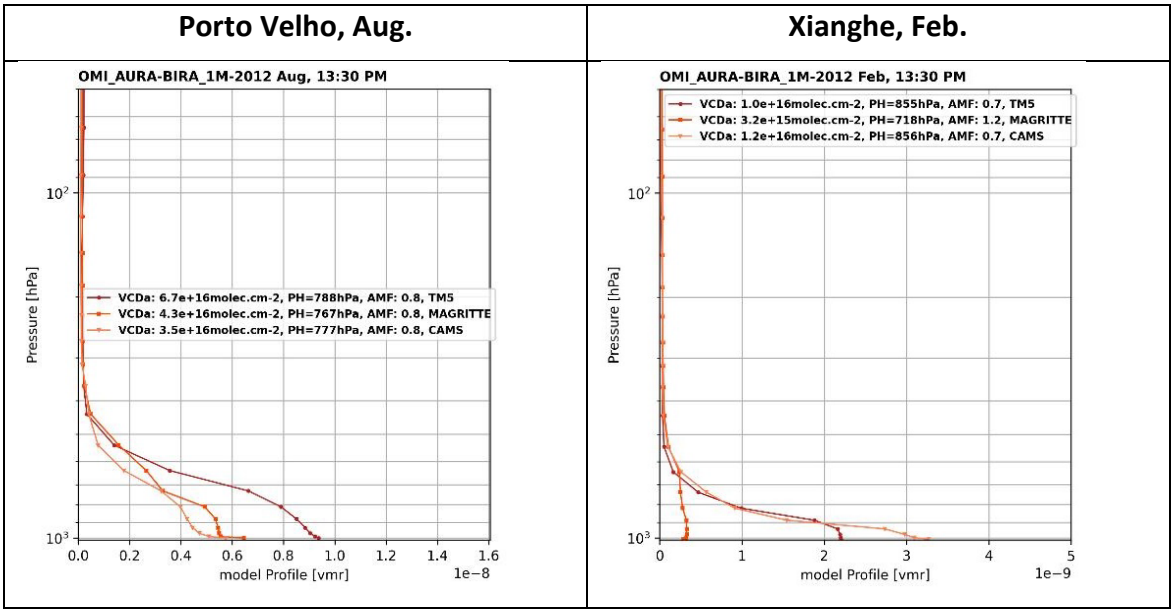


Figure 3.9: Comparison of HCHO model profiles over Porto Velho (Brazil) and Xianghe (China) respectively in August and in February 2012. Insets provide the values of the model integrated columns, the profile height (pressure under which 70% of the model column lies), and the corresponding satellite AMF.

In Figure 3.10, difference maps show the impact on the retrieved satellite HCHO columns when replacing the initial a priori profiles by either CAMS or MAGRITTE profiles. The effect ranges between ±30% on annual average. The changes in the satellite HCHO columns retrieved using different model profiles are not spatially correlated with the large differences observed between the model HCHO columns. Indeed, for DOAS-type AMFs, the relative vertical distribution of the profile (or normalized profile shape) has the most important impact on the

retrieved tropospheric columns, regardless of the model total column amount. This can be observed on Figure 3.9, that shows that equivalent AMFs are retrieved over Porto Velho despite the large differences between the model columns.

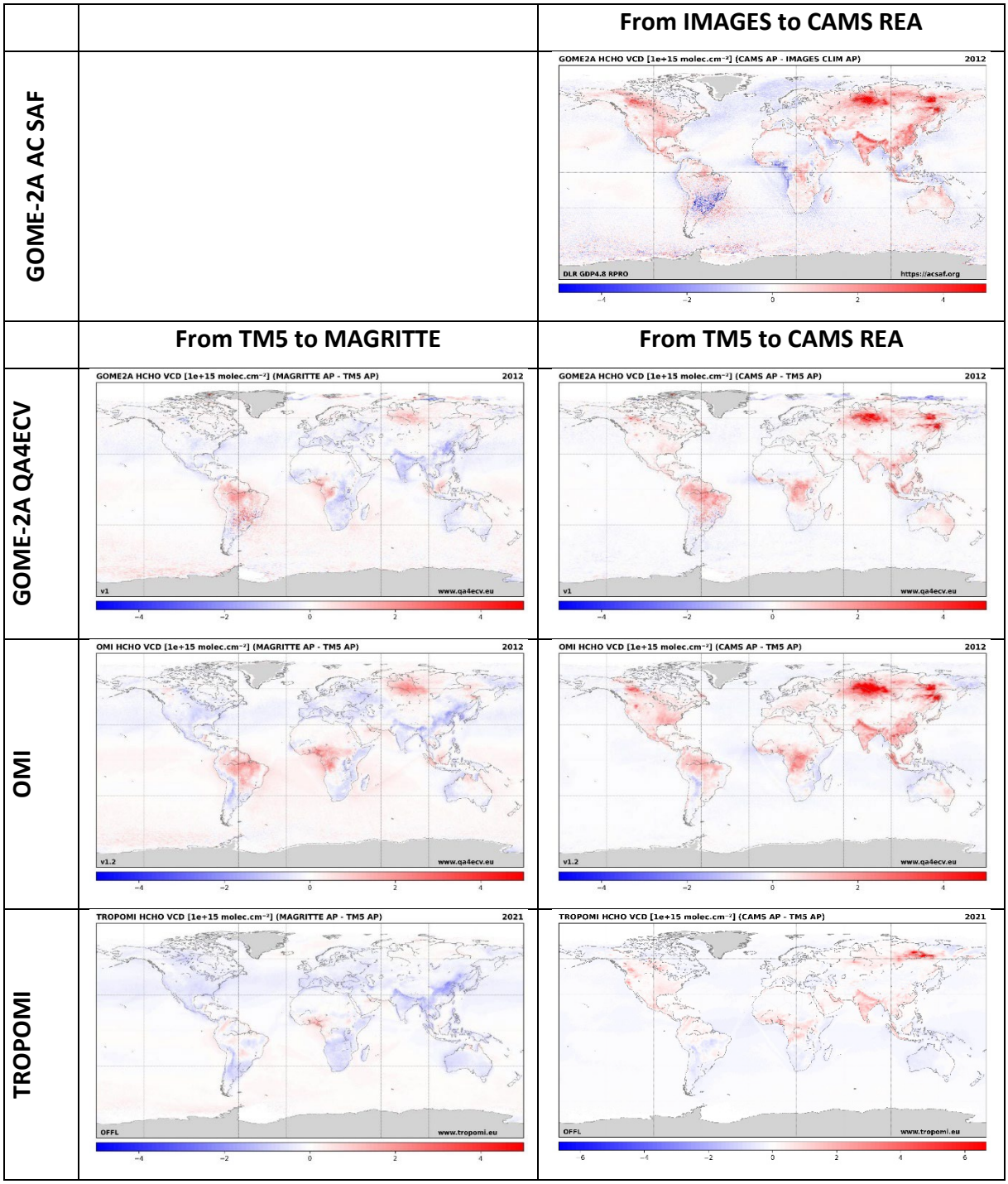


Figure 3.10: Absolute differences of satellite HCHO columns obtained when replacing TM5(-MP) profiles by MAGRITTE (first column) or CAMS REA (second column) profiles.

MAGRITTE and CAMS profiles lead to larger satellite HCHO VCDs over tropical forests and over fires. More generally, CAMS profiles lead to larger columns over all NMVOC emission regions (Asia, US), while MAGRITTE profiles generally lead to smaller satellite columns over Asia, Europe and America. The smallest impact is found when replacing the TM5-MP profiles by CAMS in TROPOMI retrievals in 2021. This points to differences between the TM5 and TM5-MP versions that have been used for OMI in 2012 and TROPOMI in 2021. No strong variation of the general patterns of the impact on the satellite columns is observed between morning and afternoon or from month to month (not shown), pointing to a systematic effect of the selected model for the a priori profiles.

3.4.3. **Clouds**

Similar to the NO_2 cloud analysis (see Section 2.4.3), comparisons of mean cloud fraction and height have been performed with GOME-2B time series over selected HCHO related regions: Northern China, the Amazon, the southeastern US, and the Equatorial Pacific.

The initial comparison has been done using the operational AC SAF OCRA/ROCINN CRB cloud algorithm (Valks, 2019) and FRESCO+ v7, used in the QA4ECV products. Both are based on the previous GOME-2 L1 data version. A second comparison was completed using the AC SAF RPRO product and FRESCO+ v8 which are both based on the new reprocessed GOME-2 L1 v6.3 data. Another important difference in Fresco+ v8 is the use of the latest GOME-2 DLER surface albedo database. See section 2.4.3 for more details on the cloud product versions.

From our analysis, the following conclusions can be drawn:

- For the cloud fraction, an improved consistency between the cloud products is systematically found when FRESCO+ v8 is used.
    - Mean cloud fraction from FRESCO+ v7 over Amazon appears with a positive relative difference of up to 50% for the months June-July-August (see Figure 3.11 a, left panel) compared to the AC SAF OFFL data. However, this discrepancy is almost removed when FRESCO+ v8 time series are compared to the AC SAF RPRO data (Figure 3.11 a, right panel).
    - The seasonal pattern variations of the cloud fraction over Amazon are strong. However, for SE US (Figure 3.11 b, left panel), the mean cloud fractions do not fluctuate so sharply with the relative differences between the mean cloud fraction of FRESCO+ v7 and OCRA/ROCINN being on average 25%. Nevertheless, also in this case the agreement for the mean cloud fraction improves significantly when using FRESCO+ v8 (Figure 3.11 b, right panel).
- For the cloud height, the consistency is only partly improved with FRESCO+ v8.
    - As can be seen in Figure 3.12 a, FRESCO cloud height has larger seasonal variations than OCRA/ROCINN. For the SE US region, a large discrepancy of about 70% remains present for the months March-April-May. For those months, higher HCHO AMFs are anticipated when cloud correction is done based on the OCRA/ROCINN algorithm. This results in lower HCHO columns compared to those found using FRESCO.

- Over Northern China, the FRESCO+ v8 mean cloud height has increased substantially for the months March-April-May. The relative differences with the OCRA/ROCINN cloud height increased to 50-80% (see Figure 3.12 b). However, for the same region, good agreement between FRESCO+ and OCRA/ROCINN cloud heights is found for the months November-December-January.

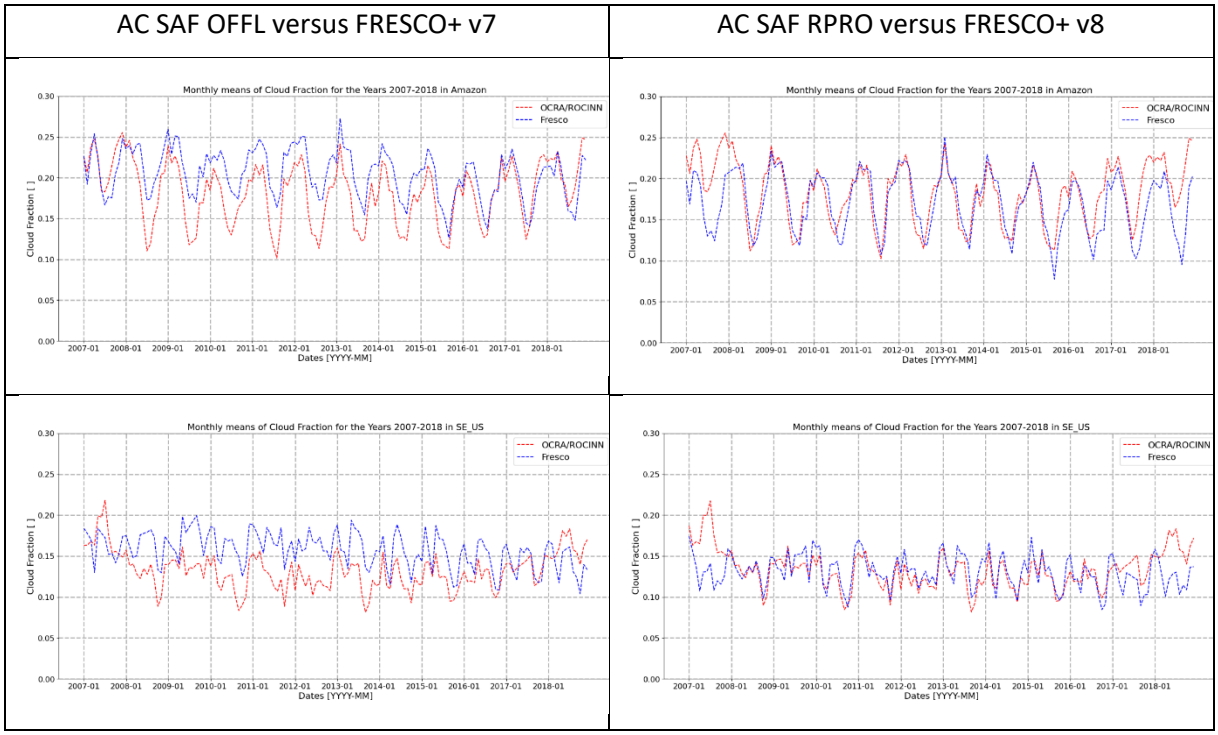


Figure 3.11: Time series of the monthly means of Cloud Fraction from AC SAF (red line) and FRESCO (blue line) over (a) Amazon, (b) SE US.

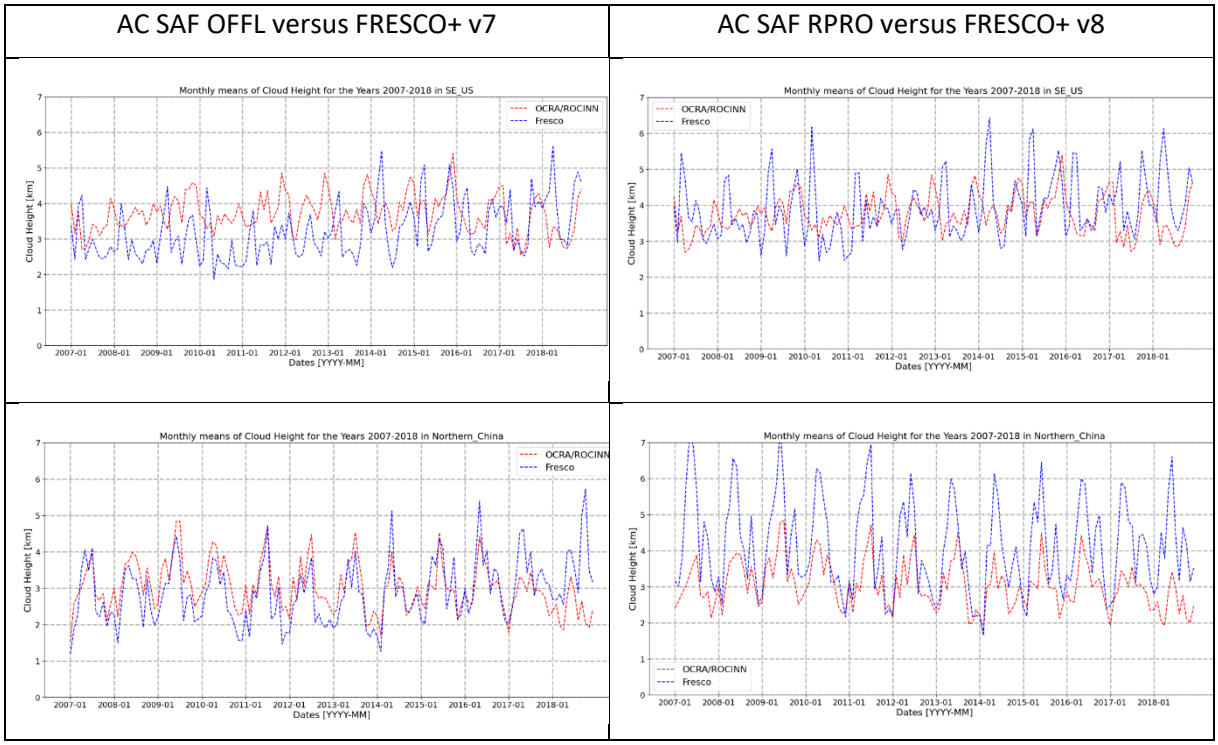


Figure 3.12: Time series of the monthly means of Cloud Height from AC SAF (red line) and FRESCO (blue line) over (a) SE US, (b) Northern China

In order to get a more general view of the HCHO retrieval sensitivity to the clouds, the impact of the cloud correction has been evaluated in the current AC SAF, QA4ECV and TROPOMI HCHO products over the full time series. To this end, we compared the HCHO columns with and without cloud correction (VCD or VCD_{clear}).

Figure 3.13 presents time series of the $\frac{VCD}{VCD_{clear}}$ ratio for all the sensors over Northern China and Amazonia. This ratio indicates the impact of the cloud correction on the final HCHO column and is a relative quantity for each product. A ratio larger than 1 indicates that the cloud correction increases the HCHO VCD.

The impact of the cloud correction ranges between ±20% and differs in sign between the different cloud products. The GOME-2A FRESCO+ v7 product leads to the largest decreases of the HCHO VCD, while the OMI O_4 cloud product leads to the largest increase of the HCHO VCD. The other cloud products have a smaller impact on the HCHO VCDs (±10%). Some trends related to trends in OCRA/ROCCIN clouds are observed in AC SAF GOME-2 datasets (RPRO).

- Over the Amazon:
  - The GOME, SCIAMACHY and GOME-2 FRESCO+ cloud products lead to a decrease of the HCHO VCD from 0 to 20%. The largest decrease is observed with GOME-2A FRESCO+ v7.
  - The OMI O_4 cloud product leads to an increase in HCHO of up to 20%.
  - The GOME-2 OCRA/ROCINN cloud product leads to an increase of HCHO VCD by 5 to 10%

- The TROPOMI OCRA/ROCINN cloud product leads to a decrease of the HCHO VCD of 0 to 10%.

- Over China:
  - The GOME and SCIAMACHY Fresco cloud products lead to moderate changes in the HCHO VCD of ±10%. The GOME-2A Fresco cloud product stands out with a larger decrease of the HCHO VCD up to -20%.
  - The OMI O_4 cloud product leads to an HCHO increase of about 10%.
  - The GOME-2 and TROPOMI OCRA/ROCINN cloud products lead to moderated changes in the HCHO VCD of ±10%.

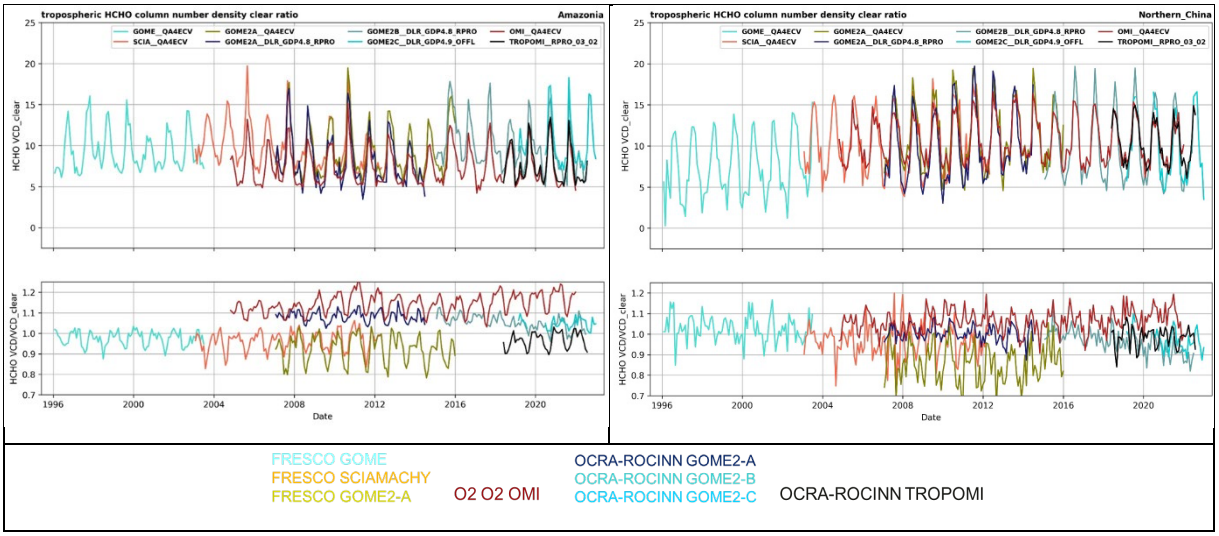


Figure 3.13: Monthly average of HCHO VCDs obtained from GOME, SCIAMACHY, GOME-2A, B, C, OMI, and TROPOMI measurements over Northern China and the Amazon regions. The upper plots present the HCHO VCD clear [units: 10^{15} molec/cm^2]. The lower plots present the ratio of VCD/VCD_clear [no unit] as a quantification of the cloud correction impact in each product.

3.5. Validation

A study was done on QA4ECV and AC SAF product variants, for the sensor GOME-2A. Two aspects were varied: (1) application of cloud correction or using clear-sky assumption, (AC SAF and QA4ECV products) and (2) the prior profile (for QA4ECV product only): TM5 (standard product), CAMS reanalysis and MAGRITTE, and with for AC SAF: IMAGES (standard product) and CAMS reanalysis. Details on the validation methods used are summarised in Annex A2 and A3.

3.5.1. Cloud-corrected vs clear-sky columns

Figure 3.14 shows the comparison of GOME-2A AC SAF columns with ground-based FTIR measurements, using the default cloud-corrected values and the clear-sky assumption. Hardly any difference is visible, over the full range from clean to polluted sites. This demonstrates that for the AC SAF product OCRA/ROCINN, the cloud correction is a very minor effect. A similar result is obtained using MAX-DOAS data (not shown).

Figure 3.15 shows the comparison of GOME-2A QA4ECV columns with MAX-DOAS, using the default cloud-corrected values and the clear-sky assumption. Generally, biases decrease by

10-20% when using the clear-sky assumption. Also, the Pearson-R correlation and the least-square regression slope improve in 6 out of 7 cases for the clear-sky case (not shown). Dispersions mostly increase, but this can be attributed to the increase in amplitude of the satellite columns.

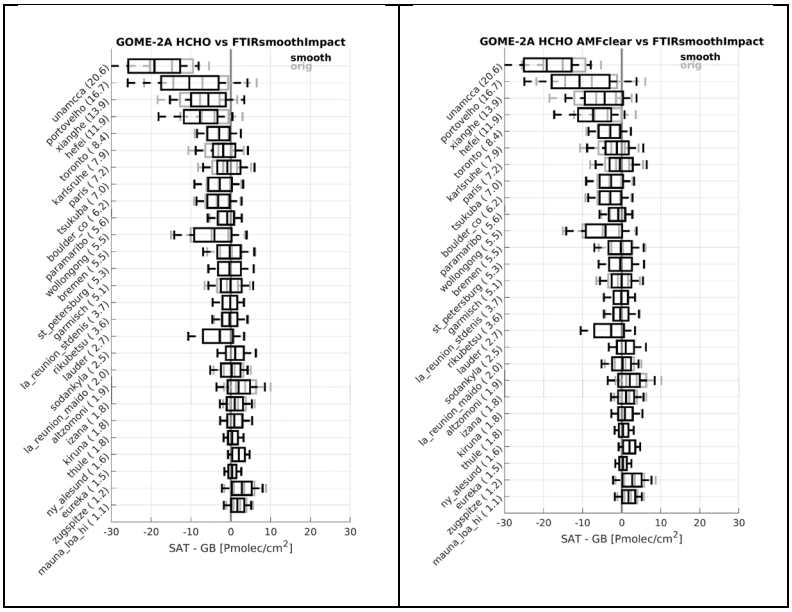


Figure 3.14: HCHO comparison of cloud-corrected (left) and not cloud corrected (right) GOME-2A AC SAF columns with ground-based FTIR measurements [Pmolec/cm²=1e15molec/cm²].

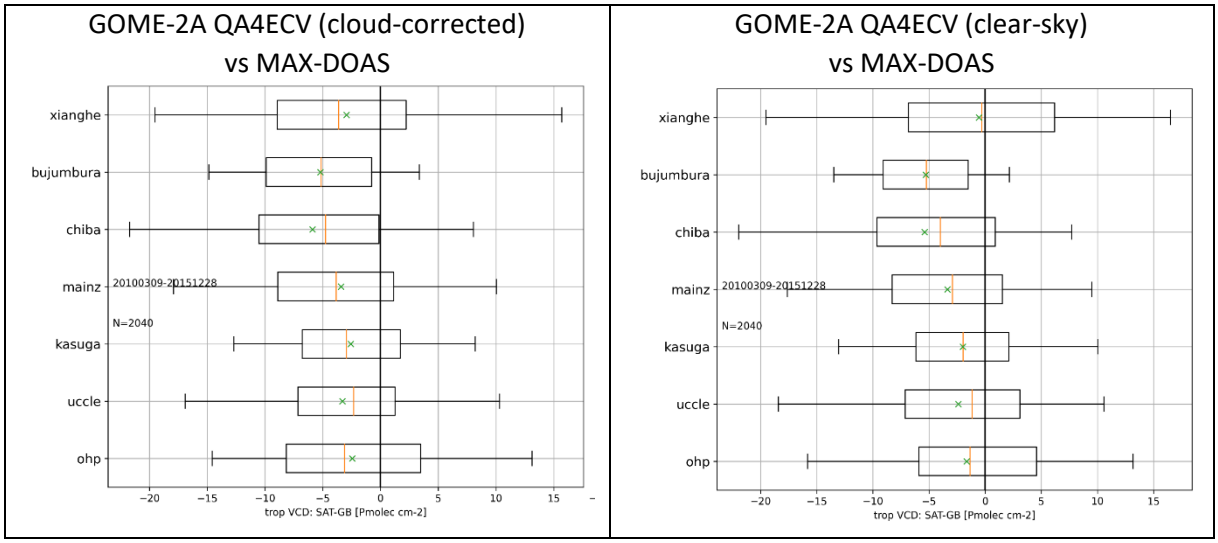


Figure 3.15. HCHO comparison of cloud-corrected (left) and not cloud corrected (right) GOME-2A QA4ECV columns with ground-based MAX-DOAS observations [Pmolec/cm²=1e15molec/cm²].

The impact on validation results of cloud-correction vs clear-sky assumption, for AC SAF (virtual no impact) and for QA4ECV (more negative bias for cloud-corrected case), is consistent with the AMF ratios in Figure 3.13, and indicates that the clear-sky assumption is in better agreement with MAX-DOAS observations.

3.5.2. **Prior profile choice**

Table 3-2 presents statistics of GOME-2A QA4ECV variant products, generated for the year 2012, vs bePRO MAX-DOAS measurements in Xianghe. Here the default cloud correction is used, and no vertical smoothing is applied in the comparisons. Comparing the impact of different prior profiles over Xianghe, one finds that TM5 and CAMS lead to similar bias and dispersion, and similar robust regression and correlation statistics. Assuming MAGRITTE as prior leads to an increased bias and lower slope, but a smaller dispersion of the differences. Those results agree with Figure 3.10 showing annual maps of HCHO VCD differences obtained when replacing TM5 profiles by MAGRITTE or CAMS REA profiles. The prior profile impact is expected to be larger in the tropics, and in the afternoon.

Table 3-2. Comparison statistics of GOME-2A QA4ECV variant products vs bePRO MAX-DOAS in Xianghe, year 2012, no vertical smoothing. Variant products are: (1) prior=TM5, (2) prior=TM5, clear-sky assumption, (3) prior=CAMS reanalysis, (4) prior=MAGRITTE.

	Bias	Dispersion	Regression	correlation
TM5 (orig)	Mean: -3.2P,-17% Median: -3.9P,-26%	Std: 12.2P ½IP68: 9.1P,79%	OLS: y=0.62x+2.51 TS: y=0.64x+1.65	Rpe=0.42 RSp=0.52
CAMS	Mean:-3.0P,-17% Median: -3.9P,-27%	Std: 11.9P ½IP68: 9.1P,78%	OLS: y=0.66x+2.08 TS: y=0.64x+1.20	Rpe=0.45 RSp=0.52
MAGRITTE	Mean:-5.2P,-32% Median: -5.2P,-41%	Std: 9.8P ½IP68: 8.0P,61%	OLS: y=0.53x+1.70 TS: y=0.52x+1.15	Rpe=0.47 RSp=0.51

Bias and standard deviation are calculated from SAT-GB and (SAT-GB)/GB. Std=standard deviation, IP68: 68 interpercentile, OLS: ordinary linear least-squares regression, TS: robust Theil-Sen regression, Rpe: Pearson correlation coefficient, RSp: Spearman correlation coefficient. [P=1e15molec/cm^2].

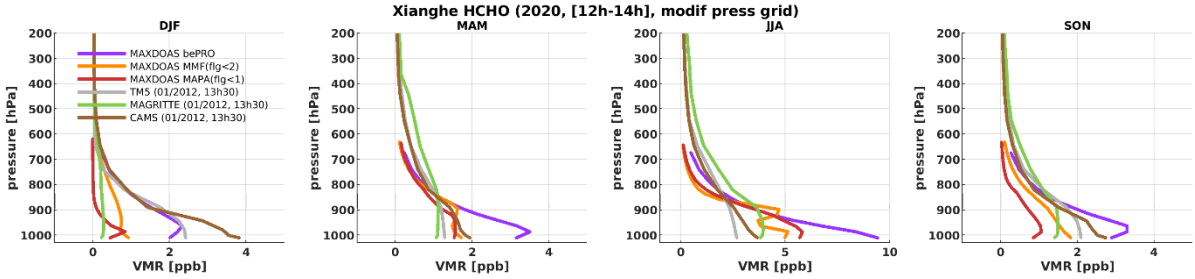


Figure 3.16. Seasonally averaged HCHO profiles at Xianghe: model profiles TM5, CAMS, MAGRITTE and MAX-DOAS profiles retrieved with different settings: bePRO, MMF and MAPA (the latter two being developed for FRM4DOAS).

Figure 3.16 shows seasonally averaged profiles of the models TM5, CAMS and MAGRITTE and of MAX-DOAS obtained with different algorithms: bePRO, MMF and MAPA (latter two being part of FRM4DOAS development). There are significant differences not only between the models used as prior, but also between the different MAX-DOAS retrieval algorithms. It would be interesting to redo the comparison statistics for Table 3-2, but this is beyond the scope of this work.

3.6. Algorithm Selection

We summarize here the main conclusions from the Round Robin study. The details of the selected retrieval settings can be found in the ATBD of the CCI+ precursors project.

The GOME-2 AC SAF prototype slant column retrieval and background correction will be mostly aligned to the QA4ECV settings. The main features are: one large fitting interval, daily averages of radiances selected in the reference sector as DOAS reference, pseudo cross-sections to minimize East/West bias and an iterative spike removal algorithm. The background correction is based on the VZA and latitudinal dependency of the slant columns in the reference sector. The background model column which is added in a final step has not been investigated. It will depend on the AMF parameters, especially on the a priori model columns.

For the AMF calculation, for all the sensors.

- The TROPOMI minimum LER is selected because of its improved spatial resolution. The same climatology is chosen for all sensors, morning and afternoon, because the discrepancies of the different databases are too large in the UV. Using different databases would lead to systematic biases in the HCHO columns retrieved from the different sensors. The climatology has a spatial resolution of 0.125° x 0.125° and, for OMI, it will be used 'as is'. For low resolution sensors as GOME and SCIAMACHY, the TROPOMI climatology will be degraded to the appropriate resolutions.
- For the a priori profiles, the CAMS reanalysis model is selected for the following reasons:
    - Validation results are expected to be improved thanks to an increase of the columns over NMVOC emission regions, and a decrease of the background columns.
    - The reanalysis dataset uses a constant version from 2003 to 2022.
    - CAMS provides an operational service that could be used for future extension.
- The main HCHO VCDs will be provided without applying a correction for the cloud altitudes, but only a cloud filtering based on the cloud fractions. This choice is made to minimize biases between sensors coming from differences between the cloud products. Furthermore, the impact of the cloud correction is generally limited and did not improve the validation results. However, the cloud-corrected values will be provided in the L3 files of all the satellite HCHO products.

From the sensitivity tests that have been performed, uncertainties can be estimated, as well as their random or systematic behaviour in time. Our approach is to describe an uncertainty as systematic if its impact is not significantly reduced by averaging the data in time. For example, the standard deviation of the HCHO SCD is greatly reduced when averaging the observations. We therefore consider measurement noise to be the dominant contribution to slant column uncertainty (although there are systematic contributions from inaccurate cross section spectra, modelling errors, etc.). The impact of the LER, a priori profile and cloud correction selection show clear spatial patterns in the annually averaged columns, as shown by figures 3.6, 3.10 and 3.13. Some of the differences between the auxiliary datasets are linked to geophysical parameters (bright surfaces, land cover, emission regions, …). Those sources of

uncertainty are therefore classified as mostly systematic, although part of them also shows a random behaviour (see the discussion in section 2.6). This is summarized in Table 3-3.

Table 3-3: Summary of HCHO sensitivity analysis for each of the retrieval tests.

Test	Sensitivity	Behaviour in time
ΔSCD	±1e16 molec./cm^2	Mostly random
LER	From -10% to +20%	Mostly systematic
Profiles	±30%	Mostly systematic
Clouds	±20%	Random and systematic
VCD_{cor}	±0-2e15 molec./cm2 *	Mostly systematic

* not evaluated during this project. Taken from De Smedt et al. 2018.

4. SO_2

4.1. Data and Algorithms used

The SO_2 algorithmic development relies on the TROPOMI Level-2 processor development (Theys et al., 2017) and, more recently, on the Covariance-Based Retrieval Algorithm (COBRA) implementation in the S5P-PAL pre-operational environment. While the TROPOMI operational processor is based on Differential Optical Absorption Spectroscopy (DOAS) technique for the spectral fitting module, it is planned to be replaced by a COBRA approach soon (end 2024). For the CCI project, it has been proposed to use the COBRA method to generate SO_2 CDRs from the GOME, SCIAMACHY, OMI and TROPOMI sensors. Therefore, it is important to evaluate the performance of COBRA against other retrievals methods. This is the objective of the first part of the SO_2 Round Robin activities (section 4.2). Here we focus on OMI data, for two reasons: (1) COBRA and DOAS results from TROPOMI have already been compared in Theys et al. (2021), (2) the comparison also includes data from the Principal Component Analysis (PCA) technique (Li et al., 2020), developed initially for OMI. The different datasets used are listed below:

- OMI COBRA CCI version 1
- NASA OMSO2 PCA version 3
- OMI DOAS Scientific product (Theys et al., 2015).

The second part of the SO_2 Round Robin activities deals with the impact on the air mass factor of using different auxiliary data for the surface albedo (section 4.3.1), the model a priori profiles (section 4.3.2) and the cloud products (section 4.3.3). The sensitivity tests are performed using COBRA results (unless stated otherwise), either from OMI or TROPOMI.

4.2. Fitting approaches

The OMI SO_2 slant column retrievals from COBRA, PCA and DOAS have been intercompared for 4 months (February, May, August, November, covering 4 seasons) at the beginning (2006), middle (2012) and end (2020) of the OMI lifetime. For this, the same pixel selection is used for

the three data sets: SZA less than 65°, cloud fraction (based on OMCLDO2.003) less than 30%, and pixels unaffected by the row anomaly.

Figure 4.1 presents a comparison of the three algorithms for the month of August 2006 over China, as an example. Overall, the three retrievals are in good agreement with each other. Comparable elevated SO_2 slant columns are found in the same emission regions, e.g., North and South-Central China, Japan, and the Philippines. As can be seen, COBRA and PCA agree particularly well. This likely reflects the fact that both methods rely on the same concept, i.e., the use of a set of radiance measurements to represent the background contribution of the spectra. The DOAS SCDs are typically larger than COBRA/PCA by ~25% and this discrepancy increases for winter months (not shown) reaching 40%. However, one can observe from Figure 4.1 that the DOAS results are also characterized by larger noise and offsets. This behavior is exacerbated at high SZA (hence for winter period in China).

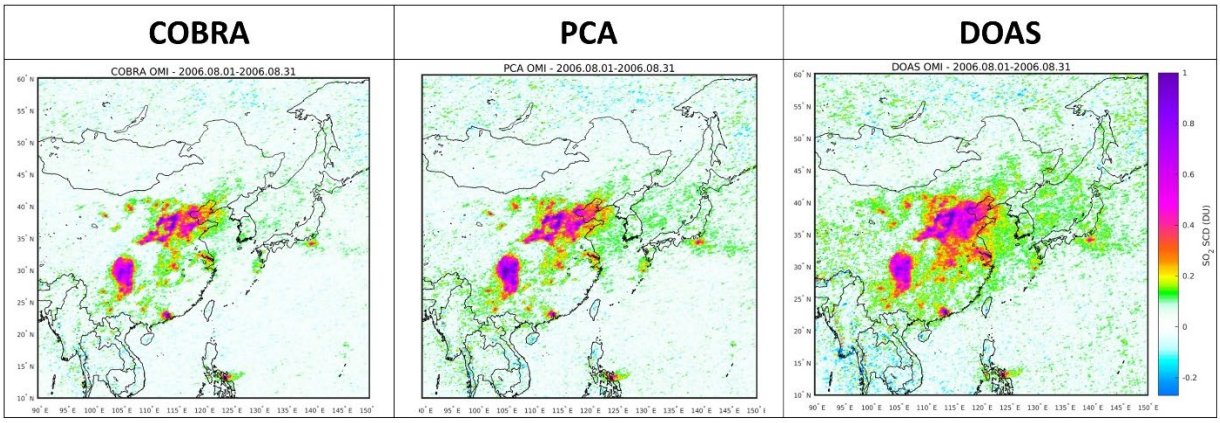


Figure 4.1: OMI 0.1°x0.1° monthly averaged SO_2 slant columns (in DU) retrieved with the (left) COBRA, (center) PCA and (right) DOAS techniques, for August 2006 over China.

To provide a quantitative assessment of the COBRA and PCA SO_2 products, Figure 4.2 shows the scatter plots of SO_2 slant columns from the two methods for 4 months in 2006 over China. This confirms the very good agreement between COBRA and PCA. From the fit parameters, it appears that COBRA is lower than PCA by ~ 10-25% depending on the season. The largest difference is found in winter, but this is also the period when the data scatter is the largest. As a matter of fact, COBRA is found to have less artefacts and lower noise than PCA.

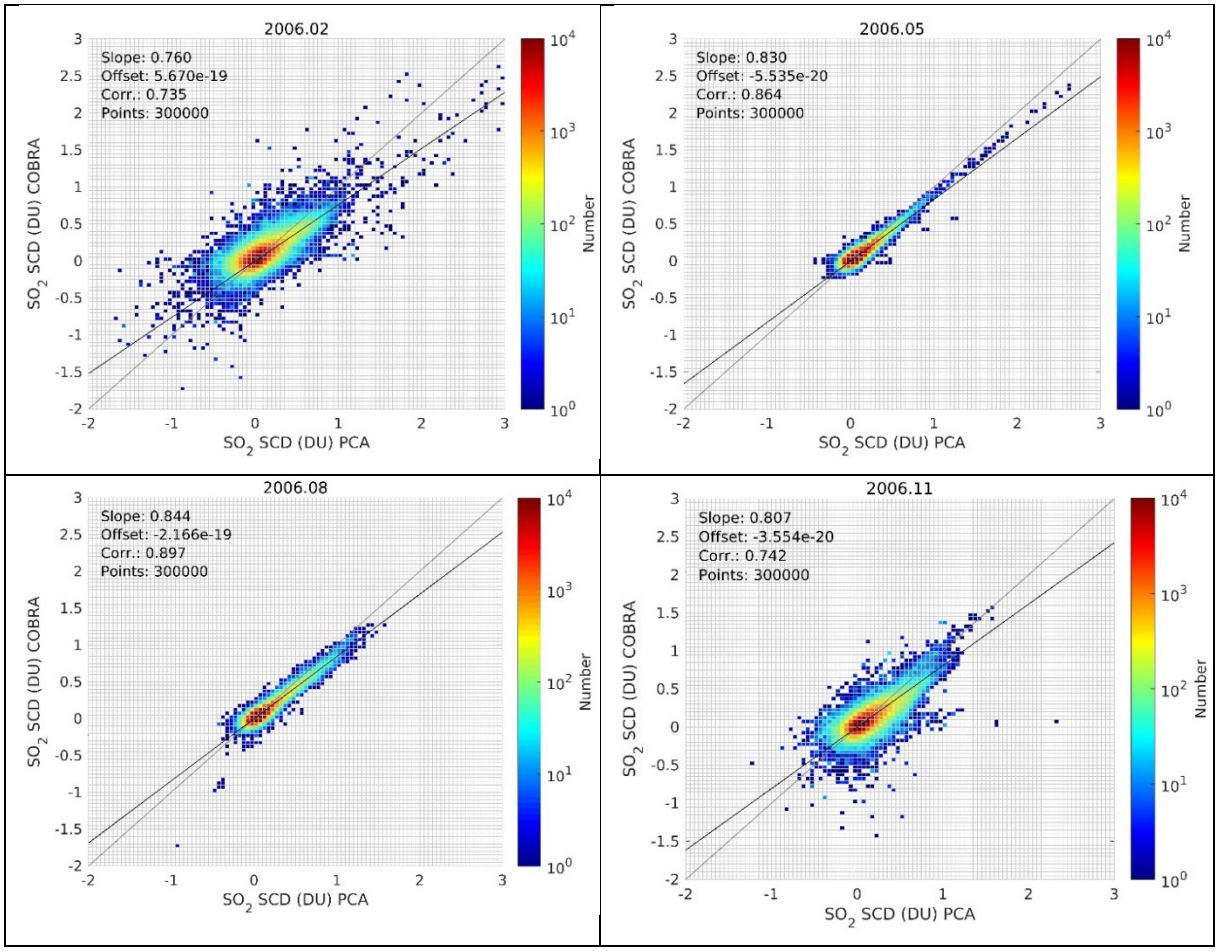


Figure 4.2: Scatter plots of (gridded) OMI SO_2 slant columns from COBRA (y-axis) and PCA (x-axis) for 4 months (Feb, May, Aug, Nov) in 2006, for the geographical region shown in Figure 4.1 (10-60°N, 90-150°E). The slope and offset of the regression line and the correlation coefficient are given as an inset for each plot.

It should be noted that, for this intercomparison exercise, the COBRA and PCA retrievals have used identical SO_2 cross-section data, to avoid any difference from the spectroscopy. However, there are slight differences in the retrieval settings, e.g., the spectral ranges used for the fit are not exactly the same (COBRA: 310.5-326 nm, PCA: 310.5-340 nm), and this can ultimately lead to differences in the results. Similarly, the DOAS settings are not fully in line with the ones from COBRA. The fitting interval is 312-326 nm. This can partly explain some of the observed differences, because the inclusion of the intense absorption band at 310.8 nm typically leads to a reduction of the noise on the data but also of the SCD values (due to a stronger impact of ozone on the SO_2 AMFs).

We have conducted the same exercise for other years (than 2006) and regions, and found comparable results.

4.3. AMF Impacts

4.3.1. LER

The impact of the Lambertian Equivalent Reflector (LER) used as surface albedo in the AMF calculation has been tested on TROPOMI data for 4 months (January, April, July, October 2020). Two albedo climatologies are used:

- OMI 0.5°x0.5° surface minimum LER of Kleipool et al. (2008) used in the operational TROPOMI SO_2 processor.
- TROPOMI 0.125°x0.125° surface minimum LER v2.0 of Tilstra et al. (2023).

Figure 4.3 shows the global maps of LER values at 328 nm from OMI and TROPOMI climatologies for two selected months, April, and July. As can be seen, the two data sets are generally consistent outside of the snow/ice regions, at high latitudes. However, the TROPOMI values are globally lower than the OMI data by 0.01 to 0.02.

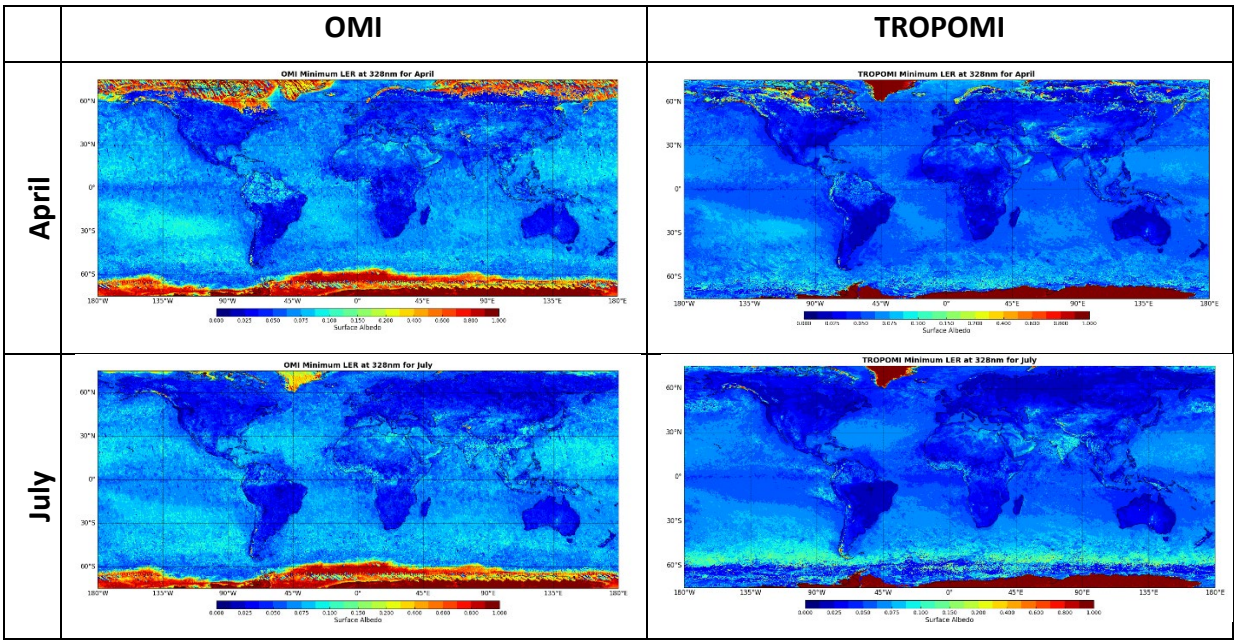


Figure 4.3: Comparison of April and July minimum LER values at 328nm from the OMI and TROPOMI albedo climatologies.

Figure 4.4 presents the relative differences between the TROPOMI SO_2 AMFs (no cloud correction) for 4 months in 2020, obtained using the TROPOMI and OMI minimum LER values. The large differences at high latitudes are due to different classification of snow/ice scenes between OMI and TROPOMI databases and are not discussed further. Overall, the use of TROPOMI LER instead of OMI LER leads to a decrease in the AMFs of 5-10%. However, in some locations and months, the relative differences can be larger, and have the opposite sign, but remain mostly within -20% to 20%.

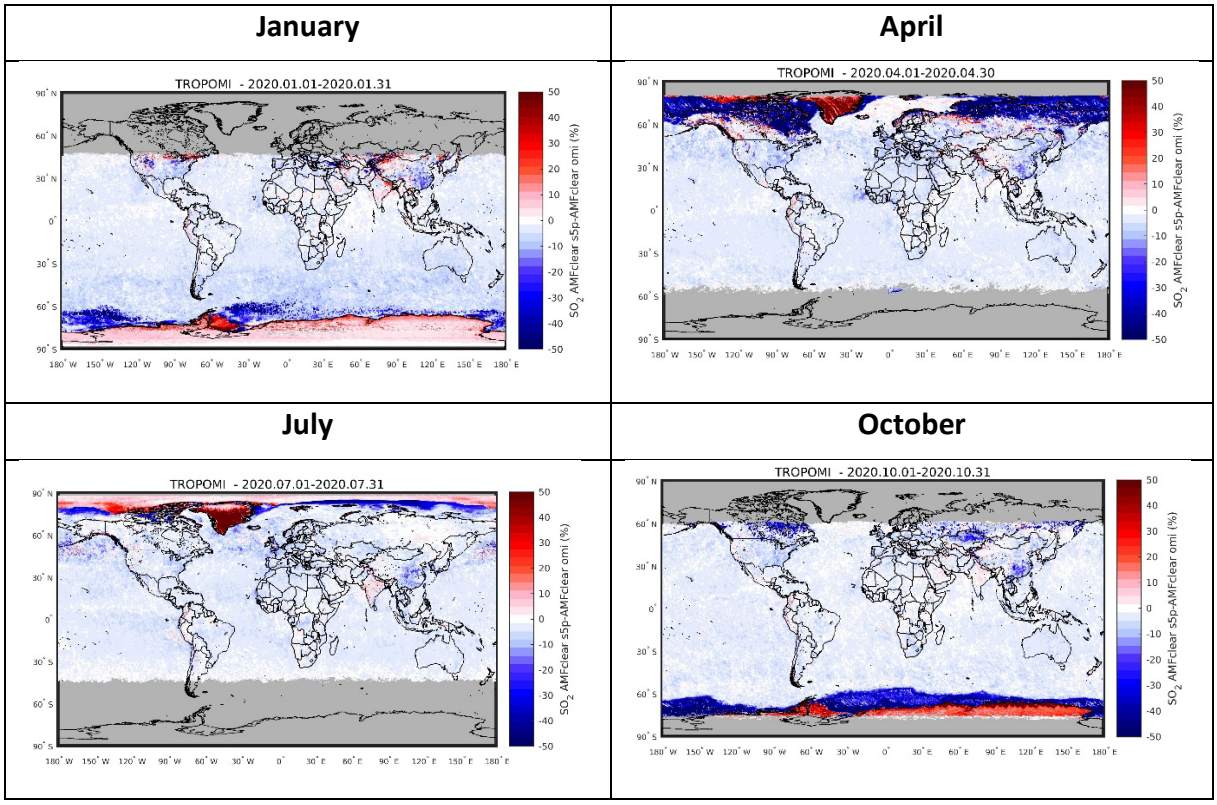


Figure 4.4: Relative differences of SO_2 clear-sky air mass factors obtained using TROPOMI and OMI minimum LER, for the months of January, April, July, and October 2020. Pixel selection criteria: SZA less than 65°, cloud fraction (CRB) lower than 30% and TROPOMI rows 26– 424.

Figure 4.5 presents the comparison of SO_2 vertical columns retrieved by TROPOMI using the two different albedo climatologies (TROPOMI and OMI minimum LER), over selected emission regions: Northeast India [20-23°N, 81-87°E], Northeast China [33-42°N, 110-124°E], South Africa [21-29°S, 24-36°E], Mexico [18-21°N, 96-103°W]. From Figure 4.5, one can conclude that the impact of changing the surface albedo on the VCDs is modest in these regions, usually not more than 5% on average.

Finally, we have performed a test where the TROPOMI minimum LER was replaced by the Directional Lambertain Equivant Reflector (the so-called DLER) of Tilstra et al. (2023). The impact on the retrieved SO_2 VCDs was found to be small, consistent with the findings for HCHO (section 3.4.1).

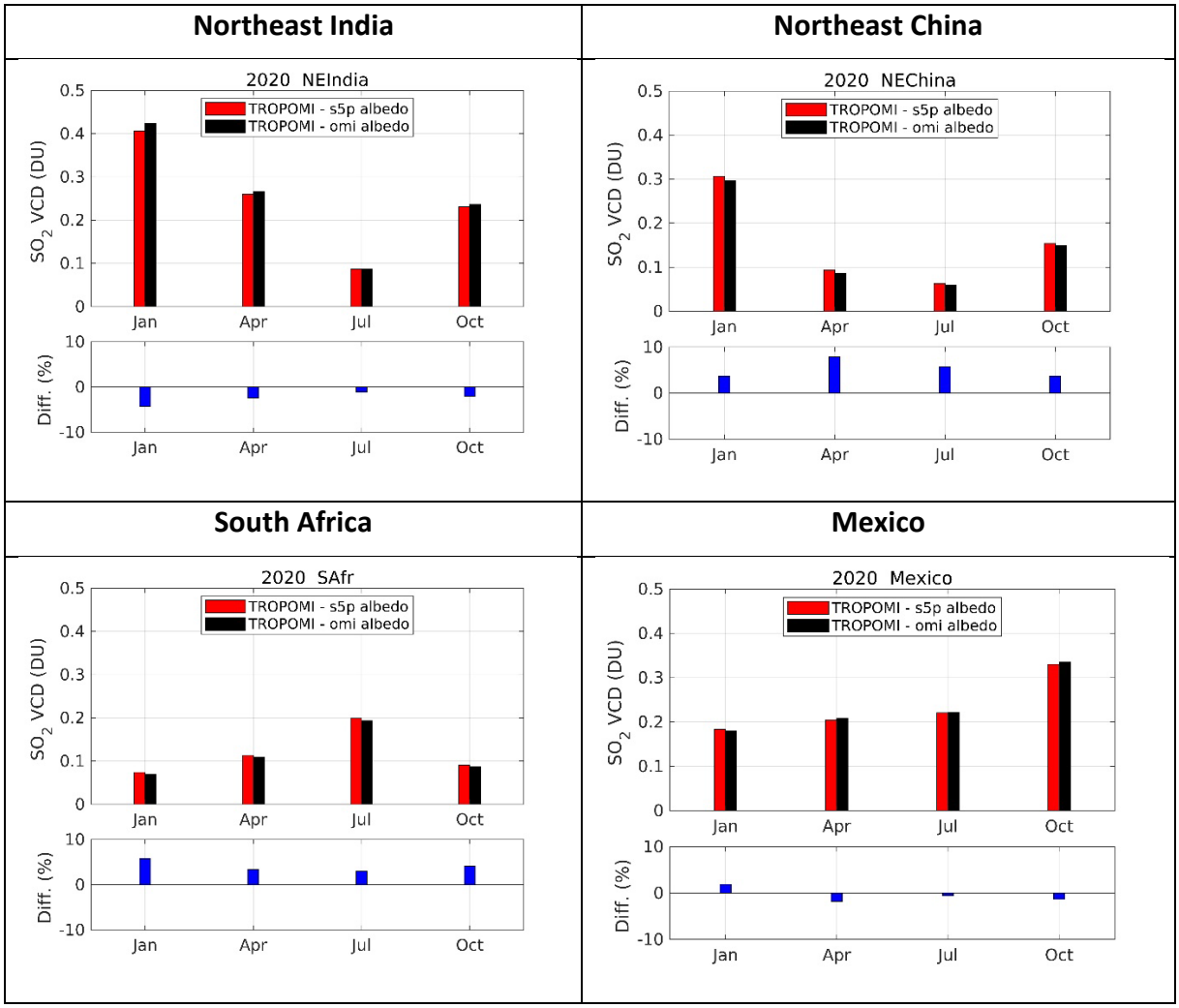


Figure 4.5: Comparison of TROPOMI monthly averaged SO_2 vertical columns using TROPOMI (red) and OMI (black) albedo climatologies as input of the (clear sky) AMF calculation, for January, April, July, and October 2020. The VCD results and relative differences are presented for 4 selected regions (Northeast India, Northeast China, South Africa, and Mexico). The same pixel selection criteria are used as in Figure 4.4.

4.3.2. **A priori profiles**

The impact of the SO_2 a priori profiles on the retrieved SO_2 columns has been assessed on one full year of OMI (2012) and two months of TROPOMI (April and October 2020), at the global scale.

Two different models were used:

- TM5(-MP) [1995-2017 (2018-now)], 1° x 1° resolution, 30 minutes time step, and 34 levels (Huijnen et al., 2010). The SO_2 profiles from TM5-MP are used as input of the TROPOMI SO_2 operational algorithm in its foreast mode.
- CAMS reanalysis (REA) N128 [2003-2022], 80 km resolution, 3hour time steps, and 60 levels (Inness et al., 2019).

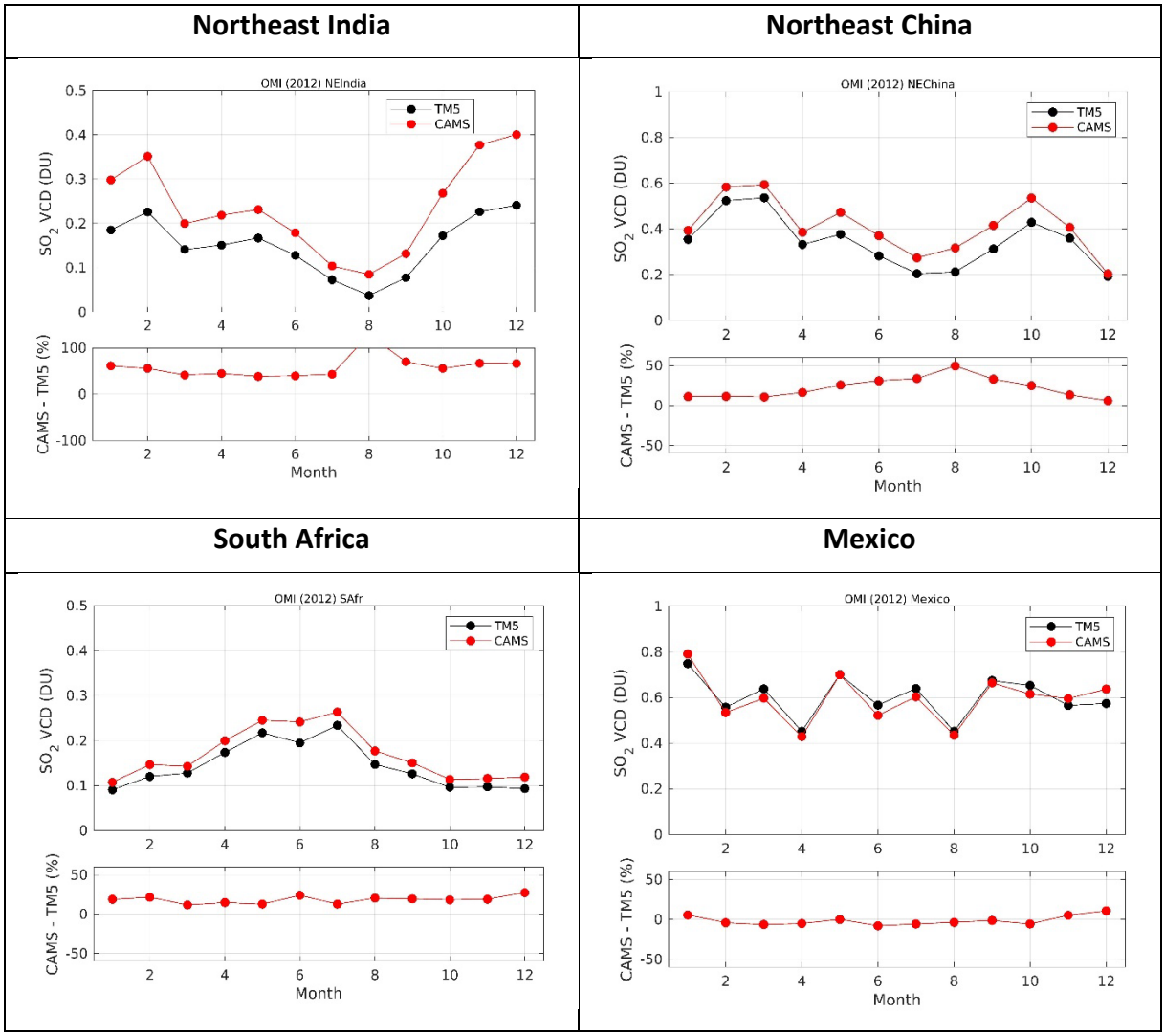


Figure 4.6: Comparison of OMI monthly averaged SO_2 vertical columns using CAMS and TM5 SO_2 apriori profiles, as input of the (clear sky) AMF calculation. The VCD results and relative differences are presented for 2012 and for the 4 selected regions used in Figure 4.5. The same pixel selection criteria are used.

We have quantified the sensitivity of the satellite SO_2 column retrievals by changing the model apriori profiles. Figure 4.6 shows the OMI monthly averaged SO_2 vertical columns using TM5 and CAMS for the 4 regions used in section 4.3.1, as well as the relative differences (OMI CAMS-OM TM5). We find a large impact of the a priori profiles on the SO_2 VCD results, in the range of 10-100%. The use of CAMS generally leads to an increase in SO_2 columns compared to TM5. Inspection of the vertical profile shapes reveals that, overall, the CAMS profiles are more peaked towards the surface than TM5 profiles. Yet, the impact on the SO_2 columns varies with the location and season. In Figure 4.6, the lowest differences are found over Mexico and the highest in Northeast India.

In Figure 4.7, we investigate the impact of changing the profiles on the SO_2 column maps retrieved by TROPOMI for October 2020 (as an example), over a region with large SO_2 emissions from oil industry (in the Middle East) and coal-fired power plants (in India). In relative values, the differences (CAMS – TM5) is mostly positive and in the 0-100% range (or

even above). There is a large variation depending on the location. E.g., in the Persian Gulf, the differences are about 40%, while in India, the SO_2 VCDs are +100 % larger with CAMS than with TM5.

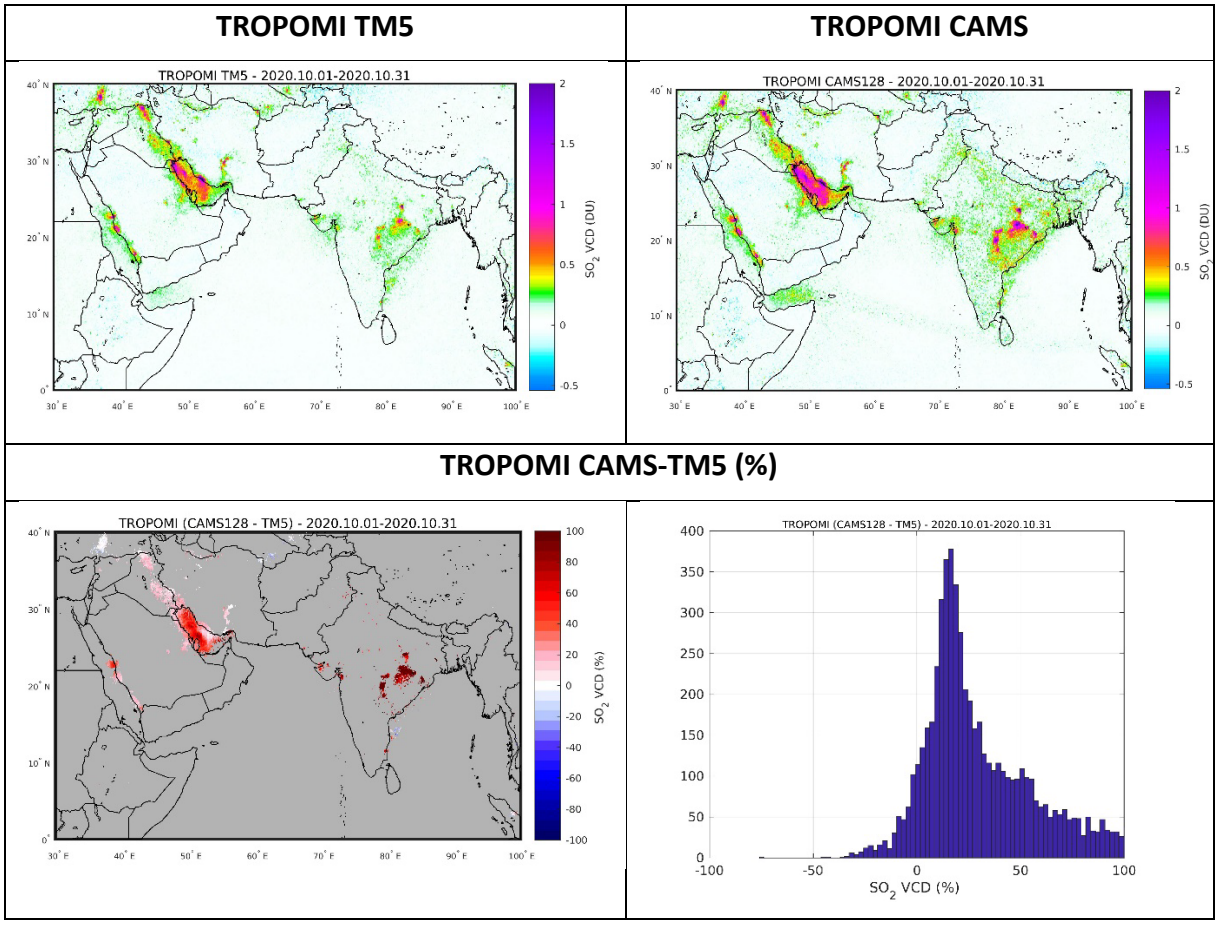


Figure 4.7: Comparison of TROPOMI monthly averaged SO_2 vertical columns (October 2020) using CAMS and TM5 SO_2 apriori profiles, as input of the (clear sky) AMF calculation. The maps of VCDs (top left: TM5, top right: CAMS) and relative differences (bottom left) are shown over the region covering the middle East and India [0-40°N, 30-100°E]. The histogram of the relative differences is displayed on the bottom right panel. The same pixel selection criteria are used as previously. Fo the relative differences, only SO_2 VCDs (based on TM5 profiles) larger than 0.3 DU are considered.

4.3.3. **Clouds**

To test the impact of clouds on the SO_2 VCD results, first a comparison was performed on 4 complete days (2018/06/30, 2019/01/05, 2019/04/04, 2019/09/20, spanning the 4 seasons) of measurements from TROPOMI. The SO_2 SCDs are based on the TROPOMI DOAS RPRO 3.2 data. For the AMFs, the calculation was made using TM5 a priori profiles and different cloud products as input:

- FRESCO: cloud fraction and cloud pressure are retrieved from O_2A NIR band measurements (cloud albedo is fixed to 0.8). The cloud parameters are obtained from the NO2 RPRO 3.2 product.

- O2O2: cloud fraction and cloud pressure are retrieved from O_2-O_2 dimer visible band measurements (cloud albedo is fixed to 0.8). The cloud parameters are obtained from the NO2 RPRO 3.2 product.
- OCRA/ROCINN: cloud fraction, cloud albedo and cloud pressure are retrieved using a combination of UV, vis and NIR spectral regions. The cloud parameters are taken from the CLD CRB RPRO 3.2 product.
- OCRA/ROCINN scaled: the settings are the same, except that the cloud albedo CA is set to 0.8 (as in the other cloud products) and an effective cloud fraction is calculated by a scaling of the original parameters, i.e., CF_{eff}=CFxCA/0.8.

In addition to the different cloud corrections listed above, results without correcting for the impact of clouds (i.e., using clear-sky AMFs) have also been considered. Figure 4.8 summarizes the SO_2 VCDs obtained with the different AMF settings, for the 4 days considered and the 4 selected regions already used in the previous sections.

From Figure 4.8 one can conclude that the change of cloud product, as input of the retrieval algorithm, has a rather small impact on the SO_2 VCDs. The differences are generally less than 20% in absolute values, and no clear pattern is observed. Moreover, switching off the cloud correction appears to have a modest impact on the results. However, it should be noted that the differences due to clouds can be very large for individual pixels. Fortunately, much of these differences average out on monthly scales over the relatively large regions used here.

A second comparison exercise was conducted on OMI data, using cloud parameters retrieved from the O2-O2 (OMCLDO2.003) and the Rotational Raman Scattering (RRS, OMCLDRR.003) cloud products. Similar as for TROPOMI, SO_2 VCDs have been calculated (with or without cloud correction) and intercompared. For the sake of conciseness, the results are not shown in this report, but the general conclusion is the same as for TROPOMI in that the impact of cloud correction on the retrievals is relatively small (on average) in the range of 0-20%.

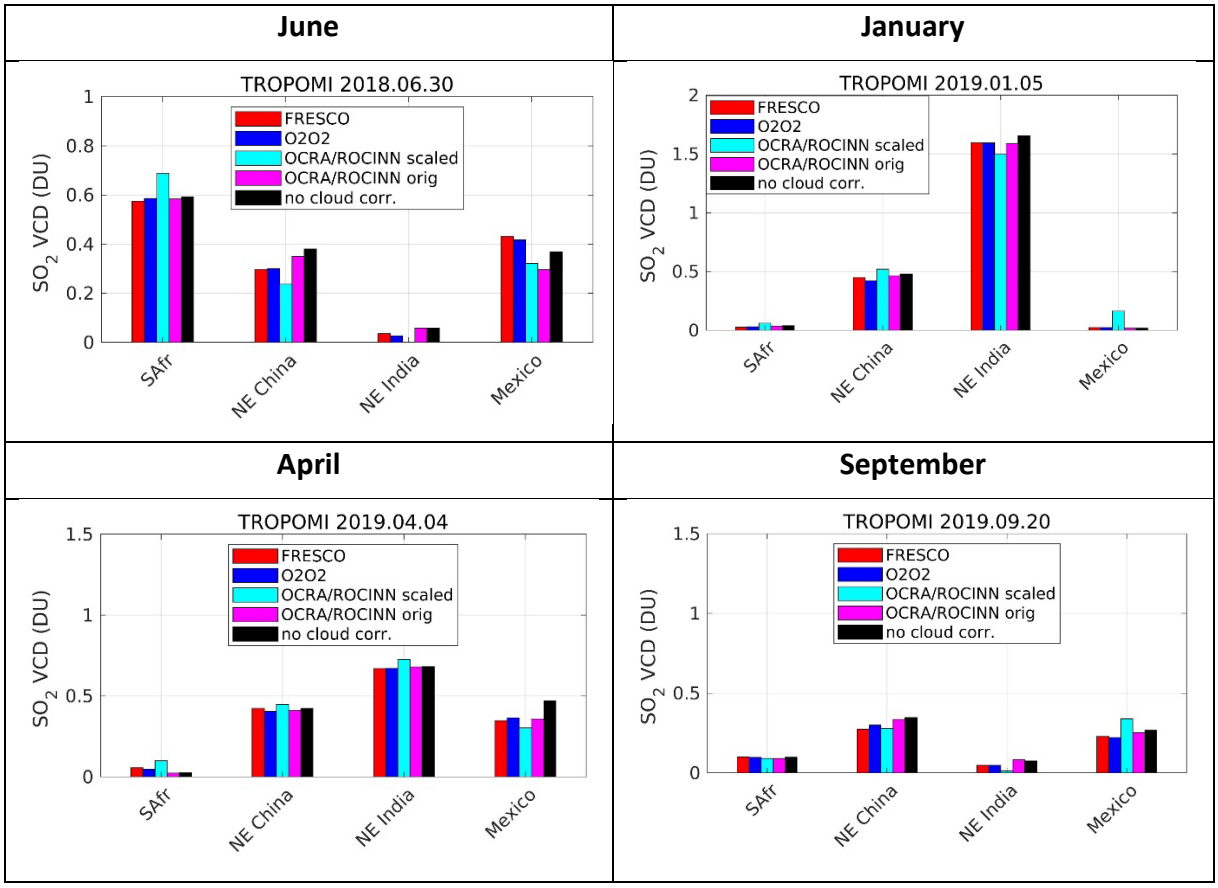


Figure 4.8: Comparison of TROPOMI daily averaged SO_2 vertical columns using cloud corrected AMFs based on different cloud products (see text) and and clear-sky AMFs, for 4 days and for 4 selected regions (Northeast India, Northeast China, South Africa, and Mexico). Pixel selection criteria: SZA less than 65°, cloud fraction (OCRA/ROCINN CRB) lower than 30% and TROPOMI rows 26– 424.

A last comparison was performed to evaluate the consistency between OMI and TROPOMI SO_2 retrievals and whether applying a cloud correction would have a positive or negative impact on the agreement. For this, we focused on the 4 regions used before, for 4 months in 2020. The results are shown in Figure 4.9. As can be seen, the SO_2 VCDs from OMI and TROPOMI agree generally very well and there is no clear conclusion that can be drawn on whether applying a cloud correction would help to make OMI and TROPOMI data more consistent or not.

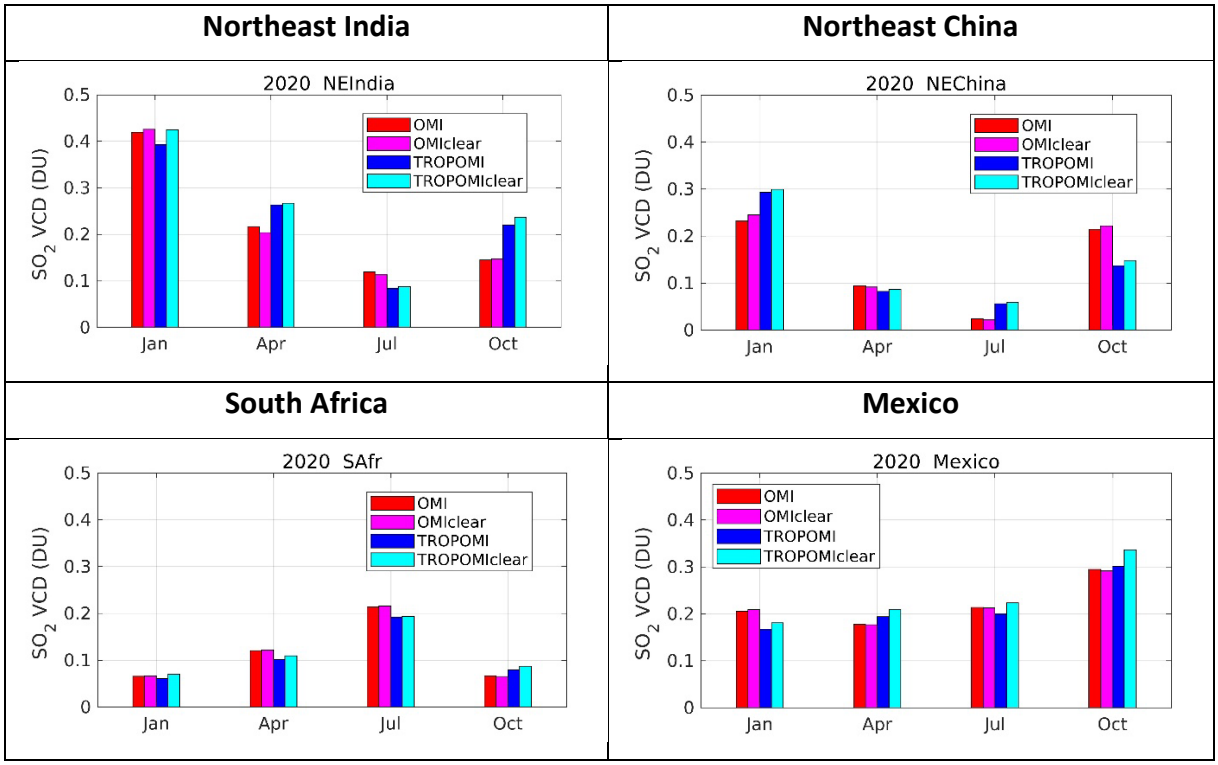


Figure 4.9: Comparison of OMI and TROPOMI monthly averaged SO_2 vertical columns, applying cloud correction (OMI: OMCLDO2, TROPOMI: OCRA/ROCINN scaled) or not ('clear') in the AMF calculation. The VCD results are presented for 4 months in 2020 and for 4 selected regions (Northeast India, Northeast China, South Africa, and Mexico).

4.4. Validation

A direct comparison of 4 OMI COBRA SO_2 variants (with prior CAMS and TM5, with and without cloud correction) was performed with MAX-DOAS measurements from Xianghe. In addition, for the cloud-corrected products, also a comparison after vertical harmonization is applied, where the satellite prior profile is replaced with the MAX-DOAS retrieved profile. Details on the validation procedure are summarised in the Annex A4. The results are presented in Figure 4.10.

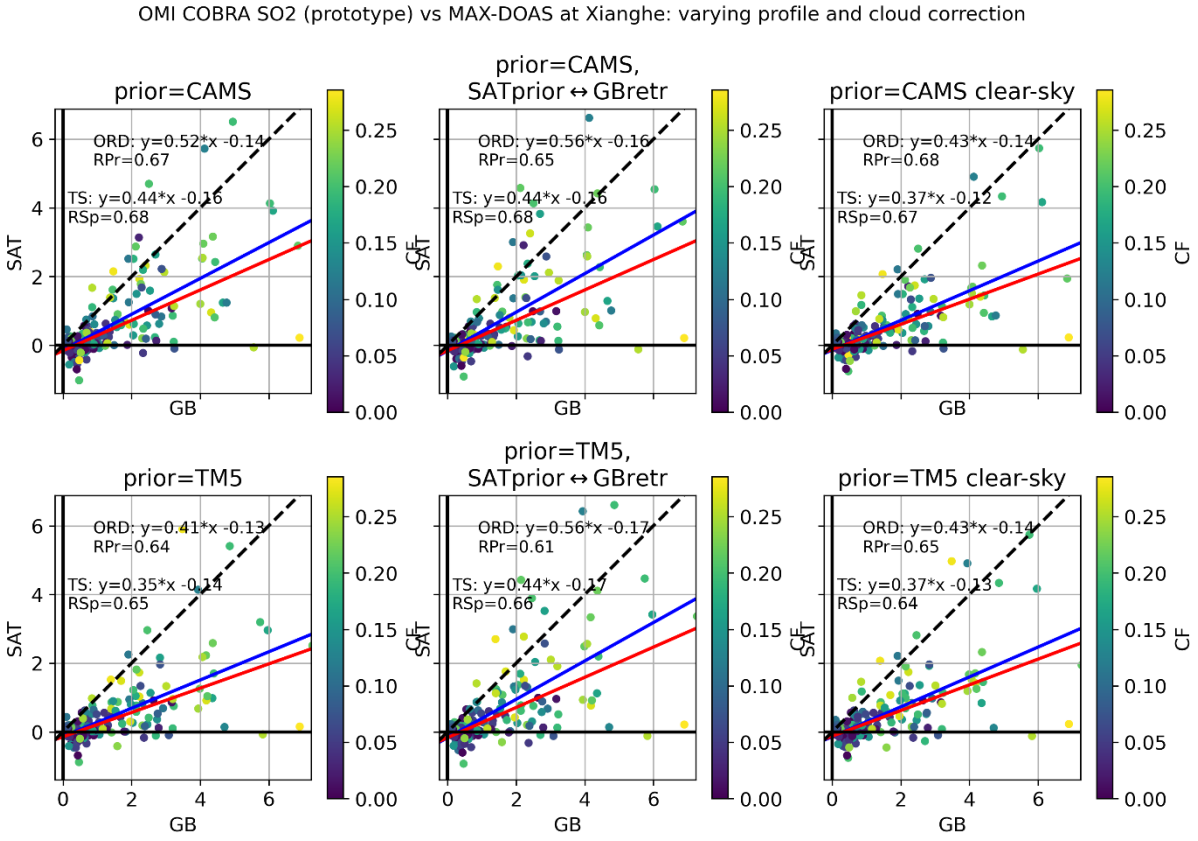


Figure 4.10: Comparison of OMI COBRA SO_2 with MAX-DOAS. Top: product with CAMS reanalysis as prior. Bottom: product with TM5 reanalysis as prior. Left: direct comparison. Middle: after replacing satellite prior with MAX-DOAS retrieved profile. Right: direct comparison, using the clear-sky product. Ordinary (ORD, blue) and robust Theil-Sen regressions (TS, red), Pearson (RPr) and Spearman (RSp) correlation coefficients are indicated. Points are coloured by satellite cloud fraction.

The conclusions are as follows:

- Substituting CAMS prior by the MAX-DOAS profile has a small impact on the comparison, indicating that the CAMS prior has already an adequate shape. The TM5 prior seems less adequate, as it leads to a larger proportional bias.
- Using clear-sky AMFs (irrespective if the prior is TM5 or CAMS) leads to a larger proportional bias, comparable to the TM5 cloud-corrected result.

The product with the lowest bias is therefore the cloud-corrected product using CAMS as prior. The bias seen for this product is likely not dominated by the prior profile assumption.

In addition, MAX-DOAS profile shapes have been compared with TM5 and CAMS profile shapes, with and without smoothing using the MAX-DOAS averaging kernel. The smallest median root-mean-square deviation was found for the CAMS profile, in agreement with the results shown above.

4.5. Algorithm Selection

We summarize here the main conclusions from the Round Robin study. The details of the selected retrieval settings can be found in the ATBD of the CCI+ precursors project.

From the sensitivity tests that have been performed, the impact on the results can be summarized (Table 4-1). The associated uncertainties can be estimated, as well as their random or systematic behaviour in time. By definition, an uncertainty is considered as systematic if its impact is not significantly reduced by averaging the data in space or time. For slant column retrievals, a dominant source of error is the noise on the spectra and its propagation on the results. However, averaging the observations enable to greatly reduce the standard deviation of the SO_2 SCDs. Other contributions to the slant column uncertainty are mostly systematic and relate to errors on the forward model and to uncertainties on the absorption cross-sections used. The impact of the LER, a priori profile and cloud correction selection shows clear spatial patterns in the averaged columns, as shown in Section 4.3. Those sources of uncertainty are therefore classified as mostly systematic, except for clouds which also show a random behaviour.

Table 4-1: Summary of the impact on the SO_2 retrievals of the Round Robin sensitivity tests.

Test	Description	Sensitivity	Behaviour in time
SCD	COBRA vs PCA	10-25%	Mostly systematic
LER	TROPOMI vs OMI minLER	5-10%	Mostly systematic
Profiles	TM5 vs CAMSrea	0-100%	Mostly systematic
Clouds	Different cloud products or no cloud correction	0-20%	Random and systematic

The SO_2 slant column retrieval will be achieved using the COBRA scheme. As anticipated at the start of the Round Robin and from earlier work on TROPOMI (see section 3 of Theys et al., 2021), the results shown in section 4.2 confirm that applying COBRA leads to a general reduction of the data scatter and local offsets, when compared to other methods. This is particularly important for studying weak emission sources. The good agreement found between COBRA and PCA for OMI is also a valuable argument to choose COBRA as it will ensure that the developed SO_2 CCI+ precursors CDR would be compatible with the widely used NASA PCA SO_2 data record. As COBRA is by essence a self-calibrating method, it is suitable for generating stable, long-term data sets. In that sense, the comparison between OMI and TROPOMI SO_2 COBRA results is encouraging and suggests that a good inter-satellite consistency will be reached. However, the SCD comparison in section 4.2 also revealed that COBRA SCDs are generally lower than DOAS results, under some conditions. At this stage, it is not entirely clear why this is so, but we cannot exclude that COBRA (and PCA) underestimate the true atmospheric SO_2 column amount.

For the AMF calculation:

- LER: the impact on the retrieved SO_2 VCDs is quite small (**Error! Reference source not found.**). However, the TROPOMI minimum LER is selected because of its improved spatial resolution. The same climatology is chosen for all sensors. The climatology has a spatial resolution of 0.125° x 0.125° and, for OMI, it will be used 'as is'. For low resolution sensors as GOME and SCIAMACHY, the TROPOMI climatology will be degraded to the appropriate resolutions.
- Profiles: among all sensitivity tests, the assumption on the vertical distribution of SO_2 has, by far, the largest impact on the retrievals. Switching from TM5 to CAMS can lead to an increase of the SO_2 VCDs up to a factor of 2 (e.g., in India). The reason for this is not completely understood. It could be due to the use of different emission inventories by the models, the difference in spatial resolution or different model settings for the transport or chemistry. The difficulty is also that little data is available to evaluate the SO_2 profiles from the models on one hand, and the satellite SO_2 vertical columns on the other hand.

  The CAMS reanalysis model is selected for the a priori profiles for the following reasons:
  - OMI validation results at Xianghe indicate that CAMS profiles lead to the lowest bias with the respect to the MAX-DOAS data.
  - The same behaviour (TM5 vs CAMS) is observed for HCHO. Validation results of HCHO columns are expected to be improved thanks to an increase of the columns over NMVOC emission regions. Thus, it is reasonable to believe that using CAMS SO_2 profile will be beneficial to the accuracy of the satellite SO_2 column retrievals.
  - The reanalysis dataset uses a constant version from 2003 to 2022.
  - CAMS provides an operational service that could be used for future extension.

- Clouds: the overall impact of the cloud correction and the use of different cloud products is rather limited (at least on monthly averaged data). The main SO_2 VCDs and averaging kernels will be provided with cloud correction applied. One motivation for this choice is that, for volcanic plumes well above the clouds, the cloud correction works well (and is less uncertain than for SO_2 in the boundary layer). Note that SO_2 clear-sky VCDs (i.e., with no cloud correction applied) will also be stored in the output files and made available to the users.

5. Summary and Conclusions

In this document, the results of the Round Robin exercise, performed for the three products NO_2, HCHO, and SO_2 as part of the Precursors CCI+ project are summarised. For each of the three trace gases, various options for the slant column retrieval were evaluated, algorithms for stratospheric and background correction compared, sensitivities of the air mass factor calculations studied and limited validation on different variants of the products performed.

For several of the critical inputs, good consistency was found between different choices, while for others, the choice made has important impacts on the product. Due to the limited scope of the activity, only some of the aspects could be evaluated in depth.

For NO_2, the main points can be summarised as

- NO_2 slant columns converge with the most recent versions of the processors, and for GOME-2, the latest lv1 data in combination with the settings of the AC-SAF prototype algorithm yield the most consistent results
- The stratospheric correction schemes STREAM and assimilation (QA4ECV) yield comparable results, in particular over polluted regions.
- Different LER inputs can lead to large differences in AMFs over snow and ice covered regions, but have much smaller impacts over most other areas. The directionality of the LER should be taken into account, but the absolute magnitude of the effect as well as how to correct it in a consistent manner needs to be further evaluated.
- In spite of large differences in the a priori columns from the different models evaluated (TM5, CAMS, CAMS reanalysis), the variations in tropospheric NO_2 columns retrieved using the various a priori data are small.
- Cloud properties retrieved with the different cloud algorithms (FRESCO, OCRA/ROCCIN) are converging in the latest algorithm versions, and now lead to relatively small differences in the tropospheric NO_2 columns computed using different cloud algorithms.

For HCHO, the results can be summarised as

- The QA4ACV fitting settings and background correction were found to perform best and will be used with adaptations.
- LER choice has a large impact on the results, and consistency between data sets is not good. Further work is needed to improve on this.
- The choice of HCHO a priori profiles (TM5, CAMS reanalysis, MAGRITTE) can lead to significant changes in the HCHO columns. The choice for CAMS reanalysis was made both because of expected better agreement with validation data and for practical reasons (availability).
- Cloud correction will not be applied as it creates inconsistencies between data from different sensors. This point should be revisited as new cloud products become available.

The main findings for SO_2 are

- The COBRA approach is superior to the classical DOAS analysis and comparable if not better than the PCA approach. There is indication for a low bias which should be evaluated further.
- The effect of using different LER climatologies is relatively small, and using TROPOMI data provides the best spatial resolution.
- The impact of the different a priori data for SO_2 (TM5, CAMS reanalysis) are large in some regions, and the lack of validation data limits the ability to decide, which of them is the best. CAMS reanalysis was chosen mainly for consistency with the other products and because of availability.
- Cloud correction has limited impact on non-volcanic SO_2, and differences between products based on different cloud products are relatively small. In contrast to the other products, volcanic SO_2 is often evaluated in the presence of thick ash clouds, where cloud correction is strictly necessary. Therefore, cloud correction is applied.

Based on the results of the comparisons, choices were made for the algorithms, input data and settings which are further detailed in the ATBDs of the individual products.

While the Round Robin exercise represents a large effort by the project consortium, only a small fraction of all algorithm aspects could be evaluated thoroughly in the time available. It is therefore to be expected that future improvements in input data and algorithms will improve the quality and consistency of future versions of the products.

6. References

Argyrouli, A.; Lutz, R.; Romahn, F.; Molina García, V.; Loyola, D.; Seo, S.; Valks, P.; De Smedt, I.; Boersma, F.; Tilstra, L.G.; et al. Comparison of Cloud Parameters from GOME-2 and Assessment of Cloud Impact on Tropospheric NO 2 and HCHO Retrievals. Environ. Sci. Proc., 26, 203.https://doi.org/10.3390/environsciproc2023026203, 2023.

Beirle, S., Hörmann, C., Jöckel, P., Liu, S., Penning de Vries, M., Pozzer, A., Sihler, H., Valks, P., and Wagner, T.: The STRatospheric Estimation Algorithm from Mainz (STREAM): estimating stratospheric NO2 from nadir-viewing satellites by weighted convolution, Atmos. Meas. Tech., 9, 2753–2779, https://doi.org/10.5194/amt-9-2753-2016, 2016

Boersma, K. F., Eskes, H. J., Veefkind, J. P., Brinksma, E. J., van der A, R. J., Sneep, M., van den Oord, G. H. J., Levelt, P. F., Stammes, P., Gleason, J. F., and Bucsela, E. J.: Near-real time retrieval of tropospheric NO 2 from OMI, Atmos. Chem. Phys., 7, 2103–2118, https://doi.org/10.5194/acp-7-2103-2007, 2007.

Chance, K. and Kurucz, R. L. L.: An improved high-resolution solar reference spectrum for earth's atmosphere measurements in the ultraviolet, visible, and near infrared, J. Quant. Spectrosc. Radiat. Transf., 111(9), 1289–1295, doi:10.1016/j.jqsrt.2010.01.036, 2010.

Compernolle et al., 2020. 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.

EUMETSAT: GOME-2 Products Guide, Tech. rep., Ref.: EUM/OPSEPS/MAN/07/0445, Issue: v2D, 2009.

EUMETSAT: GOME-2/Metop-A Level 1B Product Validation Report No. 5: Status at Reprocessing, Tech. rep., EUM/OPS/EPS/REP/09/0619, v1F, 2012

EUMETSAT, AC SAF Validation Report of Reprocessed L2 Total HCHO, GOME-2/MetopA&MetopB, doc ref: SAF/AC/IASB/VR/HCHO, October 2023. Draft.

Flemming, J., Huijnen, V., Arteta, J., Bechtold, P., Beljaars, A., Blechschmidt, A.-M., Diamantakis, M., Engelen, R. J., Gaudel, A., Inness, A., Jones, L., Josse, B., Katragkou, E., Marecal, V., Peuch, V.-H., Richter, A., Schultz, M. G., Stein, O., and Tsikerdekis, A.: Tropospheric chemistry in the Integrated Forecasting System of ECMWF, Geosci. Model Dev., 8, 975–1003, https://doi.org/10.5194/gmd-8-975-2015, 2015

Huijnen, V., Williams, J., van Weele, M., van Noije, T., Krol, M., Dentener, F., Segers, A., Houweling, S., Peters, W., de Laat, J., Boersma, F., Bergamaschi, P., van Velthoven, P., Le Sager, P., Eskes, H., Alkemade, F., Scheele, R., Nédélec, P., and Pätz, H.-W.: The global chemistry transport model TM5: description and evaluation of the tropospheric chemistry version 3.0, Geosci. Model Dev., 3, 445–473, https://doi.org/10.5194/gmd-3-445-2010, 2010.

Inness, A., Ades, M., Agustí-Panareda, A., Barré, J., Benedictow, A., Blechschmidt, A.-M., Dominguez, J. J., Engelen, R., Eskes, H., Flemming, J., Huijnen, V., Jones, L., Kipling, Z., Massart, S., Parrington, M., Peuch, V.-H., Razinger, M., Remy, S., Schulz, M., and Suttie, M.: The CAMS reanalysis of atmospheric composition, Atmos. Chem. Phys., 19, 3515–3556, https://doi.org/10.5194/acp-19-3515-2019, 2019.

Kleipool, Q.L., M.R. Dobber, J.F. de Haan, and P.F. Levelt, Earth surface reflectance climatology from 3 years of OMI data, J. Geophys. Res. 113, D18308, doi:10.1029/2008JD010290, 2008.

Li, Can, Nickolay Krotkov, Peter J. T. Leonard, Simon Carn, Joanna Joiner, Robert J. D. Spurr, and Alexander Vasilkov, Version 2 Ozone Monitoring Instrument SO2 Product (OMSO2 V2): New Anthropogenic SO2 Vertical Column Density Dataset, Atmos. Meas. Tech., 13, 61750-6191 https://doi.org/10.5194/amt-13-6175-2020, 19 Nov 2020.

Liu, S., Valks, P., Pinardi, G., De Smedt, I., Yu, H., Beirle, S., and Richter, A.: An improved total and tropospheric NO2 column retrieval for GOME-2, Atmos. Meas. Tech., 12, 1029–1057, https://doi.org/10.5194/amt-12-1029-2019, 2019.

Loyola, D. Automatic Cloud Analysis from Polar-Orbiting Satellites using Neural Network and Data Fusion Techniques. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Anchorage, AK, USA, 20–24 September 2004; Volume 4, pp. 2530–2534.

Loyola, D. A semi-transparent Lambertian cloud model for ozone retrieval. In DLR Presentation; Wessling, Germany, 2007.

Lutz, R.; Loyola, D.; Gimeno García, S.; Romahn, F. OCRA radiometric cloud fractions for GOME-2 on MetOp-A/B. Atmos. Meas. Tech., 9, 2357–2379, 2016.

Müller, J.-F., Stavrakou, T., and Peeters, J.: Chemistry and deposition in the Model of Atmospheric composition at Global and Regional scales using Inversion Techniques for Trace gas Emissions (MAGRITTE v1.1) – Part 1: Chemical mechanism, Geosci. Model Dev., 12, 2307–2356, https://doi.org/10.5194/gmd-12-2307-2019, 2019.

Munro, R., Lang, R., Klaes, D., Poli, G., Retscher, C., Lindstrot, R., Huckle, R., Lacan, A., Grzegorski, M., Holdak, A., Kokhanovsky, A., Livschitz, J., and Eisinger, M.: The GOME-2 instrument on the Metop series of satellites: instrument design, calibration, and level 1 data processing – an overview, Atmos. Meas. Tech., 9, 1279–1301, https://doi.org/10.5194/amt-9-1279-2016, 2016.

Pommereau, J. and Goutail, F.: O3 and NO2 ground-based measurements by visible spectrometry during Arctic winter and spring 1988, Geophys. Res. Lett., 15, 891–894, https://doi.org/10.1029/GL015i008p00891, 1988.

Popp, C.; Wang, P.; Brunner, D.; Stammes, P.; Zhou, Y.; Grzegorski, M. MERIS albedo climatology for FRESCO+ O2 A-band cloud retrieval. Atmos. Meas. Tech., 4, 463–483, 2011.

Seo, S., Valks, P., Boersma, K. F., De Smedt, I., Pinardi, G., Liu, S., Richter, A., and Loyola, D.: Development and evaluation of a long-term NO_2 data record from GOME-2 observations, in preparation, 2024.

Richter, A., Begoin, M., Hilboll, A., and Burrows, J. P.: An improved NO2 retrieval for the GOME-2 satellite instrument, Atmos. Meas. Tech., 4, 1147-1159, doi:10.5194/amt-4-1147-2011, 2011.

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

Theys, N., I. De Smedt, J. van Gent, T. Danckaert, T. Wang, F. Hendrick, T. Stavrakou, S. Bauduin, L. Clarisse, C. Li, N. A. Krotkov, H. Yu, M. Van Roozendael, 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., 120, doi:10.1002/2014JD022657, 2015.

Theys, N., I. De Smedt, H. Yu, T. Danckaert, J. van Gent, C. Hörmann, T. Wagner, P. Hedelt, H. Bauer, F. Romahn, M. Pedergnana, D. Loyola, M. Van Roozendael: Sulfur dioxide operational retrievals from TROPOMI onboard Sentinel-5 Precursor: Algorithm Theoretical Basis, Atmos. Meas. Tech., 10, 119-153, doi:10.5194/amt-10-119-2017, 2017.

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., and Van Roozendael, M.: A sulfur dioxide Covariance-Based Retrieval Algorithm (COBRA): application to TROPOMI reveals new emission sources, Atmos. Chem. Phys., 21, 16727–16744, https://doi.org/10.5194/acp-21-16727-2021, 2021.

Tilstra, L.G., O.N.E. Tuinder, P. Wang, and P. Stammes, Surface reflectivity climatologies from UV to NIR determined from Earth observations by GOME-2 and SCIAMACHY, J. Geophys. Res. Atmos. 122, 4084-4111, doi:10.1002/2016JD025940, 2017.

Tilstra, L.G., M. de Graaf, V. Trees, P. Litvinov, O. Dubovik, and P. Stammes, A directional surface reflectance climatology determined from TROPOMI observations, Atmos. Meas. Tech. Discuss. [preprint], doi:10.5194/amt-2023-222, in review, 2023.

Valks, P. ACSAF ATBD GOME-2 Total Column Products of Ozone, NO2, BrO, HCHO, SO2, H2O, OClO and Cloud Properties, Tech. Rep., Issue 3/B, 11 November 2019, German Aerospace Centre (DLR), Microsoft Word-DLR_GOME-2_ATBD_3C_GDP49_Nov2019. Available online: acsaf.org (accessed on 24 May 2023).

Valks, P., Pinardi, G., Richter, A., Lambert, J.-C., Hao, N., Loyola, D., Van Roozendael, M., and Emmadi, S.: Operational total and tropospheric NO 2 column retrieval for GOME-2, Atmos. Meas. Tech., 4, 1491–1514, https://doi.org/10.5194/amt-4-1491-2011, 2011.

Van Geffen, J., Eskes, H., Compernolle, S., Pinardi, G., Verhoelst, T., Lambert, J. C., Sneep, M., Linden, M. Ter, Ludewig, A., Folkert Boersma, K. and Pepijn Veefkind, J.: Sentinel-5P TROPOMI NO2retrieval: impact of version v2.2 improvements and comparisons with OMI and ground-based data, Atmos. Meas. Tech., 15(7), 2037–2060, doi:10.5194/amt-15-2037-2022, 2022.

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., Folkert Boersma, K., 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., Cárdenas, C. R., De Miguel, L. S., Sinyakov, V. P., Stremme, W., Strong, K., Van Roozendael, M., Pepijn Veefkind, J., Wagner, T., Wittrock, F., Yela González, M. and Zehner, C.: Ground-based validation of the Copernicus Sentinel-5P TROPOMI NO2 measurements with the NDACC ZSL-DOAS, MAX-DOAS and Pandonia global networks, Atmos. Meas. Tech., 14(1), 481–510, doi:10.5194/amt-14-481-2021, 2021.

Wang, P., Stammes, P., van der A, R., Pinardi, G., and van Roozendael, M.: FRESCO+: an improved O2 A-band cloud retrieval algorithm for tropospheric trace gas retrievals, Atmos. Chem. Phys., 8, 6565–6576, https://doi.org/10.5194/acp-8-6565-2008, 2008.

Williams, J. E., Boersma, K. F., Le Sager, P. and Verstraeten, W. W.: The high-resolution version of TM5-MP for optimized satellite retrievals: description and validation, Geosci. Model Dev., 10(2), 721–750, doi:10.5194/gmd-10-721-2017, 2017.

Zara, M., Boersma, K. F., De Smedt, I., Richter, A., Peters, E., van Geffen, J. H. G. M., Beirle, S., Wagner, T., Van Roozendael, M., Marchenko, S., Lamsal, L. N., and Eskes, H. J.: Improved slant column density retrieval of nitrogen dioxide and formaldehyde for OMI and GOME-2A from QA4ECV: intercomparison, uncertainty characterisation, and trends, Atmos. Meas. Tech., 11, 4033–4058, https://doi.org/10.5194/amt-11-4033-2018, 2018.

Annex: Validation methodology

A1 Stratospheric NO2 QA4ECV & AC SAF prototype validation methodology

The validation of the stratospheric NO_2 products follows the methodology already published in Compernolle et al., 2020 and Verhoelst et al., 2021, which builds upon previous developments in the AC SAF, H2020 QA4ECV, and the S5P MPC VDAF.

A1.1 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°). The QA4ECV data were additionally filtered on the snow-ice flag for some of the analyses. No filtering was done on the ZSL-DOAS data.

A1.2 Co-location

- 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 are preferentially based on the sunrise ZSL-DOAS data, and vice versa for the afternoon sounders.

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

A1.3 Harmonization

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

A1.4 Analysis

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

A2 GOME-2A AC SAF HCHO validation methodology

The validation methodology of GOME-2A AC SAF HCHO vs FTIR and MAX-DOAS follows the common AC SAF validation methodology. This is reproduced below from (EUMETSAT, 2023).

Satellite HCHO VCD daily means are calculated using all pixels falling within a radius of 150 km around the stations and meeting the following selection criteria:

- HCHO slant column densities (SCDs) larger than -4x10^{16} molec/cm^2,
- solar zenith angle (SZA) lower than 70°,
- and cloud fraction lower than 40%.

For the ground-based data, daily means are performed for FTIR (due to their sometimes limited number of direct-sun solar measurements in one day), while MAXDOAS data are averaged for data within +/- 1h of the satellite overpass time.

In order to allow direct comparison between GOME-2 and ground-based observations, the difference in vertical sensitivity between both measurement types is taken into account by applying the satellite column averaging kernels to the ground-based HCHO profiles when available (i.e. for the FTIR dataset). Smoothed ground-based HCHO VCDs ($VCD_{GB,smoothed}$) are derived for each day by averaging retrieved ground-based profiles falling within the time-selection and convolving the mean profile ($\mathbf{x}_{GB}$) with the corresponding satellite column averaging kernel (AK_{sat}):

$$VCD_{GB,smoothed} = AK_{sat} * \mathbf{x}_{GB}$$

If the first altitude level of the satellite column averaging kernel is above the altitude of the station, then the averaging kernel is extrapolated down to the altitude of the station.

A3 GOME-2A QA4ECV HCHO validation methodology

The validation methodology of GOME-2A QA4ECV HCHO vs MAX-DOAS is as follows:

Satellite pixel filtering. Pixels are kept if `processing_error_flag`==0, `rms_fit`<1.5e-3, SZA<70°, `surface_albedo_HCHO`<0.3, SCD>-4e16 molec/cm2, cloud fraction<0.5, `snow_ice_flag` not in [100 (sea-ice covered),101 (permanent ice),103 (snow)].

Co-location and comparison pair handling. Pixels must be within 100 km radius of sensor. Ground-based data must be within ±1 h of satellite pixel. All pixels co-located with the same ground-based measurement are averaged.

A4 SO_2 validation methodology

Satellite pixel filtering. Pixels kept if AMF>0.15, 10 DU>VCD>-3.5 DU, SZA<65, CF<0.3, xtrack_flag equal to zero.

Co-location. Keep satellite pixels within 50 km of MAX-DOAS sensor location. Keep ground-based data within ±1 h of satellite pixel.

Vertical harmonization.
1/ Direct comparison: No action
2/ Profile harmonization: SAT VCD is recalculated by using the GB prior: Mass-conserved re-gridding of MAX-DOAS profile to satellite pressure grid, and extended with satellite prior, and then multiplied with satellite averaging kernel to obtain a smoothed column. Finally, the satellite VCD is obtained as (SAT VCD)*(GB VCD)/(GB smoothed VCD).

Comparison pair handling
All MAX-DOAS measurements co-located with the same pixel are averaged.
Daily averages of MAX-DOAS and satellite data are taken.