Atmospheric
Measurement
Techniques

TROPOMI–Sentinel-5 Precursor formaldehyde validation using an extensive network of ground-based Fourier-transform infrared stations

Corinne Vigouroux[1], Bavo Langerock[1], Carlos Augusto Bauer Aquino[2], Thomas Blumenstock[3], Zhibin Cheng[4], Martine De Mazière[1], Isabelle De Smedt[1], Michel Grutter[5], James W. Hannigan[6], Nicholas Jones[7], Rigel Kivi[8], Diego Loyola[4], Erik Lutsch[9], Emmanuel Mahieu[10], Maria Makarova[11], Jean-Marc Metzger[12], Isamu Morino[13], Isao Murata[14], Tomoo Nagahama[15], Justus Notholt[16], Ivan Ortega[6], Mathias Palm[16], Gaia Pinardi[1], Amelie Röhling[3], Dan Smale[17], Wolfgang Stremme[5], Kim Strong[9], Ralf Sussmann[18], Yao Té[19], Michel van Roozendael[1], Pucai Wang[20], and Holger Winkler[16]

[1]Royal Belgian Institute for Space Aeronomy (BIRA-IASB), Brussels, Belgium
[2]Instituto Federal de Educaçao, Ciência e Tecnologia de Rondônia (IFRO), Porto Velho, Brazil
[3]Karlsruhe Institute of Technology (KIT), Institute for Meteorology and Climate Research (IMK-ASF), Karlsruhe, Germany
[4]German Aerospace Centre (DLR), Remote Sensing Technology Institute, Oberpfaffenhofen, 82234 Weßling, Germany
[5]Centro de Ciencias de la Atmósfera, Universidad Nacional Autónoma de México (UNAM), Mexico City, México
[6]Atmospheric Chemistry, Observations & Modeling, National Center for Atmospheric Research (NCAR), Boulder, CO, USA
[7]Centre for Atmospheric Chemistry, University of Wollongong, Wollongong, Australia
[8]Finnish Meteorological Institute (FMI), Sodankylä, Finland
[9]Department of Physics, University of Toronto, Toronto, Canada
[10]Institut d'Astrophysique et de Géophysique, Université de Liège, Liège, Belgium
[11]Saint Petersburg State University, Atmospheric Physics Department, St. Petersburg, Russia
[12]Observatoire des Sciences de l'Univers Réunion (OSU-R), UMS 3365, Université de la Réunion, Saint-Denis, France
[13]National Institute for Environmental Studies (NIES), Tsukuba, Ibaraki 305-8506, Japan
[14]Graduate School of Environment Studies, Tohoku University, Sendai 980-8578, Japan
[15]Institute for Space-Earth Environmental Research (ISEE), Nagoya University, Nagoya, Japan
[16]Institute of Environmental Physics, University of Bremen, Bremen, Germany
[17]National Institute of Water and Atmospheric Research Ltd (NIWA), Lauder, New Zealand
[18]Karlsruhe Institute of Technology, IMK-IFU, Garmisch-Partenkirchen, Germany
[19]LERMA-IPSL, Sorbonne Université, CNRS, Observatoire de Paris, PSL Université, 75005 Paris, France
[20]Institute of Atmospheric Physics, Chinese Academy of Sciences (CAS), Beijing, China

Correspondence: Corinne Vigouroux (corinne.vigouroux@aeronomie.be)

Abstract. TROPOMI (the TROPOspheric Monitoring Instrument), on board the Sentinel-5 Precursor (S5P) satellite, has been monitoring the Earth's atmosphere since October 2017 with an unprecedented horizontal resolution (initially $7\,km^2 \times 3.5\,km^2$, upgraded to $5.5\,km^2 \times 3.5\,km^2$ in August 2019). Monitoring air quality is one of the main objectives of TROPOMI; it obtains measurements of important pollutants such as nitrogen dioxide, carbon monoxide, and formaldehyde (HCHO). In this paper we assess the quality of the latest HCHO TROPOMI products versions 1.1.(5-7), using ground-based solar-absorption FTIR (Fourier-transform infrared) measurements of HCHO from 25 stations around the world, including high-, mid-, and low-latitude sites. Most of these stations are part of the Network for the Detection of Atmospheric Composition Change (NDACC), and they provide a wide range of observation conditions, from very clean

remote sites to those with high HCHO levels from anthropogenic or biogenic emissions. The ground-based HCHO retrieval settings have been optimized and harmonized at all the stations, ensuring a consistent validation among the sites.

In this validation work, we first assess the accuracy of TROPOMI HCHO tropospheric columns using the median of the relative differences between TROPOMI and FTIR ground-based data (BIAS). The pre-launch accuracy requirements of TROPOMI HCHO are 40 %–80 %. We observe that these requirements are well reached, with the BIAS found below 80 % at all the sites and below 40 % at 20 of the 25 sites. The provided TROPOMI systematic uncertainties are well in agreement with the observed biases at most of the stations except for the highest-HCHO-level site, where it is found to be underestimated. We find that while the BIAS has no latitudinal dependence, it is dependent on the HCHO concentration levels: an overestimation ($+26 \pm 5$ %) of TROPOMI is observed for very low HCHO levels ($< 2.5 \times 10^{15}$ molec. cm^{-2}), while an underestimation (-30.8 % ± 1.4 %) is found for high HCHO levels ($> 8.0 \times 10^{15}$ molec. cm^{-2}). This demonstrates the great value of such a harmonized network covering a wide range of concentration levels, the sites with high HCHO concentrations being crucial for the determination of the satellite bias in the regions of emissions and the clean sites allowing a small TROPOMI offset to be determined. The wide range of sampled HCHO levels within the network allows the robust determination of the significant constant and proportional TROPOMI HCHO biases (TROPOMI = $+1.10 \pm 0.05$ $\times 10^{15} + 0.64 \pm 0.03 \times$ FTIR; in molecules per square centimetre).

Second, the precision of TROPOMI HCHO data is estimated by the median absolute deviation (MAD) of the relative differences between TROPOMI and FTIR ground-based data. The clean sites are especially useful for minimizing a possible additional collocation error. The precision requirement of 1.2×10^{16} molec. cm^{-2} for a single pixel is reached at most of the clean sites, where it is found that the TROPOMI precision can even be 2 times better (0.5–0.8×10^{15} molec. cm^{-2} for a single pixel). However, we find that the provided TROPOMI random uncertainties may be underestimated by a factor of 1.6 (for clean sites) to 2.3 (for high HCHO levels). The correlation is very good between TROPOMI and FTIR data ($R = 0.88$ for 3 h mean coincidences; $R = 0.91$ for monthly means coincidences). Using about 17 months of data (from May 2018 to September 2019), we show that the TROPOMI seasonal variability is in very good agreement at all of the FTIR sites.

The FTIR network demonstrates the very good quality of the TROPOMI HCHO products, which is well within the pre-launch requirements for both accuracy and precision. This paper makes suggestions for the refinement of the TROPOMI random uncertainty budget and TROPOMI quality assurance values for a better filtering of the remaining outliers.

1 Introduction

TROPOMI (the TROPOspheric Monitoring Instrument), on board the Sentinel-5 Precursor (S5P) satellite, has been monitoring the column amounts of atmospheric constituents since October 2017 at the unprecedented horizontal resolution of $7 \, \text{km}^2 \times 3.5 \, \text{km}^2$, which was upgraded to $5.5 \, \text{km}^2 \times 3.5 \, \text{km}^2$ in August 2019. This huge number of data, delivered to the public and the scientific community, represents a big step towards improving our knowledge of chemical and dynamical processes in the atmosphere. It is crucial to validate the quality of these new satellite data to trust and benefit from their scientific exploitation. This paper focuses on the first quality assessment of the latest publicly available TROPOMI HCHO data products (v.1.1.[5-7]).

In the past, the HCHO satellite products have been validated at a few locations only, mainly using aircraft in situ measurements (Martin et al., 2004; Zhu et al., 2016, 2020), MAX-DOAS (multi-axis differential optical absorption spectroscopy) measurements over land (Wittrock et al., 2006; De Smedt et al., 2015), or ship-based (Peters et al., 2012; Tan et al., 2018) and FTIR (Fourier-transform infrared) measurements (Jones et al., 2009; Vigouroux et al., 2009; De Smedt et al., 2015). However, given the high spatial heterogeneity of HCHO concentrations due to its short lifetime (a few hours), there is a crucial need for a more extended world coverage to unambiguously assess the satellites' achieved accuracy and precision. Furthermore, increasing the number of ground-based locations is not sufficient; it is also important to harmonize the reference data obtained at all the stations in order to facilitate the interpretation of the satellite validation by minimizing the site-to-site biases. In this view and in particular in the framework of the TROPOMI calibration and validation (Cal/Val) activities, we have developed HCHO retrieval settings that are suitable for any ground-based FTIR site and have been consistently applied in Vigouroux et al. (2018) at 21 FTIR stations, most of them affiliated with the NDACC (Network for the Detection of Atmospheric Composition Change). Vigouroux et al. (2018) described in detail the retrieval settings and the harmonized HCHO time series obtained at these stations, which cover a large range of HCHO concentrations, from very clean Arctic and oceanic sites to high-HCHO-level sites, such as polluted cities (e.g. Paris or Mexico City) and sites close to large biogenic emissions, like the Amazon basin (Porto Velho).

This paper presents the validation of the TROPOMI HCHO product (v.1.1.[5-7]) using an updated network of 25 ground-based FTIR stations. In the first section, the TROPOMI HCHO data are introduced with their uncertainty budget and their quality flag criteria. The second section describes the ground-based FTIR HCHO network and the characterization of these reference data (uncertainties and averaging kernels). Then, the validation procedure (collocation criteria, smoothing technique, definition of the quantities to be used in the quality assessment) is explained in Sect. 4. Fi-

nally, Sect. 5 shows the validation results using comparisons between TROPOMI and FTIR ground-based network data, leading to an assessment of the TROPOMI HCHO accuracy and precision and the observed TROPOMI bias.

2 TROPOMI HCHO data

TROPOMI, on the S5P platform, is in a low-Earth afternoon polar orbit with a swath of 2600 km, resulting in daily global coverage (Veefkind et al., 2012). Operational Level 2 (L2) products include vertical columns of O_3, SO_2, NO_2, HCHO, CO, and CH_4 as well as O_3 profile, aerosol layer height, cloud information and aerosol index. The spatial resolution of originally $3.5 \, km^2 \times 7 \, km^2$ was increased to $3.5 \, km^2 \times 5.5 \, km^2$ on 6 August 2019.

The prototype algorithm of the formaldehyde product is being developed at the Royal Belgian Institute for Space Aeronomy (BIRA-IASB), and the corresponding operational processor is being developed at the Remote Sensing Technology Institute (IMF) of the German Aerospace Center (DLR). The product has been declared operational and was released to the public at the end of 2018. At the time of writing this paper, the latest product versions 1.1.(5-7) provided a consistent time series of reprocessed + offline (RPRO + OFFL) data, covering the period between May 2018 and (at least) December 2019 (last access). The detailed validation results shown in Sect. 5 are obtained using this consistent time series (RPRO + OFFL, from 14 May 2018 to 31 December 2019). The version numbers and their dates of change are given in Table 1, and further details are given in the Readme file.[1] The near-real-time (NRTI) product for the same versions 1.1.(5-7) spanned from December 2018 up to December 2019 (last access). This product has also been validated, but because the results are very similar to the RPRO+OFFL validation, we do not show them in detail in this paper.

The S5P HCHO retrieval algorithm is based on the DOAS method and is directly inherited from the OMI QA4ECV product retrieval algorithm (https://doi.org/10.18758/71021031, De Smedt et al., 2017). It consists of a three-step method (slant column retrieval, air mass factor calculation, and conversion to tropospheric column), which is fully described in De Smedt et al. (2018). The retrieval of the slant columns (N_s) is performed in the UV part of the spectra (in TROPOMI channel 3) in a fitting interval of 328.5–359 nm. The HCHO cross section is from Meller and Moortgat (2000). Together with the HCHO cross section, the absorptions of NO_2, BrO, O_3 (at two temperatures), and O_4 are fitted. A Ring cross section and two pseudo-cross sections to account for non-linear O_3 absorption effects are also included in the fit. References are given in De Smedt et al. (2018).

All cross sections have been pre-convolved for every row separately, with an instrumental slit function adjusted just after launch. The DOAS reference spectrum is updated daily with an average of Earth radiances selected in the Equatorial Pacific region on the previous day. The result of the fit is therefore a differential slant column showing increases over continental sources compared to the remote background. The conversion from slant to tropospheric columns (N_v) is performed using a lookup table of vertically resolved air mass factors (M) calculated at 340 nm with the radiative transfer model VLIDORT v2.6 (Spurr, 2008). Parameters for each ground pixel are the observation geometry; the surface elevation and reflectivity, including the clouds (that are treated as reflecting surfaces); and a priori tropospheric profiles. The surface albedo is taken from the monthly OMI albedo climatology (minimum Lambertian equivalent reflectivity; Kleipool et al., 2008) at the spatial resolution of $1° \times 1°$. A priori vertical profiles are specified using the TM5-MP daily forecast at the same spatial resolution (Williams et al., 2017). Cloud properties are provided by the S5P operational product in its cloud as reflecting boundary (CRB) mode (Loyola et al., 2018). A cloud correction based on the independent pixel approximation (Boersma et al., 2004) is applied for cloud fractions larger than 0.1. In order to correct for any remaining global offset and stripes, a background correction is applied based on HCHO slant columns from the 5 previous days in the Pacific Ocean ($N_{(s,0)}$), as described in De Smedt et al. (2018). Finally, the background vertical column of HCHO due to methane oxidation is taken from the TM5 model in the reference region ($N_{(v,0)}^{CTM}$). The equation of the tropospheric HCHO vertical column can be written as follows:

$$N_v = \frac{(N_s - N_{(s,0)})}{M} + \frac{M_0}{M} \cdot N_{(v,0)}^{CTM}, \tag{1}$$

with M_0 being the average of the air mass factors M of the slant columns selected in the reference sector, the Pacific Ocean ($N_{(s,0)}$). Intermediate quantities and auxiliary data are all provided in the L2 files (http://www.tropomi.eu/sites/default/files/files/Sentinel-5P-Level-2-Product-User-Manual-Formaldehyde_v1.01.01_20180716.pdf, last access: 7 July 2020).

Several diagnostic variables are provided together with the measurements. Quality assurance (QA) values are defined to perform a quick selection of the observations. QA > 0.5 filters out most observations presenting an error flag or a solar zenith angle (SZA) larger than 70°, a cloud radiance fraction larger than 0.6 at 340 nm, or an air mass factor smaller than 0.1. The product Readme file reports that, in the current version, the QA values are not always correctly set over snow and ice regions or above an SZA of 75°. They also need to be further checked over cloudy scenes. In the forthcoming S5P version 2, QA values will be refined and will exclude data with a surface albedo larger than 0.2 and a snow or ice warning as well as remaining SZAs larger than 75°.

---

[1] http://www.tropomi.eu/sites/default/files/files/publicSentinel-5P-Formaldehyde-Readme_20191213.pdf (last access: 7 July 2020)

**Table 1.** TROPOMI RPRO+OFFL complete time series versions 1.1.(5-7) used in the present work.

Date	Processor version	Relevant improvements (see Readme file*).
2018-05-14 to 2018-11-28	RPRO v.1.1.5	Alignment of the configuration for NRTI, OFFL, and RPRO chains regarding the chemistry
2018-11-28 to 2019-03-28	OFFL v.1.1.5	transport model input, leading to the same product quality
2019-03-28 to 2019-04-23	OFFL v.1.1.6	– Surface classification climatology updated – Fixed a bug in the interpolation of the surface albedo climatology – Fixed a problem regarding the retrieved CLOUD product parameters being too close to the a priori values; this might have affected the calculation of the HCHO in cloudy cases
2019-04-23 to 2019-12-31	OFFL v.1.1.7	No changes (for HCHO) with respect to previous version

* http://www.tropomi.eu/sites/default/files/files/publicSentinel-5P-Formaldehyde-Readme_20191213.pdf (last access: 7 July 2020).
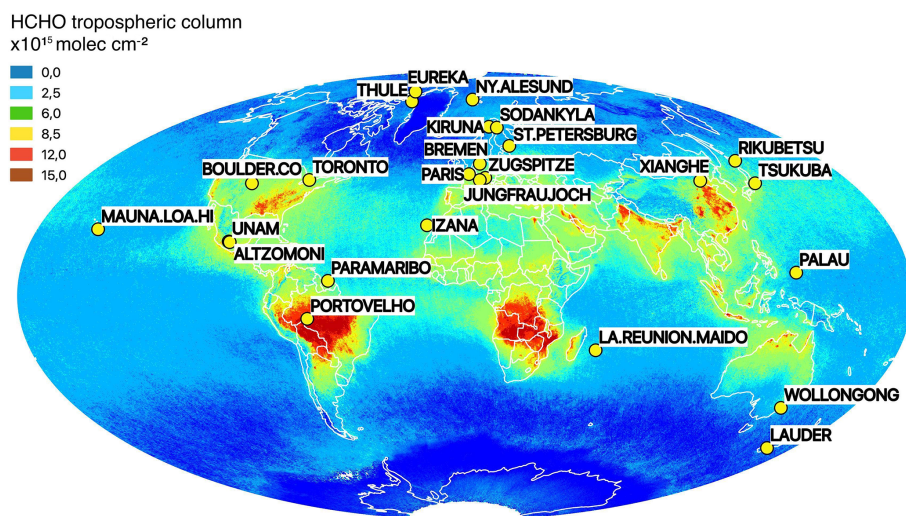
Figure 1. Network of ground-based FTIR stations providing HCHO total column data. The background is the September 2018 monthly mean of TROPOMI HCHO tropospheric columns, averaged on a $0.2° \times 0.2°$ grid using the HARP tool v.1.5 (https://atmospherictoolbox.org, last access: 7 July 2020).

The tropospheric column uncertainty is divided into random (precision) and systematic (accuracy) components and is provided per pixel. It varies with the observation conditions. Over remote regions at a moderate solar zenith angle, the precision of an individual observation is about 5×10^{15} molec. cm^{-2}. This value agrees with the standard deviation of the columns in the same region for a particular day. The random uncertainty is dominated by the random error in the slant columns. The tropospheric column accuracy is the combined systematic uncertainty resulting from the slant column, the air mass factor, and the background correction errors. It varies between 30 % and 60 % of the columns. The column averaging kernel and the a priori profiles are provided for every observation.

3 Ground-based FTIR HCHO data

We show in Fig. 1 a map of the ground-based FTIR stations used in this TROPOMI validation. The background image represents the global TROPOMI monthly mean tropospheric columns for September 2018, illustrating the different HCHO levels sampled by the ground-based network: from clean Arctic and oceanic sites to very-high-concentration sites such as Porto Velho, in the Amazon basin.

Table 2 lists the ground-based FTIR stations, their coordinates and altitude, the spectrometer type, the retrieval code, and the team involved in the measurements and/or the retrievals of HCHO. For more details on the monitoring of FTIR solar absorption spectra at these stations, we refer to Vigouroux et al. (2018) and references therein, and for the FTIR retrieval principles we refer to e.g. Vigouroux et al. (2009).

The same retrieval settings are used at all the stations to avoid introducing possible bias in the HCHO total columns between the stations and inconsistent comparisons with the satellite. Details are given in Vigouroux et al. (2018). The main settings that might be responsible for internal biases within the network are the spectroscopic database and the fitted spectral windows, the spectroscopic parameters being the main source of the FTIR HCHO systematic uncertainties. The HCHO spectral signatures lie in the 3.6 μm region and belong to the ν_1 and ν_5 bands (fitted windows are 2763.42–2764.17, 2765.65–2766.01, 2778.15–2779.1, and 2780.65–2782.0 cm^{-1}). The spectroscopic database used is the atm16 linelist by Geoffrey Toon (JPL), which can be found at http://mark4sun.jpl.nasa.gov/toon/linelist/linelist.html (last access: 7 July 2020). This linelist is optimized for the main absorbing gases in the fitted windows (HDO, CH_4, O_3, N_2O, CO_2) and is based on HITRAN 2012 (Rothman et al., 2013) for HCHO, which used the work of Jacquemart et al. (2010).

The retrieval codes used in the FTIR NDACC community are PROFITT9 (Hase et al., 2006) and SFIT4.0.9.4 (updated from SFIT2; Pougatchev et al., 1995), which are both based on the optimal estimation method (Rodgers, 2000). A past comparison exercise has shown a very good agreement between the retrieved products obtained with these two codes (Hase et al., 2004). Based on a priori profile information (from the Whole Atmosphere Community Climate Model, WACCM; Garcia et al., 2007) and an L1 Tikhonov regularization matrix (Tikhonov, 1963), low-vertical-resolution profiles can be retrieved in principle as well as total columns. However, as described in Vigouroux et al. (2018), the degrees of freedom for signal are very low for HCHO (median value of 1.1 for all FTIR sites), meaning that we essentially have one piece of information. The FTIR total column averaging kernel shows a decrease in the sensitivity at the surface, which is quite similar to the TROPOMI sensitivity. This can be seen in Fig. 2 as an example for the Maïdo station. We also show in Fig. 2 the FTIR a priori profile at Maïdo, which is based on a climatology (1980–2020) from the WACCM model calculated at Maïdo. A single profile is used for the whole time series at a specific station (Vigouroux et al., 2018), while TROPOMI uses daily a priori profiles from TM5 (Sect. 2). An example is shown in Fig. 2 for 18 January 2019.

The FTIR uncertainty budget is calculated following the formalism of Rodgers (2000) and is described in Vigouroux et al. (2018). It is separated into random and systematic components. The random uncertainty is dominated at all sites by the measurement noise uncertainty, which can vary from site to site depending on the spectrometer. The uncertainty in the retrieved FTIR total columns for individual sites is given in Vigouroux et al. (2018) for the 21 sites involved at that time. We obtain a median random uncertainty of 2.3×10^{14} molec. cm^{-2}, with a large value of 11.1×10^{14} molec. cm^{-2} only at Mexico City, where a lower-resolution instrument is used (Vertex 80). The smoothing

uncertainty in the total column has a non-negligible random component (median value of 1.2×10^{14} molec. cm^{-2}). With the inclusion of the smoothing error in the uncertainty budget, the median total random uncertainty is 2.9×10^{14} molec. cm^{-2}, which is very close to our empirical uncertainty estimation of 2.8×10^{14} molec. cm^{-2} based on the standard deviation of the differences between two individual subsequent FTIR measurements (within an interval of maximum 30 min), confirming our theoretical uncertainty calculation. Since the Vigouroux et al. (2018) paper, five more sites have joined the HCHO harmonized network. The mean random errors of individual FTIR measurements are 1.4×10^{14}, 2.7×10^{14}, 2.2×10^{14}, 5.2×10^{14}, and 5.6×10^{14} molec. cm^{-2} for Jungfraujoch, Rikubetsu, Tsukuba, Palau, and Xianghe, respectively.

The forward-model parameter median systematic uncertainty in the HCHO FTIR total columns is 13 % in the network described by Vigouroux et al. (2018). As already mentioned, the dominating systematic uncertainty sources are the spectroscopic parameters: the line intensities and the pressure-broadening coefficients of the fitted HCHO absorption lines. We use 10 % for the three parameters: the line intensity and the air- and self-broadening coefficients. The systematic uncertainty can be larger (up to 21 %–26 %) at the stations using the PROFFIT9 retrieval code due to an assumed uncertainty in the channeling that is not taken into account yet in the SFIT4 code. However, this channeling uncertainty can also be negligible at some sites (it depends on each instrument), and more investigation is needed at each station to avoid its underestimation or overestimation. The median smoothing systematic uncertainty is 3.4 %. For the five added sites, the median total systematic uncertainty is 13 % (Jungfraujoch, Tsukuba, Palau) or 14 % (Rikubetsu, Xianghe), commensurate with the other sites.

4 Validation method

4.1 Collocation criteria

The precision of a single pixel of the TROPOMI HCHO measurement is expected to be below 1.2×10^{16} molec. cm^{-2} (pre-launch requirements) or even better, reaching 5×10^{15} molec. cm^{-2} for remote areas (after-launch uncertainty analysis; see Sect. 2). These values are quite large compared to the measured levels of HCHO (around 1.5×10^{15} molec. cm^{-2} for very clean sites to around 9×10^{15} molec. cm^{-2} for a city such as Paris). It is therefore necessary to average several pixels in order to reduce the random uncertainty of the TROPOMI mean HCHO data, improve the detection level, and increase the TROPOMI sensitivity to day-to-day variability. For this reason, we choose to average the TROPOMI pixels located within 20 km of the FTIR station. Once we filter out the TROPOMI pixels that do not reach the recommended quality criteria (QA flag > 0.5; see

Table 2. FTIR stations that contribute to the present work: location, altitude (in kilometres above sea level), instrument type, retrieval code, and team.

Station	Latitude	Longitude	Altitude	Instrument	Code	Team
Eureka	80.05° N	86.42° W	0.61	Bruker 125 HR	SFIT4	U. of Toronto
Ny-Ålesund	78.92° N	11.92° E	0.02	Bruker 120 HR	SFIT4	U. of Bremen
Thule	76.52° N	68.77° W	0.22	Bruker 125 HR	SFIT4	NCAR
Kiruna	67.84° N	20.40° E	0.42	Bruker 120/5 HR	PROFFIT	KIT–ASF; IRF Kiruna
Sodankylä	67.37° N	26.63° E	0.19	Bruker 125 HR	SFIT4	FMI; BIRA
St. Petersburg	59.88° N	29.83° E	0.02	Bruker 125 HR	SFIT4	SPbU
Bremen	53.10° N	8.85° E	0.03	Bruker 125 HR	SFIT4	U. of Bremen
Paris	48.97° N	2.37° E	0.06	Bruker 125 HR	PROFFIT	Sorbonne U.
Zugspitze	47.42° N	10.98° E	2.96	Bruker 120/5 HR	PROFFIT	KIT–IFU
Jungfraujoch	46.55° N	7.98° E	3.58	Bruker 120 HR	SFIT4	U. of Liège
Toronto	43.60° N	79.36° W	0.17	Bomem DA8	SFIT4	U. of Toronto
Rikubetsu	43.46° N	143.77° E	0.38	Bruker 120/5 HR	SFIT4	Nagoya U.; NIES
Boulder	40.04° N	105.24° W	1.61	Bruker 125 HR	SFIT4	NCAR
Xianghe	39.75° N	116.96° E	0.05	Bruker 125 HR	SFIT4	CAS; BIRA
Tsukuba	36.05° N	140.12° E	0.03	Bruker 125 HR	SFIT4	NIES; Tohoku U.
Izaña	28.30° N	16.48° W	2.37	Bruker 120/5 HR	PROFFIT	AEMET; KIT–ASF
Mauna Loa	19.54° N	155.57° W	3.40	Bruker 125 HR	SFIT4	NCAR
Mexico City (UNAM)	19.33° N	99.18° W	2.26	Bruker Vertex 80	PROFFIT	UNAM
Altzomoni	19.12° N	98.66° W	3.98	Bruker 120/5 HR	PROFFIT	UNAM
Palau	7.34° N	134.47° E	0.03	Bruker 120/5 M	SFIT4	U. of Bremen
Paramaribo	5.81° N	55.21° W	0.03	Bruker 120/5 M	SFIT4	U. of Bremen
Porto Velho	8.77° S	63.87° W	0.09	Bruker 125 M	SFIT4	BIRA
Maïdo (LA.REUNION.MAIDO)	21.08° S	55.38° E	2.16	Bruker 125 HR	SFIT4	BIRA
Wollongong	34.41° S	150.88° E	0.03	Bruker 125 HR	SFIT4	U. of Wollongong
Lauder	45.04° S	169.68° E	0.37	Bruker 120 HR	SFIT4	NIWA
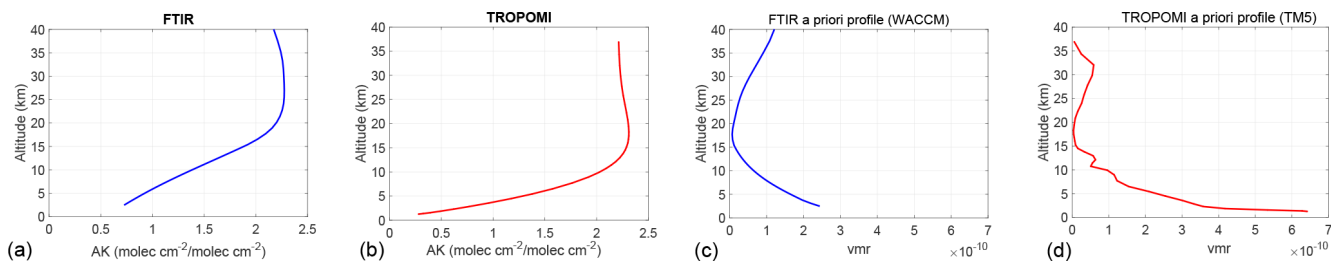
Figure 2. (a, b) Typical total column averaging kernel (AK) from FTIR (blue) and TROPOMI (red) measurements at the Maïdo station (altitude: 2.2 km). **(c, d)** A priori profile used in the FTIR retrievals for the whole time series at Maïdo (blue), and an example of TROPOMI a priori profile at Maïdo (the 18 January 2019).

Sect. 2), we obtain a median value of 34 pixels to average. In cloudy conditions, this number can be smaller. A collocation pair is kept when at least 10 pixels can be averaged. A higher number of pixels can be averaged for Arctic stations (around 45–60), which is useful due to the very low HCHO levels to be detected there. At sub-tropical and tropical stations, the median number of pixels is around 20–29. The higher number of pixels in the Arctic is due to the fact that each FTIR measurement is collocated to all S5P pixels that match the collocation criteria, even if these pixels originate from different orbits with different overpass times.

Before choosing the 20 km collocation criterion, we tested several distances (10, 20, 30, 40, and 50 km). The 10 km criterion was discarded because of the poor number of remaining coincidences, leading to less robust statistics. The 20 to 50 km criteria give similar biases between TROPOMI and FTIR. The standard deviations of the comparisons usually decrease slightly with increasing collocation distance due to a smaller TROPOMI random uncertainty (more pixels to average) except at the most polluted sites. However, the ratio between the standard deviations and the random uncertainty budgets increases with the collocation distance at all sites, pointing to an increased random error due to the collocation.

We therefore choose the 20 km distance to reduce the random spatial collocation error.

The time coincidence criterion is set to ± 3 h. This choice is a compromise to obtain a significant number of coincidences between TROPOMI and FTIR data, noting that the median FTIR measurement frequency is five per day (with a range of 3 to 10 depending on the station). A shorter time coincidence criterion significantly decreases the numbers of sampled collocated days and sometimes months, which is a limitation for checking the TROPOMI seasonality (sampled months: 267 for 1 h and 305 for 3 h criteria). Note that a 6 h criterion would provide 20 additional sampled months: the critical stations are Mauna Loa, Altzomoni, and Paramaribo, for which we would have coincidences back to May 2018. The standard deviations of the TROPOMI–FTIR comparisons are usually smaller with a longer time coincidence criterion, but this can be explained by the increased number of pixels (improved TROPOMI precision of the mean) in the 6 h collocation, mainly at Arctic sites with an increased number of multiple orbits. Despite the smaller standard deviations usually obtained within a 6 h criterion, we finally choose 3 h to reduce the possible impact of some passing plumes and of the HCHO diurnal cycle on the comparisons. The diurnal cycle at most of the FTIR stations can be found in Vigouroux et al. (2018) and its Supplement. At many stations no significant diurnal cycle was observed, but in some cases, mainly polluted sites, we obtained a maximum around 12:00–01:00 p.m. local time, close to the TROPOMI overpass time. At the Mexico City station, where the diurnal cycle amplitude is the greatest, the effect of collocation time (6 h vs 3 h) on the statistical bias is 4 %.

4.2 Building inter-comparable products

Some manipulation of the original data products is needed before looking at the differences between TROPOMI and FTIR data. Both measurements provide total columns (for FTIR) or tropospheric columns (for TROPOMI) that have a lower sensitivity near the ground (see Fig. 2), and their retrievals use a priori profile information that has been chosen differently (TROPOMI: daily a priori profiles from TM5; FTIR: single a priori profile from climatology of WACCM). To correct for this, for each S5P individual pixel collocated with each FTIR measurement, we use the comparison method described in Rodgers and Connor (2003). First, the a priori substitution is applied using the S5P a priori profile $x_{S,a}$ as the common a priori profile. For this, the S5P a priori profile is re-gridded to the FTIR retrieval grid ($x_{S,a/F}$) using a mass conservation algorithm (Langerock et al., 2015). In the rare situation that the satellite pixel elevation is above the FTIR site, the S5P a priori profile is extended to the FTIR instrument's altitude. The re-gridded S5P a priori $x_{S,a/F}$ is then substituted following Rodgers and Connor (2003), and we finally use the corrected FTIR-retrieved profile $x\prime_F$ in the

comparisons:

$$x\prime_F = x_F + (\mathbf{A}_F - \mathbf{I})(x_{F,a} - x_{S,a/F}), \tag{2}$$

where x_F is the original FTIR-retrieved profile, $\mathbf{A}_F$ is the FTIR averaging kernel matrix, $\mathbf{I}$ is the unit matrix, and $x_{F,a}$ is the FTIR a priori profile.

The next step, following Rodgers and Connor (2003), is to smooth the corrected FTIR profile with the S5P column averaging kernel a_S. For that purpose we re-grid the corrected FTIR profile $x\prime_F$ to the S5P column averaging kernel grid ($x\prime_{F/S}$) and apply the smoothing equation:

$$c_F^{smoo} = c_{S,a} + a_S(x\prime_{F/S} - x_{S,a}), \tag{3}$$

with $c_{S,a}$ being the S5P a priori column derived from the S5P a priori profile. We obtain a smoothed FTIR column c_F^{smoo} associated with a collocated TROPOMI pixel. In the case of mountain sites where the pixel altitude is below the instrument's height, the re-gridding of the FTIR profile $x\prime_{F/S}$ is done such that the FTIR profile is extended with the S5P a priori profile (such an extension is invariant under the latter smoothing equation). Note that this FTIR re-gridding to the satellite grid also has the advantage that only the FTIR profile up to the altitude of the satellite product (which is only a tropospheric column) remains in the re-gridded column: we therefore finally compare tropospheric columns in both products.

Next, we need to take into account that, for mountain stations, the difference between satellite columns and the original ground-based columns can be significant. To bring both satellite and smoothed FTIR column c_F^{smoo} (which is calculated as a column valid at the satellite's pixel surface) values to the scale of the original FTIR columns, we apply a scaling factor f representative of the fraction of the partial column between the satellite pixel altitude and the FTIR station altitude. This scaling factor is derived from the satellite a priori profile and is defined as

$$f = 1 - \frac{c_{S,a}^{\Delta z}}{c_{S,a}}, \tag{4}$$

where $c_{S,a}^{\Delta z}$ denotes the partial column derived from the S5P a priori profile between the pixel surface and the FTIR station. The TROPOMI column c_S and its random and systematic uncertainties are also scaled with the same factor so that finally the collocated products are all expressed at the altitude of the FTIR site (and not of the pixel surface). For mountain stations, the scaling factor f, calculated for each satellite's pixel, can reach a minimum of 0.5 for stations located at about 2 km altitude from the satellite's pixel surface (Maïdo, Izaña, or Altzomoni) or even 0.3 at the higher sites Jungfraujoch and Zugspitze, while at sea-level sites it is of course close to 1.0. In the rare cases where the satellite pixel is above the FTIR station, we apply the conversion factor

$f = 1 + c_{S,a}^{\Delta z}/c_{S,a}$, where the satellite a priori profile is extrapolated to the station surface in order to calculate the a priori partial column between both altitudes.

The final step is to average the individual smoothed and scaled FTIR columns $c_F^{smoo} \times f$ that are taken within 3 h and the TROPOMI $c_S \times f$ individual pixel columns that are available within 20 km (which can belong to different orbits) to form the collocated pair $FTIR_i$ and $TROP_i$ used in the next section.

4.3 Estimation of the TROPOMI accuracy and precision

In Sect. 5.1, we assess whether the TROPOMI accuracy is compliant with pre-launch requirements (40 %–80 %, as reported in the official ESA document S5P-RS-ESA-SY-164, 2014, Table 3, p. 19). The accuracy of the TROPOMI HCHO measurements will be estimated by deriving the median of the relative differences (BIAS) between the collocated $TROP_i$ and the reference $FTIR_i$ data at each station:

$$BIAS = med\left(\frac{(TROP_i - FTIR_i)}{FTIR_i}\right). \tag{5}$$

We can note that the applied scaling factor f (see previous section) does not affect the BIAS estimation even at high mountain stations because it cancels in the division.

For robust statistics, the median is preferred to the mean due to the presence of outliers (a few remaining TROPOMI outliers after the QA filter and some very small FTIR values that give very large relative differences after the division in Eq. 5). The presence of TROPOMI outliers is minimized by using the median, but they should ideally be removed by the QA filter. An improvement of the QA value is foreseen in the next product version, which should improve, for example, the filtering at Arctic sites (SZA > 75°).

In the next section, we also compare the obtained BIAS with the systematic uncertainty in the difference σ_{syst} to evaluate the TROPOMI uncertainty budget:

$$\sigma_{syst}^2 = (\sigma_{S,syst})^2 + a_S^T S_{F,syst} a_S$$
$$+ a_S^T (I - A_F) S_{var,syst} (I - A_F)^T a_S, \tag{6}$$

where $\sigma_{S,syst}$ is the systematic uncertainty of TROPOMI columns, as provided in the public release database (but scaled for altitude; see Sect. 4.2); a_S is the TROPOMI total column averaging kernel; and $S_{F,syst}$ is the FTIR systematic covariance matrix provided in volume mixing ratio units in the standardized GEOMS format, converted to partial column units in Eq. 6. The last term is the impact of different low-vertical-resolution profile measurements (the smoothing error) on the comparisons (see Eq. 27 in Rodgers and Connor, 2003), where, for the systematic uncertainty part, we account for possible bias in $x_{S,a}$ by following von Clarmann (2014):

$$S_{var,syst} = (x_{S,a} - <x>)(x_{S,a} - <x>)^T. \tag{}$$

The $x_{S,a} - <x>$ is not known, and we follow Vigouroux et al. (2018), with $x_{S,a} - <x> = -50\%, -20\%, -10\%, +10\%, +8\%$, and $+5\%$ for the ground–4, 4–8, 8–13, 13–25, 25–40, and 40–120 km layers, respectively (expressed in molecules per square centimetre). The last term of Eq. (6) is found to be of the order of a few per cent and therefore negligible in σ_{syst}. In practice, the systematic uncertainty in the difference σ_{syst} is dominated by the TROPOMI systematic uncertainty of about 40 %, FTIR having a median systematic uncertainty of only 13 % and a maximum of 26 % (see Sect. 3).

Similarly, the precision of the TROPOMI HCHO products is estimated in Sect. 5.2 not with the usual standard deviation, which is not robust in the case of outliers, but with the median absolute deviation (MAD; see Huber, 1981) of the differences ($DIFF_i = TROP_i\text{-}FTIR_i$):

$$MAD = k \times med(abs(DIFF_i - med(DIFF_i))), \tag{7}$$

where $k = 1.4826$ for a correspondence with the 1σ standard deviation for normal distribution without outliers.

In Sect. 5.2, we compare the obtained MAD to the random uncertainty in the differences σ_{rand}, which is calculated following Rodgers and Connor (2003):

$$\sigma_{rand}^2 = (\sigma_{S,rand})^2 + a_S^T S_{F,rand} a_S$$
$$+ a_S^T (I - A_F) S_{var,rand} (I - A_F)^T a_S, \tag{8}$$

where $\sigma_{S,rand}$ is the random uncertainty of TROPOMI columns, as provided in the public release database (but scaled for altitude; see Sect. 4.2); $S_{F,rand}$ is the FTIR random covariance matrix; and $S_{var,rand}$, to take into account the impact of low vertical resolution on the random part of the uncertainty, is the natural variability matrix chosen to be 50 %, 50 %, 40 %, 35 %, 30 %, 30 %, and 10 % for the ground–4, 4–8, 8–13, 13–25, 25–40 km, and 40–120 km layers, respectively (expressed in molecules per square centimetre). As for the systematic uncertainty part, the random uncertainty in the difference is dominated by the TROPOMI random uncertainty (median of about 1.1×10^{15} molec. cm^{-2} for $TROP_i$ within 20 km), while $FTIR_i$ has a median random uncertainty of 2.0×10^{14} molec. cm^{-2}. The last term of Eq. (8) is comparable to the FTIR one (median value of 2.4×10^{14} molec. cm^{-2}).

We can use MAD as an upper limit of the TROPOMI precision since collocation in space and time of the sounded air masses is never perfect. It is compared in the next section to the pre-launch precision requirement. The MAD estimation is influenced by the scaling factor f, which is important only for high-altitude sites (Sect. 4.2). It should be interpreted as an estimation of the precision of a TROPOMI column that would be measured at the altitude of the FTIR site. The random uncertainty in the differences is also expressed at the altitude of the FTIR site so that the comparison between MAD and σ_{rand} is always valid.

The observed BIAS between TROPOMI and the reference FTIR data is statistically significant if it exceeds its statistical error: $\text{ERR}_B = 2 \times \text{MAD}/\sqrt{n}$ (with n being the number of coincidences).

5 Validation results

In this section, we provide a table and plots for the offline (RPRO+OFFL) HCHO TROPOMI product. We do not show detailed results for the near-real-time (NRTI) product versions 1.1.(5-7) because they are very similar to the offline version. Numbers for the main conclusions will be given in the text for this NRTI product.

5.1 TROPOMI observed BIAS and accuracy

In Table 3, we provide, at each individual FTIR station, the mean of the FTIR HCHO total columns (mean FTIR), the obtained median of the relative differences (BIAS; in per cent to compare with the pre-launch TROPOMI accuracy requirements of 40 %–80 %; Eq. 5), the error in the BIAS (ERR$_B$), and the number of collocated pairs n. The systematic uncertainty in a single difference is also given (in per cent; calculated from Eq. 6, where each term has been expressed in per cent; dividing by each individual instrument HCHO column).

We have ordered the stations not according to decreasing latitude as in Table 2 but to increasing mean HCHO FTIR column. The reason is that we observe a tendency of the BIAS between TROPOMI and FTIR: while the BIAS is always (with the exception of Eureka) positive or not significant (if BIAS < ERR$_B$) for very clean to clean sites (with mean HCHO levels lower than 6.5×10^{15} molec. cm^{-2}), it is negative and very consistent for the stations with higher HCHO levels ranging from 8.7 to 28.6×10^{15} molec. cm^{-2} (-29 % to -36 %), with a small error in the bias (2 % to 6 %). Note that the BIAS is also consistent at Paramaribo (-26 %) but with a larger error (14 %) due to a small number of collocations. This dependence of the TROPOMI bias on the HCHO concentration levels is visualized in Fig. 3, where the BIAS at each station is plotted as a function of the mean FTIR columns. It is therefore not appropriate to use the median bias obtained using the data from all stations together (-10 %) if one wants to correct the TROPOMI HCHO data in model inversion studies. If we calculate the median of the differences for HCHO FTIR columns $> 8.0 \times 10^{15}$ molec. cm^{-2}, we obtain a significant negative bias of -30.8 ± 1.4 %. The detection of this bias is especially important for modelling studies that use satellite data to optimize the volatile organic compound emission sources, as done in e.g. Fortems-Cheiney et al. (2012) and Stavrakou et al. (2015) with OMI and GOME-2. The bias for clean HCHO levels ($< 2.5 \times 10^{15}$ molec. cm^{-2}) is significantly positive ($+26 \pm 5$ %).

The validation results for the NRTI TROPOMI products give very similar results: a negative BIAS (-31.7 ± 1.8 %) for the high HCHO levels ($> 8.0 \times 10^{15}$ molec. cm^{-2}) and a positive one ($+22 \pm 7$ %) for low HCHO levels ($< 2.5 \times 10^{15}$ molec. cm^{-2}). The small differences are mainly due to the different sampling of the comparisons (NRTI data have been retrieved since December 2018, while the OFFL data have been retrieved since May 2018).

The different TROPOMI BIAS at different HCHO levels points to the presence of two kinds of bias: a constant one and a proportional one. They can be obtained by using the scatter plot of the two instruments shown in Fig. 4: the constant bias is the intercept of the linear relationship between TROPOMI and FTIR, while the proportional bias is given by its slope. But this has to be done carefully: a usual linear regression by ordinary least squares (OLS) is not statistically robust and can give spurious results in the presence of outliers and/or heteroscedasticity. We are confronted with both problems in our scatter plot: we do have outliers, and the uncertainty increases with HCHO levels. Therefore, we use the robust Theil–Sen estimator (Sen, 1968), where the slope s of the scatter plot is the median of the slopes of the lines through all pairs of data points $(\text{TROP}_j - \text{TROP}_i)/(\text{FTIR}_j - \text{FTIR}_i)$, with $\text{FTIR}_j \neq \text{FTIR}_i$. The intercept b is then the median of $\text{TROP}_i - s \times \text{FTIR}_i$. Using this robust estimator, we obtain the relation $\text{TROP} = 0.64 \times \text{FTIR} + 1.10 \times 10^{15}$ molec. cm^{-2}. We have calculated the uncertainties in s and b using $2 \times \text{MAD}/\sqrt{n}$, with MAD being the median absolute deviation of the slopes and intercepts of the pairs of data points and n being the number of pairs. We obtain an uncertainty of 0.03 and 0.05×10^{15} molec. cm^{-2} for s and b, respectively. Therefore, both the constant ($1.10 \pm 0.05 \times 10^{15}$ molec. cm^{-2}) and proportional (0.64 ± 0.03 %) biases are significant.

Using the scatter plot to derive the constant and proportional biases is very sensitive to the range of observed values. As an example, if one would only use HCHO FTIR data $> 8.5 \times 10^{15}$ molec. cm^{-2}, one would obtain a slope of 0.51 and an intercept of 3.2×10^{15} molec. cm^{-2}, which would point to a strong overestimation and underestimation of the constant and proportional biases, respectively. This would lead to erroneous prediction of overall TROPOMI bias outside the range of observed values. This demonstrates why such a ground-based network covering very clean sites to high-HCHO-level sites is crucial to providing a good estimate of both constant and proportional biases of TROPOMI.

The BIAS values given in Table 3 are a combination of the constant and proportional biases and can be used to statistically assess the TROPOMI HCHO overall accuracy. We can easily see from Table 3 that all BIAS values are within the upper limit of the pre-launch requirement of 80 %, and they are within the 40 % lower limit for 20 of the 25 stations. The five stations exceeding a 40 % BIAS are clean (Arctic or mountains) sites, with mean HCHO columns below 2.5×10^{15} molec. cm^{-2}. But these are sites where the systematic uncertainty in the differences (see Table 3 and Eq. 6)

Table 3. Validation of TROPOMI RPRO + OFFL. Please note that the ordering of the sites is by increasing mean HCHO column. For each station: mean of the HCHO FTIR total columns (in molecules per square centimetre); median of the relative differences, BIAS = med $((\text{TROP}_i\text{-FTIR}_i)/\text{FTIR}_i)$ and its error ERR_B (in per cent; see text); number of collocated pairs n; systematic uncertainty in a single difference σ_{syst} (in per cent; Eq. 6); median absolute deviation (MAD; in molecules per square centimetre; Eq. 7); random uncertainty in a single difference σ_{rand} (in molecules per square centimetre; Eq. 8); and pre-launch TROPOMI precision requirements associated with the choice of 20 km around the station Requ = $1.2 \times 10^{16}/\sqrt{n_{\text{pix}}}$ molec. cm^{-2}, with n_{pix} being the mean number of pixels averaged in the collocated TROPOMI data. The Pearson correlation coefficient R is given for individual coincidences (± 3 h) and for monthly means of coincident data.

Station	Mean FTIR (molec. cm^{-2})	BIAS ($\pm \text{ERR}_B$ %)	n	σ_{syst} (%)	MAD (molec. cm^{-2})	σ_{rand} (molec. cm^{-2})	Requ. (molec. cm^{-2})	n_{pix}	R (indiv.)	R (monthly)
Jungfraujoch	1.24E+15	19 ± 15	87	58	9.0E14	5.6E14	2.5E15	24	0.61	0.70
Zugspitze	1.36E+15	52 ± 10	184	59	7.8E14	5.0E14	2.1E15	33	0.71	0.86
Mauna Loa	1.60E+15	52 ± 22	52	54	1.3E15	8.8E14	2.5E15	23	−0.09	−0.05
Eureka	1.65E+15	−40 ± 11	114	97	1.1E15	5.3E14	1.8E15	45	0.22	0.43
Maïdo	1.86E+15	5 ± 9	155	43	1.0E15	7.1E14	3.0E15	16	0.45	0.53
Ny-Ålesund	1.90E+15	43 ± 20	47	41	1.2E15	4.9E14	1.7E15	52	0.35	0.38
Thule	2.06E+15	−3 ± 5	346	57	9.6E14	4.8E14	1.6E15	60	0.56	0.82
Izaña	2.07E+15	13 ± 10	97	83	8.9E14	6.4E14	2.5E15	24	0.47	0.79
Altzomoni	2.44E+15	66 ± 18	67	42	1.6E15	8.9E14	2.6E15	22	0.50	0.86
Kiruna	2.44E+15	50 ± 12	146	67	1.5E15	8.8E14	1.7E15	60	0.64	0.72
Lauder	2.54E+15	−11 ± 14	225	78	2.6E15	1.3E15	2.1E15	33	0.38	0.65
Rikubetsu	3.16E+15	26 ± 40	16	50	2.8E15	1.0E15	1.9E15	41	0.45	0.60
Palau	3.80E+15	0 ± 15	10	36	9.8E14	8.2E14	2.7E15	20	0.15	0.33
Sodankyla	4.15E+15	8 ± 7	307	51	2.5E15	1.2E15	1.7E15	48	0.51	0.69
Boulder	5.91E+15	−1 ± 9	103	50	2.2E15	1.3E15	2.2E15	31	0.79	0.90
St. Petersburg	6.21E+15	−4 ± 8	158	44	3.0E15	1.2E15	1.9E15	42	0.68	0.78
Wollongong	6.36E+15	9 ± 8	322	54	3.3E15	1.9E15	2.3E15	27	0.78	0.94
Tsukuba	7.05E+15	−23 ± 12	34	44	3.1E15	1.2E15	2.2E15	31	0.68	0.51
Bremen	7.77E+15	−5 ± 12	46	39	3.2E15	1.4E15	1.8E15	43	0.63	0.68
Paramaribo	8.43E+15	−26 ± 14	15	36	3.3E15	1.3E15	2.5E15	23	0.12	0.14
Paris	8.72E+15	−29 ± 6	128	44	3.1E15	1.2E15	1.9E15	41	0.76	0.79
Toronto	1.06E+16	−34 ± 4	251	38	3.7E15	1.6E15	2.3E15	29	0.73	0.95
Xianghe	1.43E+16	−33 ± 2	384	38	5.5E15	2.1E15	2.0E15	36	0.86	0.97
Mexico City (UNAM)	1.92E+16	−27 ± 4	154	32	5.9E15	2.6E15	2.5E15	25	0.34	0.27
Porto Velho	2.86E+16	−36 ± 3	81	31	8.3E15	3.6E15	2.2E15	29	0.81	1.00
All stations BIAS	6.60E+15	−10 ± 2	3529	48	2.4E15	1.2E15	2.1E15	34	0.87	0.91
Low HCHO BIAS FTIR $< 2.5 \times 10^{15}$	1.64E+15	+26 ± 5	1321	52	1.3E15	7.7E14	2.2E15	31	0.40	
High HCHO BIAS FTIR $> 8.0 \times 10^{15}$	1.57E+16	−30.8 ± 1.4	952	46	4.8E15	2.1E15	2.1E15	33	0.88	

is usually also the largest, leading to a good correspondence between observed higher BIAS and higher calculated uncertainty for three of these five stations (Zugspitze, Mauna Loa, and Kiruna).

Therefore, we can conclude that the TROPOMI HCHO accuracy satisfies the pre-launch requirements and that the systematic uncertainty budget is in very good agreement with observed bias except at very few stations (Ny-Ålesund: 43 > 41 %; Altzomoni: 71 > 42 %; and Porto Velho: 36 > 31 %). At most of the other stations, the reported systematic uncertainty tends to be larger than the BIAS. We find the same conclusions with TROPOMI accuracy when making comparisons with the NRTI products.

The systematic uncertainties leading to the observed constant and proportional biases of our study have been calculated as described in Sect. 3 of De Smedt et al. (2018). From the error propagation of the HCHO TROPOMI tropospheric

columns (see Eq. 1), it can be found that the proportional bias is more likely due to air mass factor (M) uncertainties σ_M, while the constant bias is more likely due to the uncertainties of the slant column uncertainties $\sigma_{N,S}$ and to the uncertainty of the background correction of the slant columns. This can be seen in Eq. (13) of De Smedt et al. (2018), where σ_M is proportional to $N_s - N_{(s,0)}$. We can list some known difficulties of the satellite product:

- The negative bias over high-HCHO-level sites (biomass burning or megacities) could be due to aerosol effects. There is no plan to include a correction for aerosols in the operational product, but specific studies are foreseen to check its impact in a scientific product.

- The positive bias over clean polar sites could be due to the solar zenith angle (SZA) dependency of the slant column fit results (because of spectral interferences with
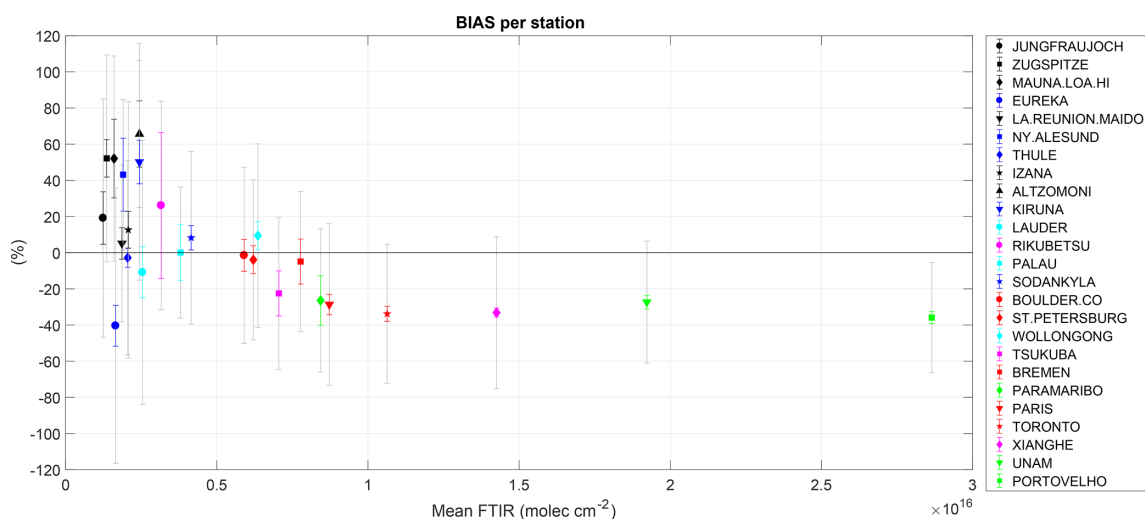
Figure 3. BIAS at each station (in per cent) as a function of the mean FTIR total columns (molec. cm^{-2}). The grey bars are the systematic uncertainty of the differences σ_{syst}, and the coloured error bars are the 2σ error in the bias ERR_B. Black markers are for mountain stations; blue markers are for Arctic stations; cyan markers are for Oceania, Australia, and New Zealand; magenta markers are for China and Japan; red markers are for mid-latitude Europe and North America; and green markers are for Central and South America.
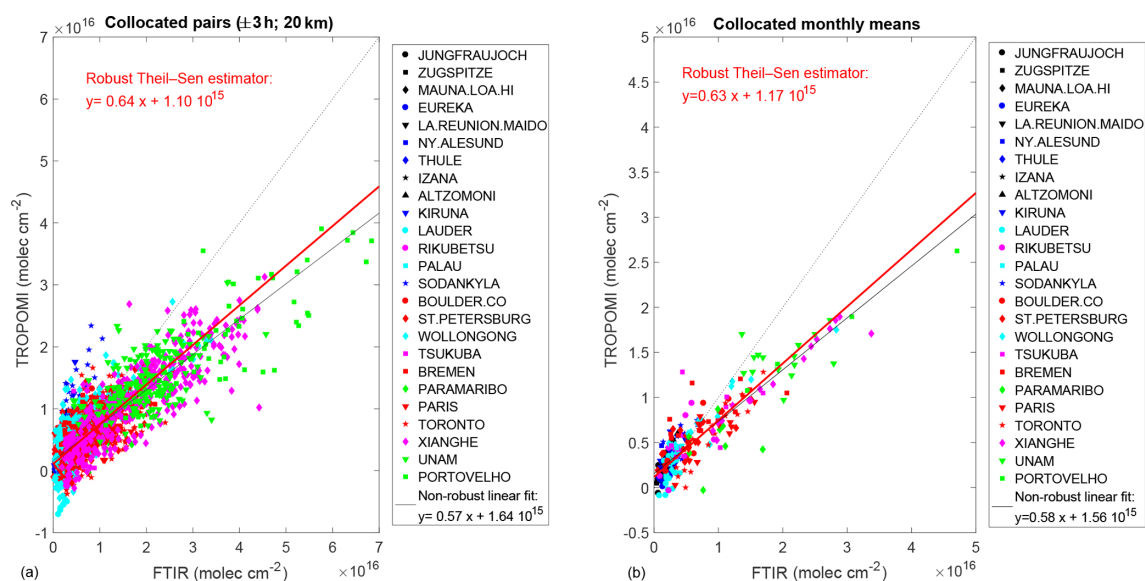
Figure 4. Scatter plots of TROPOMI versus FTIR data for individual collocated pairs (± 3 h; **a**), and for the monthly means of collocated data (**b**). The non-robust OLS fit of the data is given in the legend and plotted as a black line, while the slope and intercept obtained with the robust Theil–Sen estimator are given by the red line and text.

ozone and BrO). As explained in Sect. 2, the QA values need to be improved at large SZAs, which is foreseen in the next version.

– The current albedo climatology is too coarse for TROPOMI, which could especially be a problem for polar, mountain, or coastal sites. A climatology based on TROPOMI measurements is under development.

– It is also foreseen to test a regional model at higher spatial resolution for an improvement of the a priori HCHO profiles. This should improve the TROPOMI-retrieved product, especially at polluted sites. However, the validation presented here already takes the a priori information and averaging kernels into account. We therefore do not expect an important effect of the improved a priori profiles on the validation results.
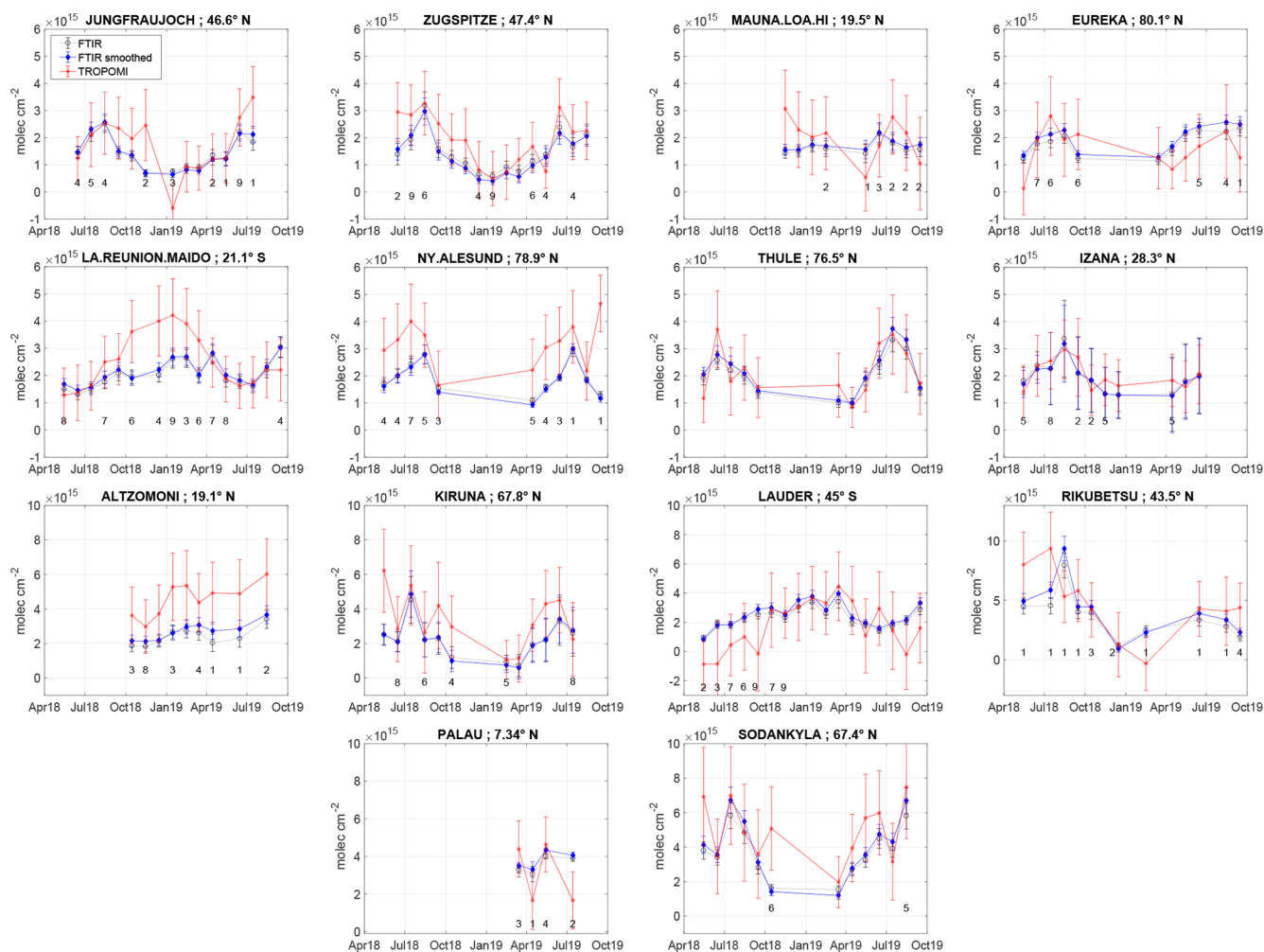
Figure 5.

5.2 Observed TROPOMI precision

For discussing the observed TROPOMI precision, we provide in Table 3 the MAD for each station (in absolute value to compare with the pre-launch precision requirement of 1.2×10^{16} molec. cm^{-2} for a single pixel; Eq. 7). Indeed, for each site, the MAD is an upper limit for the TROPOMI precision as determined by our validation (see Sect. 4.3), while the σ_{rand} given in Table 3 is an approximation of the precision as provided in the satellite product (because the FTIR random uncertainty is much smaller than that of the TROPOMI). The detection limit, usually defined as being 3 times the precision, can then be obtained at each station for an average of TROPOMI pixels within 20 km by multiplying either the σ_{rand} (theoretical estimation by TROPOMI data providers, which is probably underestimated, as seen below) or the MAD (upper limit determined by the validation), both given in Table 3, by 3. The precision pre-launch requirement is provided at each site, taking into account the mean number of pixels n_{pix}

involved in the collocated TROPOMI data (Requ. $= 1.2 \times 10^{16}$ molec. cm$^{-2}/\sqrt{n}_{\mathrm{pix}}$). We see that for all the cleanest sites ($< 2.5 \times 10^{15}$ molec. cm^{-2}), where an additional collocation uncertainty is expected to be small, the MAD is well within the pre-launch requirements. The MAD for these cleanest sites has a median of 1.3×10^{15} molec. cm^{-2} and a minimum of 0.9×10^{15} molec. cm^{-2}. This is a good estimate of the precision that TROPOMI can reach in remote conditions. For a single pixel, the best TROPOMI precision under remote conditions is therefore 5–8×10^{15} molec. cm^{-2}.

It must be noted that the pre-launch HCHO precision requirements were chosen based on pre-launch requirements for the instrument signal-to-noise ratio (equivalent to OMI). The actual signal-to-noise ratio of the measurements appears to be better than the requirements, especially in the HCHO wavelength fitting range. Furthermore, the good quality of the recorded spectra allowed for an increase in the size of the TROPOMI HCHO fitting spectral interval just after launch, further improving the precision of the slant columns. Indeed,
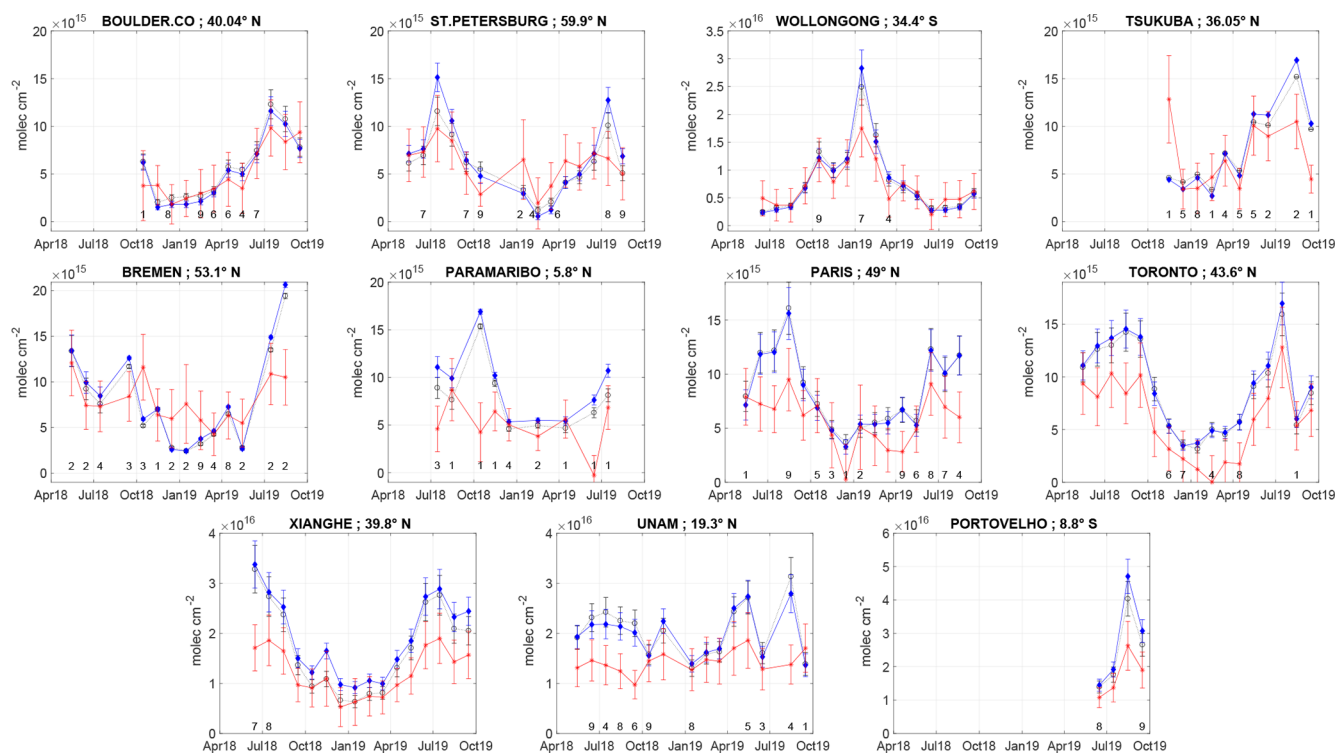
Figure 5. Monthly mean time series of FTIR raw data (black), FTIR data smoothed with the TROPOMI column averaging kernel (blue), and TROPOMI (red) at each site. Only data in coincidence are included in the monthly mean to avoid sampling bias. When the number of coincidences within 1 month is smaller than 10, it is written below the monthly mean.

as seen in Table 3, only at the three highest-HCHO-level sites (Xianghe, Mexico City, and Porto Velho) are the provided random uncertainties as high as the pre-launch requirements. The actual provided random uncertainty is smaller, and we can see that even for clean sites the observed MAD is larger than the random uncertainty in the differences by a factor of 1.6. This factor increases up to 1.8 if we take into account all the stations, but this is expected due to a collocation uncertainty that should have more impact at high-level sites (the factor rises up to 2.3 for high HCHO levels $> 8.0 \times 10^{15}$ molec. cm^{-2}). Our comparisons suggest that the TROPOMI random uncertainty is underestimated by at least a factor of 1.6 and up to a maximum of 2.3 (if one were to assume the collocation uncertainty to be smaller than the TROPOMI uncertainty). This underestimation could be due to the fact that the uncertainties associated with the air mass factor calculation and with the background correction step are currently assumed to be fully systematic. The discrimination between the random and systematic part of the uncertainties might be refined in the future based on such validation results.

5.3 Observed TROPOMI monthly variability

The Pearson correlation coefficient is very good for the collocated monthly means of TROPOMI and FTIR data (0.91; see Table 3 and Fig. 4) and is usually good for individual sites. However, Pearson correlation is not robust and can give a wrong conclusion when only few data are coincident, especially when outliers are present. We have 17 months of coincident TROPOMI and FTIR measurements in the best cases, while there are only 4 for the newest stations Palau and Porto Velho. We therefore verify that the TROPOMI precision allows the seasonal variability to be well captured even at very clean sites, which can be at the limit of the satellite detection, by plotting the individual monthly mean time series in Fig. 5.

The seasonal variability, with a maximum in July–August, is well observed at all the Arctic sites (Eureka, Ny-Ålesund, Thule, Kiruna, and Sodankylä). The monthly mean correlation is better than 0.69 except at Eureka and Ny-Ålesund. It can be seen in Fig. 5 that September 2019 is very high in TROPOMI data at Ny-Ålesund, and only one coincidence is found for this month. Removing this last outlier gives a 0.76 correlation coefficient at this station. The northern mid-latitude clean sites (mountains: Jungfraujoch, Zugspitze, Izaña) also display a seasonal variability in very good agreement, with correlation coefficients higher than 0.70. The clean Japanese site Rikubetsu shows poorer correlation (0.60), but only few data are in coincidence. The stations where we find the poorer correlations are the oceanic sites. The poorest one is Mauna Loa, but this is expected due to the very small seasonal variability there and the small number

of coincidences. A similar situation is observed at the other recent oceanic site Palau, where only 4 months of data are available. At the oceanic site Maïdo, we find a good agreement in most of the months but not in October–December, which are the predominant biomass-burning months in the region, so the collocation of the plumes might play a role there. Finally, at Lauder, TROPOMI shows many negative values in the beginning of the period (May–September 2018), which is responsible for a lower correlation (0.65) and for the negative bias there (although not significant), while other clean sites usually show positive ones (see Table 3).

The higher-HCHO-level sites show a TROPOMI seasonal variability in very good agreement with FTIR, with correlation larger than 0.90 for Boulder, Wollongong, Toronto, Xianghe, and Porto Velho. At Tsukuba, removing the outlier of one coincidence in November 2018 increases the correlation to 0.93. The poorest correlation (0.14) is found at the coastal site Paramaribo, where usually only one coincidence per month is found. Looking at the highest-HCHO-level sites, these monthly mean time series also confirm that TROPOMI has more difficulty reproducing the months with the highest enhancements, which is responsible for the significant negative bias (-31%) found in the previous section for high HCHO levels ($> 8.0 \times 10^{15}$ molec. cm^{-2}).

6   Conclusions

We have used a network of 25 FTIR stations, most of them affiliated with the NDACC, to validate the latest TROPOMI HCHO tropospheric columns (v.1.1.[5-7]). This network covers a wide range of concentrations, from very clean Arctic, oceanic, and mountain sites with columns that can be lower than 10^{14} molec. cm^{-2} to high-HCHO-level sites such as Mexico City or Porto Velho, near the Amazon rainforest, where columns up to 7×10^{16} molec. cm^{-2} have been observed.

We found an overestimation ($+26 \pm 5\%$) of TROPOMI OFFL products for very small HCHO columns ($< 2.5 \times 10^{15}$ molec. cm^{-2}) and an underestimation of TROPOMI of about -30.8% ($\pm 1.4\%$) for high HCHO columns ($> 8.0 \times 10^{15}$ molec. cm^{-2}), which can be used, for example, to correct TROPOMI data near emission sources. The results are very similar for NRTI products ($+22 \pm 7\%$ and $-31.7 \pm 1.8\%$ for small and high columns, respectively), and the differences are mainly due to the different period of available TROPOMI v.1.1.(5-7) products. Our wide range of HCHO levels and the use of the Theil–Sen method allow us to derive robust and significant constant (intercept) and proportional (slope) biases of TROPOMI (TROP $= +1.10 \pm 0.05 \times 10^{15} + 0.64 \pm 0.03 \times$ FTIR; in molecules per square centimetre). Such different BIAS values for low and high target species concentration levels due to the presence of both constant and proportional biases were also recently observed (although with fewer FTIR sites involved) in another nadir

satellite product, the formic acid observed by the Infrared Atmospheric Sounding Interferometer (IASI; Supporting Information in Franco et al., 2020). The NDACC FTIR network, which covers a large number of atmospheric species at wide ranges of concentrations, is a powerful source of reference data to detect such nadir satellites' biases.

Although significant, the observed overestimation and underestimation of TROPOMI are within the lower limits of the pre-launch requirements ($\pm 40\%$), as are the biases at individual sites for 20 of the 25 stations. The TROPOMI systematic uncertainty budget is in very good agreement with the observed bias, larger uncertainties being reported at stations where the bias exceeds the 40 % requirements. Possible improvements in the TROPOMI biases could be achieved by taking into account aerosol effects over polluted sites, improving the QA values at high SZAs, and using an albedo climatology and a priori HCHO profiles at the TROPOMI spatial resolution. Except for the former, these improvements are foreseen in the next versions of operational TROPOMI data.

The precision of TROPOMI OFFL products is estimated by the median absolute deviation (MAD) at the clean sites, where the collocation effect is expected to be small. For FTIR HCHO levels lower than 2.5×10^{15} molec. cm^{-2}, the MAD is 1.3×10^{15} molec. cm^{-2}, corresponding to a single-pixel precision of 7×10^{15} molec. cm^{-2} (5 to 8×10^{15} at individual sites), which is well below the pre-launch precision requirements of 1.2×10^{16} molec. cm^{-2}. However, the provided TROPOMI random uncertainties (after launch) were indeed found to be better than the pre-launch requirements, but they are too small by a factor of 1.6 compared to the MAD at the clean sites. There is a difference by a factor of 2.3 between MAD and the random uncertainty in the comparisons (dominated by TROPOMI random uncertainty) at the high-level sites, where an additional effect of collocation might play a role as well. The underestimation of the TROPOMI random uncertainty could be due to a random effect of the uncertainty associated with the air mass factor calculation that is not currently included in the budget. This would also explain a larger underestimation of random error at high-level sites (factor of 2.3 vs 1.6 at clean sites). Furthermore, a systematic uncertainty component over a short term (so not included in the TROPOMI random uncertainty) can have a random effect on our longer-term comparisons.

We have shown that the TROPOMI data capture the HCHO seasonal variability very well, even at very clean sites. The Pearson correlation coefficient for monthly mean coincident data is 0.91. Although we have found room for a refinement of the TROPOMI random uncertainty estimation and for an improvement of the QA values for a better filtering of the remaining few outliers and negative columns (exceeding the expected statistical distribution), this validation work has demonstrated the very good quality of the TROPOMI HCHO product, which is well within the pre-launch requirements for both accuracy and precision. This work has also

shown the high value of the FTIR HCHO network, providing harmonized and well-characterized data covering a wide range of HCHO columns. These ground-based FTIR data are continuously extended by new measurements and will be used in the coming years for the routine S5P validation within the ESA-dedicated validation server (https://mpc-vdaf-server.tropomi.eu/, last access: 7 July 2020). The FTIR network will also be used in the near future for the validation of previous satellite missions such as OMI or GOME-2. New FTIR measurements are continuously performed and can be used in the coming years for the validation of new satellite generations, such as TEMPO, GEMS, Sentinel 5P, or Sentinel 4.

An extension of this TROPOMI HCHO validation with ground-based MAX-DOAS and Pandora instruments, especially at sites where both FTIR and UV–visible techniques are available (e.g. Xianghe, Maïdo, and Lauder) or in uncovered regions (Africa), would bring additional knowledge. However, there is first a need for data product harmonization within the MAX-DOAS network as was done with the FTIR network used here. This work is ongoing as part of the ESA FRM4DOAS and Pandonia projects.

Review statement. This paper was edited by Hartmut Boesch and reviewed by two anonymous referees.

References

Boersma, K. F., Eskes, H. J., and Brinksma, E. J.: Error analysis for tropospheric NO_2 retrieval from space, J. Geophys. Res., 109, D04311, https://doi.org/10.1029/2003JD003962, 2004.

De Smedt, I., Stavrakou, T., Hendrick, F., Danckaert, T., Vlemmix, T., Pinardi, G., Theys, N., Lerot, C., Gielen, C., Vigouroux, C., Hermans, C., Fayt, C., Veefkind, P., Müller, J.-F., and Van Roozendael, M.: Diurnal, seasonal and long-term variations of global formaldehyde columns inferred from combined OMI and GOME-2 observations, Atmos. Chem. Phys., 15, 12519–12545, https://doi.org/10.5194/acp-15-12519-2015, 2015.

De Smedt, I., Yu, H., Richter, A., Beirle, S., Eskes, H., Boersma, K.F., Van Roozendael, M., Van Geffen, J., Lorente, A., and Peters, E.: QA4ECV HCHO tropospheric column data from OMI (Version 1.1) [Data set], Royal Belgian Institute for Space Aeronomy, https://doi.org/10.18758/71021031, 2017.

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

Fortems-Cheiney, A., Chevallier, F., Pison, I., Bousquet, P., Saunois, M., Szopa, S., Cressot, C., Kurosu, T. P., Chance, K., and Fried, A.: The formaldehyde budget as seen by a global-scale multi-constraint and multi-species inversion system, Atmos. Chem. Phys., 12, 6699–6721, https://doi.org/10.5194/acp-12-6699-2012, 2012.

Franco, B., Clarisse, L., Stavrakou, T., Müller, J.-F., Taraborrelli, D., Hadji-Lazaro, J., Hannigan, J. W., Hase, F., Hurtmans, D., Jones, N., Lutsch, E., Mahieu, E., Ortega, I., Schneider, M., Strong, K., Vigouroux, C., Clerbaux, C., and Coheur, P.-F.: Spaceborne Measurements of Formic and Acetic Acid: A Global View of the Regional Sources, Geophys. Res. Lett., 47, e60094, https://doi.org/10.1029/2019GL086239, 2020.

Garcia, R. R., Marsh, D. R., Kinnison, D. E., Boville, B. A., and Sassi, F.: Simulation of secular trends in the middle atmosphere, 1950–2003, J. Geophys. Res., 112, D09301, https://doi.org/10.1029/2006JD007485, 2007.

Hase, F., Hannigan, J. W., Coffey, M. T., Goldman, A., Höpfner, M., Jones, N. B., Rinsland, C. P., and Wood, S. W.: Intercomparison of retrieval codes used for the analysis of high-resolution, ground-based FTIR measurements, J. Quant. Spectros. Ra., 87, 25–52, 2004.

Hase, F., Demoulin, P., Sauval, A. J., Toon, G. C., Bernath, P. F., Goldman, A., Hannigan, J. W., and Rinsland, C. P.: An empirical line-by-line model for the infrared solar transmittance spectrum from 700 to $5000 \, cm^{-1}$, J. Quant. Spectros. Ra., 102, 450–463, 2006.

Huber, P. J.: Robust Statistics, Wiley, New York, 1981.

Jacquemart, D., Laraia, A., Kwabia Tchana, F., Gamache, R. R., Perrin, A., and Lacome, N: Formaldehyde around 3.5 and 5.7 µm: measurement and calculation of broadening coefficients, J. Quant. Spectrosc. Ra., 111, 1209–1222, 2010.

Jones, N. B., Riedel, K., Allan, W., Wood, S., Palmer, P. I., Chance, K., and Notholt, J.: Long-term tropospheric formaldehyde concentrations deduced from ground-based fourier transform solar infrared measurements, Atmos. Chem. Phys., 9, 7131–7142, https://doi.org/10.5194/acp-9-7131-2009, 2009.

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

Langerock, B., De Mazière, M., Hendrick, F., Vigouroux, C., Desmet, F., Dils, B., and Niemeijer, S.: Description of algorithms for co-locating and comparing gridded model data with remote-sensing observations, Geosci. Model Dev., 8, 911–921, https://doi.org/10.5194/gmd-8-911-2015, 2015.

Loyola, D. G., Gimeno García, S., Lutz, R., Argyrouli, A., Romahn, F., Spurr, R. J. D., Pedergnana, M., Doicu, A., Molina García, V., and Schüssler, O.: The operational cloud retrieval algorithms from TROPOMI on board Sentinel-5 Precursor, Atmos. Meas. Tech., 11, 409–427, https://doi.org/10.5194/amt-11-409-2018, 2018.

Martin, R. V., Parrish, D. D., Ryerson, T. B., Nicks, D. K. Jr., Chance, K., Kurosu, T. P., Jacob, D. J., Sturges, E. D., Fried, A., and Wert, B. P.: Evaluation of GOME satellite measurements of tropospheric NO_2 and HCHO using regional data from aircraft campaigns in the southeastern United States, J. Geophys. Res., 109, D24307, https://doi.org/10.1029/2004JD004869, 2004.

Meller, R. and Moortgat, G. K.: Temperature dependence of the absorption cross section of HCHO between 223 and 323 K in the wavelength range 225–375 nm, J. Geophys. Res., 105, 7089–7102, https://doi.org/10.1029/1999JD901074, 2000.

Peters, E., Wittrock, F., Großmann, K., Frieß, U., Richter, A., and Burrows, J. P.: Formaldehyde and nitrogen dioxide over the remote western Pacific Ocean: SCIAMACHY and GOME-2 validation using ship-based MAX-DOAS observations, Atmos. Chem. Phys., 12, 11179–11197, https://doi.org/10.5194/acp-12-11179-2012, 2012.

Pougatchev, N. S., Connor, B. J., and Rinsland, C. P.: Infrared measurements of the ozone vertical distribution above Kitt Peak, J. Geophys. Res., 100, 16689–16697, 1995.

Rodgers, C. D.: Inverse methods for atmospheric sounding: Theory and Practice, Series on Atmospheric, Oceanic and Planetary Physics – Vol. 2, World Scientific Publishing Co., Singapore, 2000.

Rodgers, C. D. and Connor, B. J.: Intercomparison of remote sounding instruments, J. Geophys. Res., 108, 4116–4129, 2003.

Rothman, L. S., Gordon, I. E., Babikov, Y., Barbe, A., Benner, D. C., Bernath, P. F., Birk, M., Bizzocchi, L., Boudon, V., Brown, L. R., Campargue, A., Chance, K., Cohen, E. A., Coudert, L. H., Devi, V. M., Drouin, B. J., Fayt, A., Flaud, J.-M., Gamache, R. R., Harrison, J. J., Hartmann, J.-M., Hill, C., Hodges, J. T., Jacquemart, D., Jolly, A., Lamouroux, J., Roy, R. J. L., Li, G., Long, D. A., Lyulin, O. M., Mackie, C. J., Massie, S. T., Mikhailenk, S., Müller, H. S. P., Naumenko, O. V., and Nikitin, A. V.: The HITRAN2012 molecular spectroscopic database, J. Quant. Spectrosc. Ra., 130, 4–50, https://doi.org/10.1016/j.jqsrt.2013.07.002, 2013.

Sen, P. K.: Estimates of the regression coefficient based on Kendall's tau, J. Am. Stat. Assoc., 63, 1379–1389, https://doi.org/10.2307/2285891, 1968.

Spurr, R. J. D.: LIDORT and VLIDORT: Linearized pseudo-spherical scalar and vector discrete ordinate radiative transfer models for use in remote sensing retrieval problems, in: Light Scattering Reviews, edited by: Kokhanovsky, A., 229–271, Berlin, Springer, 2008.

Stavrakou, T., Müller, J.-F., Bauwens, M., De Smedt, I., Van Roozendael, M., De Mazière, M., Vigouroux, C., Hendrick, F., George, M., Clerbaux, C., Coheur, P.-F., and Guenther, A.: How consistent are top-down hydrocarbon emissions based on formaldehyde observations from GOME-2 and OMI?, Atmos. Chem. Phys., 15, 11861–11884, https://doi.org/10.5194/acp-15-11861-2015, 2015.

Tan, W., Liu, C., Wang, S., Xing, C., Su, W., Zhang, C., Xia, C., Liu, H., Cai, Z., and Liu, J.: Tropospheric NO_2, SO_2, and HCHO over the East China Sea, using ship-based MAX-DOAS observations and comparison with OMI and OMPS satellite data, Atmos. Chem. Phys., 18, 15387–15402, https://doi.org/10.5194/acp-18-15387-2018, 2018.

Tikhonov, A.: On the solution of incorrectly stated problems and a method of regularization, Dokl. Acad. Nauk SSSR, 151, 501–504, 1963.

Veefkind, J.P., Aben, I., McMullan, K., Förster, H., De Vries, J., Otter, G., Claas, J., Eskes, H.J., De Haan, J.F., Kleipool, Q., Van Weele, M., Hasekamp, O., Hoogeveen, R., Landgraf, J., Snel, R., Tol, P., Ingmann, P., Voors, R., Kruizinga, B., Vink, R., Visser, H., and Levelt, P. F.: TROPOMI on the ESA Sentinel-5 Precursor: A GMES mission for global observations of the atmospheric composition for climate, air quality and ozone layer applications, Remote Sens. Environ., 120, 70–83, https://doi.org/10.1016/j.rse.2011.09.027, 2012.

Vigouroux, C., Hendrick, F., Stavrakou, T., Dils, B., De Smedt, I., Hermans, C., Merlaud, A., Scolas, F., Senten, C., Vanhaelewyn, G., Fally, S., Carleer, M., Metzger, J.-M., Müller, J.-F., Van Roozendael, M., and De Mazière, M.: Ground-based FTIR and MAX-DOAS observations of formaldehyde at Réunion Island and comparisons with satellite and model data, Atmos. Chem. Phys., 9, 9523–9544, https://doi.org/10.5194/acp-9-9523-2009, 2009.

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

von Clarmann, T.: Smoothing error pitfalls, Atmos. Meas. Tech., 7, 3023–3034, https://doi.org/10.5194/amt-7-3023-2014, 2014.

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, 721–750, https://doi.org/10.5194/gmd-10-721-2017, 2017.

Wittrock, F., Richter, A., Oetjen, H., Burrows, J. P., Kanakidou, M., Myriokefalitakis, S., Volkamer, R., Beirle, S., Platt, U., and Wagner, T.: Simultaneous global observations of glyoxal and formaldehyde from space, Geophys. Res. Lett., 33, L16804, https://doi.org/10.1029/2006GL026310, 2006.

Zhu, L., Jacob, D. J., Kim, P. S., Fisher, J. A., Yu, K., Travis, K. R., Mickley, L. J., Yantosca, R. M., Sulprizio, M. P., De Smedt, I., González Abad, G., Chance, K., Li, C., Ferrare, R., Fried, A., Hair, J. W., Hanisco, T. F., Richter, D., Jo Scarino, A., Walega, J., Weibring, P., and Wolfe, G. M.: Observing atmospheric formaldehyde (HCHO) from space: validation and intercomparison of six retrievals from four satellites (OMI, GOME2A, GOME2B, OMPS) with SEAC[4]RS aircraft observations over the southeast US, Atmos. Chem. Phys., 16, 13477–13490, https://doi.org/10.5194/acp-16-13477-2016, 2016.

Zhu, L., González Abad, G., Nowlan, C. R., Chan Miller, C., Chance, K., Apel, E. C., DiGangi, J. P., Fried, A., Hanisco, T. F., Hornbrook, R. S., Hu, L., Kaiser, J., Keutsch, F. N., Permar, W., St. Clair, J. M., and Wolfe, G. M.: Validation of satellite formaldehyde (HCHO) retrievals using observations from 12 aircraft campaigns, Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2019-1117, in review, 2020.