Atmospheric
Measurement
Techniques

Validation of MOPITT carbon monoxide using ground-based Fourier transform infrared spectrometer data from NDACC

Rebecca R. Buchholz[1], Merritt N. Deeter[1], Helen M. Worden[1], John Gille[1], David P. Edwards[1], James W. Hannigan[1], Nicholas B. Jones[2], Clare Paton-Walsh[2], David W. T. Griffith[2], Dan Smale[3], John Robinson[3], Kimberly Strong[4], Stephanie Conway[4], Ralf Sussmann[5], Frank Hase[6], Thomas Blumenstock[6], Emmanuel Mahieu[7], and Bavo Langerock[8]

[1]Atmospheric Chemistry Observations & Modeling, National Center for Atmospheric Research, Boulder, Colorado, USA
[2]Centre for Atmospheric Chemistry, University of Wollongong, Wollongong, Australia
[3]National Institute of Water and Atmospheric Research, Lauder, New Zealand
[4]Department of Physics, University of Toronto, Toronto, Canada
[5]Karlsruhe Institute of Technology, IMK-IFU, Garmisch-Partenkirchen, Germany
[6]Karlsruhe Institute of Technology, IMK-ASF, Karlsruhe, Germany
[7]Institute of Astrophysics and Geophysics, University of Liège, Liège, Belgium
[8]Royal Belgian Institute for Space Aeronomy (BIRA-IASB), Brussels, Belgium

Correspondence to: Rebecca R. Buchholz (buchholz@ucar.edu)

Abstract. The Measurements of Pollution in the Troposphere (MOPITT) satellite instrument provides the longest continuous dataset of carbon monoxide (CO) from space. We perform the first validation of MOPITT version 6 retrievals using total column CO measurements from ground-based remote-sensing Fourier transform infrared spectrometers (FTSs). Validation uses data recorded at 14 stations, that span a wide range of latitudes (80° N to 78° S), in the Network for the Detection of Atmospheric Composition Change (NDACC). MOPITT measurements are spatially co-located with each station, and different vertical sensitivities between instruments are accounted for by using MOPITT averaging kernels (AKs). All three MOPITT retrieval types are analyzed: thermal infrared (TIR-only), joint thermal and near infrared (TIR–NIR), and near infrared (NIR-only). Generally, MOPITT measurements overestimate CO relative to FTS measurements, but the bias is typically less than 10 %. Mean bias is 2.4 % for TIR-only, 5.1 % for TIR–NIR, and 6.5 % for NIR-only. The TIR–NIR and NIR-only products consistently produce a larger bias and lower correlation than the TIR-only. Validation performance of MOPITT for TIR-only and TIR–NIR retrievals over land or water scenes is equivalent. The four MOPITT detector element pixels are validated separately to account for their different uncertainty characteristics. Pixel 1 produces the highest standard deviation and lowest correlation for all three MOPITT products. However, for TIR-only and TIR–NIR, the error-weighted average that includes all four pixels often provides the best correlation, indicating compensating pixel biases and well-captured error characteristics. We find that MOPITT bias does not depend on latitude but rather is influenced by the proximity to rapidly changing atmospheric CO. MOPITT bias drift has been bound geographically to within ±0.5 % yr^{-1} or lower at almost all locations.

1 Introduction

Atmospheric carbon monoxide (CO) is useful for studying both transported and local sources of pollution. CO is directly emitted from incomplete combustion, such as from biomass burning and fossil fuel use. CO is also chemically produced from the oxidation of methane and volatile organic compounds. The approximate global lifetime of 2 months makes CO an ideal atmospheric constituent to track atmo-

spheric physical and chemical processes over a range of spatial scales (Edwards et al., 2006; Duncan et al., 2007).

The Measurements of Pollution in the Troposphere (MOPITT) is the longest running satellite sensor measuring atmospheric CO globally, measuring since 2000 aboard the satellite Terra from low-Earth orbit using thermal infrared (TIR). MOPITT is the only satellite instrument measuring CO in both TIR and near infrared (NIR). A long record presents an opportunity to analyze temporal changes in atmospheric CO. For example, long-term CO trends from satellite records were compared in (Worden et al., 2013). However, continued validation of the instrument is necessary to ensure that observed temporal changes are due to changes in the atmospheric state rather than changes in the instrument. Validation is performed against an independent measure of atmospheric CO over a long time period to help determine any instrument drift.

MOPITT has been extensively validated with in situ measurements at the ground and by aircraft (Deeter et al., 2010, 2014; Emmons et al., 2004, 2009). Validation and intercomparison were also completed using other satellite products (Martínez-Alonso et al., 2014; de Laat et al., 2014; George et al., 2009; Yurganov et al., 2008). However, there has been a lack of systematic validation using ground-based remote-sensing measurements of total column CO. Additionally, the recent addition of MOPITT multispectral retrievals – which provide improved lower tropospheric sensitivity due to reflected solar wavelengths – have a limited validation history compared to the TIR-only product (Worden et al., 2010). Finally, updates to the MOPITT retrieval algorithm and ongoing measurements require updated validation.

Ground-based solar Fourier transform infrared spectrometer (FTS) measurements provide a valuable source of long-term total column CO data for validating MOPITT. Many of the stations have records that predate the launch of MOPITT. Biases between stations are minimized by standardizing retrieval procedures (Rinsland et al., 2007; Hase et al., 2004). Such time series have proved useful in previous validation studies of CO measurements from space (e.g., Sussmann and Buchwitz, 2005; Clerbaux et al., 2008; de Laat et al., 2010; Kerzenmacher et al., 2012). FTS total column CO data at the Southern Hemisphere stations Lauder and Arrival Heights have been briefly compared with MOPITT values in (Morgenstern et al., 2012), who used version 4 level 3 MOPITT data and did not account for sensitivity differences between instruments.

This research presents the first systematic validation of MOPITT CO with ground-based solar FTS measurements. Section 2 introduces the MOPITT CO version 6 satellite product and ground-based measurements. Section 3 details the comparison methodology, including the smoothing of FTS measurements with MOPITT averaging kernels (AKs). Validation results are presented in Sect. 4, with an analysis at each ground station (Sect. 4.2), different surface-type scenes (Sect. 4.3), and investigation of detector element differences

(Sect. 4.4). The geographic and temporal dependence of validation results is discussed in Sect. 5.

2 Instruments and retrieval products

2.1 MOPITT satellite instrument

MOPITT measures CO with a correlation radiometer aboard the satellite Terra, completing ~ 14 Sun-synchronous polar orbits per day, flying at a nominal altitude of 705 km, with equator overpasses at $\sim 10{:}30$ and $\sim 22{:}30$ LT (local time). MOPITT is nadir-viewing with a ground resolution of 22×22 km. Each of the four indium antimonide (InSb) detector elements is cooled by a Stirling cycle cooler to 100 K. Ground coverage is achieved by cross-track scanning, producing a swath width of around 600 km. Complete Earth coverage is achieved in approximately 3 days (Drummond, 1992; Drummond and Mand, 1996; Deeter et al., 2003).

The MOPITT correlation radiometer modulates either the pressure or length of a correlation cell filled with target gas in order to determine spectral line differences. The fundamental CO infrared band (4.7 μm) is measured from terrestrial thermal infrared radiation (Drummond, 1992; Drummond et al., 2010; Drummond and Mand, 1996; Deeter et al., 2003). More recently, incorporating the overtone band (2.3 μm) from reflected solar radiation in the near infrared provides multispectral retrievals (Worden et al., 2010; Deeter et al., 2009; Deeter, 2013). The additional NIR signal enhances retrieval information, particularly for near-surface concentrations. However, enhanced retrievals from solar reflection are only available during the day and over land (Deeter et al., 2010). Here, we use the three retrieval products available from MOPITT version 6 (V6; Deeter et al., 2014): thermal infrared only (TIR-only; MOPITT Science Team, 2013), joint thermal and near infrared (TIR–NIR; MOPITT Science Team, 2013), and near infrared only (NIR-only; MOPITT Science Team, 2013). MOPITT products are publicly available through several repositories linked via http://terra.nasa.gov/about/terra-instruments/mopitt.

The retrieval process for MOPITT follows the maximum a posteriori method (Rodgers, 2000; Pan et al., 1998), combined with a fast radiative transfer model of the atmosphere to invert the radiometric signals (Edwards et al., 1999). Cloud-free scenes are used in retrievals (Warner et al., 2001), apart from the case where low warm clouds do not affect the radiative balance. A priori profiles are combined with radiance measurements and meteorological parameters to statistically determine a maximum a posteriori solution. The V6 monthly a priori profiles are averages from a 2000–2009 climatological run of the Community Atmosphere Model with Chemistry (CAM-Chem) gridded at 1° resolution (Lamarque et al., 2012). Meteorological parameters such as temperature and water vapor profiles are from MERRA (Modern-Era Retrospective Analysis for Research Applications, http:

//gmao.gsfc.nasa.gov/merra/) reanalysis (Deeter et al., 2014; Deeter, 2013; Rienecker et al., 2011). The V6 retrievals improve upon version 5 by eliminating a geolocation error, using an updated climatology of a priori profiles, using meteorological data with higher spatial and temporal resolution, and updating radiance correction factors. A full description of the MOPITT V6 products can be found in (Deeter et al., 2014).

Profiles of volume mixing ratio (VMR) are retrieved on 10 vertical levels that include a floating surface pressure plus nine equally spaced pressure levels from 900 to 100 hPa (Deeter et al., 2010). Profile values are reported on levels with uniform vertical weighting above each level (Deeter, 2011). Trace gas values are initially retrieved as $\log_{10}$(VMR), which means averaging kernel values are associated with $\log_{10}$(VMR) (Deeter et al., 2003, 2010). AKs are produced for every measurement and describe the sensitivity of retrieved CO values to the true state vector (Deeter et al., 2003, 2010). Reported total column CO values are calculated from integrating the VMR profiles.

2.2 Ground-based Fourier transform infrared spectrometers

We use measurements of CO from ground-based FTSs to validate MOPITT retrievals. Trace gas retrievals from solar absorption in the mid-infrared region (MIR, 2–14 µm) contribute to the Network for the Detection of Atmospheric Composition Change (NDACC, http://www.ndacc.org/; Hannigan et al., 2009). We use publicly available data from 14 NDACC stations, which span the globe from 80° N to 78° S. Figure 1 indicates the location of each NDACC station, and Table 1 provides station-specific information, including location, measurement period, instrument type, and associated site reference. Records from almost all stations predate the launch of MOPITT. The spectrometers used are the Bomem model DA8, Bruker IFS 120HR, Bruker IFS 120M, or Bruker IFS 125HR. In some cases, instruments have been updated during the station lifetime. Spectral resolution is typically 0.004 cm^{-1} or higher.

A solar tracker is coupled with the spectrometer to track the solar disk for measurements during clear-sky conditions. The FTS instrument comprises a Michelson interferometer combined with either a KBr or CaF$_2$ beamsplitter. Detectors are liquid-nitrogen-cooled indium antimonide (InSb) and mercury cadmium telluride (MCT; Bacsik et al., 2004). Optical bandpass filters are used to improve the signal-to-noise ratio in the CO spectral region. The simultaneous use of spectral microwindows during trace gas retrieval further improves the signal-to-noise ratio by minimizing the influence of interfering species. Microwindow regions used in the standard CO retrieval strategy are 2057.7–2058.0, 2069.56–2069.76, and 2157.50–2159.15 cm^{-1} (Rinsland et al., 2007), with some minor differences between stations. More information can be found at http://www2.acom.ucar.edu/irwg.

The most recent retrievals available in 2016 are used for this validation study. The retrieval code may be SFIT4 (10 stations), SFIT2 (2 stations; Pougatchev et al., 1995), or PROFFIT (2 stations; Hase, 2000). The SFIT and PROFFIT retrieval codes were intercompared by (Hase et al., 2004), who show retrievals and AKs are consistent to within 1 % or below. A method based on optimal estimation is used to retrieve trace gas information (Rodgers, 2000). The retrieval combines a radiative transfer model with spectroscopic line parameters from the HITRAN database (Rothman et al., 2009), a priori profiles from the Whole Atmosphere Community Climate Model (WACCM, Garcia et al., 2007), information about the instrument line shape (unique at each station), meteorological parameters such as pressure and temperature from the National Centers for Environmental Prediction (NCEP), and information about the atmospheric water content. Synthetic spectra are produced and parameters are iteratively improved to minimize the cost function. Interfering gases are simultaneously adjusted in the retrieval process.

Trace gases are retrieved on a vertical grid of ∼ 40 layers, with some site-specific adjustments to spacing in the lower atmosphere near the observation site altitude. The retrieved profile is reported as VMR, with each value describing a pressure-weighted average value at the midpoint within a layer for SFIT or level values in the case of PROFFIT. AK matrices are associated with VMR and are reported for every measurement. Column AKs are also reported and describe the response of column values to a change in the partial column profile.

3 Comparison methodology

We analyze all three MOPITT retrieval products: TIR-only, TIR–NIR, and NIR-only (Deeter et al., 2014). Although MOPITT has been measuring CO since March 2000, one of the two optical benches became non-operational in May 2001 as a result of cooler failure. The period before the optical bench loss is known as Phase 1 and from August 2001 onward is known as Phase 2 (Deeter et al., 2004; Emmons et al., 2004). While instrument changes between phases are accounted for in the forward model and retrieval algorithm, a small step change remains between Phase 1 and 2 retrievals. Consequently, we focus on the validation of Phase 2.

Recently, a calibration issue was found with the NIR radiances that affects retrievals after February 2012. Therefore, TIR–NIR and NIR-only are validated between August 2001 and February 2012. In contrast, TIR-only is validated between August 2001 and the station-specific end dates, determined by the available FTS data at each station (see Table 1).

Validation is performed for a range of conditions to assess whether parameters that are known to affect the MOPITT sensitivity and AKs will affect the validation results. Specifically, separate validation is performed for the four MOPITT

Table 1. Information about the 14 NDACC ground-based remote-sensing FTS sites. Locations are ordered by latitude.

FTS station name (3-letter acronym)	Alt. (m a.s.l.)	Lat.	Long.	Observation period	Instrument[1]	Code[2]	Site reference
Northern Hemisphere							
Eureka, Canada (EUR)	610	80.05° N	82.42° W	2006–2014	Br125	SFIT4	(Batchelor et al., 2009)
Ny-Ålesund, Norway (NYA)	15	78.92° N	11.93° E	1992–2014	Br120	SFIT4	(Notholt et al., 1993)
Thule, Greenland (THU)	225	76.53° N	68.74° W	1999–2014	Br120M	SFIT4	(Hannigan et al., 2009)
Kiruna, Sweden (KIR)	419	67.84° N	20.41° E	1996–2007	Br120	PRO	(Blumenstock et al., 2006)
				2007–2012	Br125		
Bremen, Germany (BRE)	27	53.10° N	8.85° E	2003–2014	Br125	SFIT4	(Velazco et al., 2007)
Zugspitze, Germany (ZUG)	2964	47.42° N	10.98° E	1995–2014	Br125	SFIT2	(Sussmann and Schäfer, 1997)
Jungfraujoch, Switzerland (JFJ)	3580	46.55° N	7.98° E	1989–2012	Br120	SFIT2	(Mahieu et al., 1997)
Toronto, Canada (TAO)	174	43.66° N	79.40° W	2002–2014	BoDA8	SFIT4	(Wiacek et al., 2007)
Izaña, Spain (IZA)	2367	28.30° N	16.48° W	1999–2005	Br120M	PRO	(Schneider et al., 2005)
				2005–2014	Br125		
Mauna Loa, USA (MLO)	3397	19.54° N	155.58° W	1995–2007	Br120M	SFIT4	(David et al., 1993)
				2007–2014	Br125		
Southern Hemisphere							
La Réunion[3], France (LRN)	10	20.90° S	55.50° E	2004–2011	Br120M	SFIT4	(Senten et al., 2008)
				2011–2013	Br125		
Wollongong, Australia (WOL)	30	34.41° S	150.88° E	1996–2008	BoDA8	SFIT4	(Paton-Walsh et al., 2005)
				2008–2013	Br125		
Lauder, New Zealand (LAU)	370	45.04° S	169.68° E	2001–2014	Br120	SFIT4	(Morgenstern et al., 2012)
Arrival Heights, Antarctica (AHS)	250	77.82° S	166.65° E	1997–2014	Br120M	SFIT4	(Zeng et al., 2012)

[1] BoDA8 = Bomem DA8, Br120 = Bruker 120HR, Br120M = Bruker 120M, Br125 = Bruker 125HR. [2] Code versions are as follows: SFIT4 is version V0.9.4.4, except for at Ny-Ålesund and Bremen, where it is version V0.9.4.3_BF_MP; SFIT2 at Jungfraujoch is v3.91 and at Zugspitze V3.90i; PRO = PROFFIT96. [3] Data investigated are only from the St. Denis station and not from Maido.
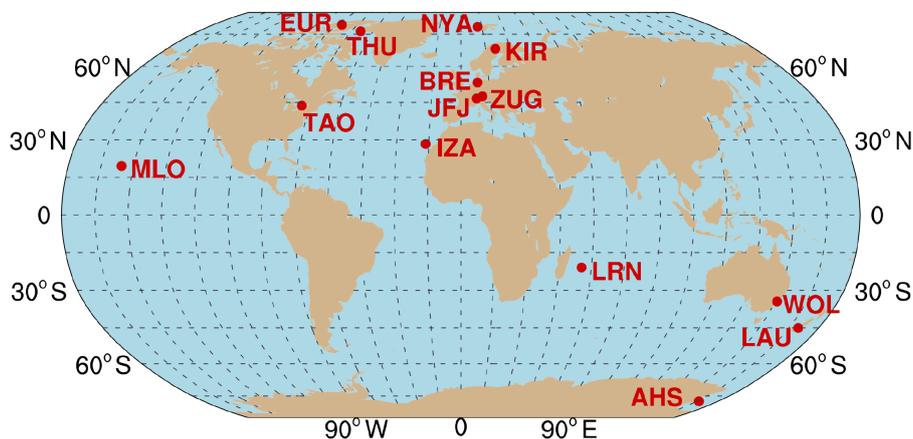
Figure 1. Location of the 14 NDACC ground-based remote-sensing FTS sites used in this study. The three-letter acronyms correspond to the information in Table 1.

detector elements, for land scenes versus water scenes, and over a range of latitudes.

In order to accurately compare measurements between instruments, equivalent air masses must be compared. This involves co-locating measurements in time and space and accounting for the relative sensitivity of each instrument.

3.1 Co-location criteria

Temporal co-location is defined as comparing daytime measurements from MOPITT with FTS measurements retrieved within the same day as the MOPITT overpass. All FTS measurements within the same day as the MOPITT overpass time of ~ 10:30 LT are considered. While the MOPITT overpass also occurs at ~ 22:30 LT, the daytime-only MOPITT measurements are used in order to include enhanced infor-

mation from the reflected solar NIR. Another constraint is that ground-based instruments only measure during daytime clear-sky conditions.

MOPITT retrievals are spatially co-located with the FTS by selecting MOPITT data within a 1° radius around each ground station, a distance criterion that is suggested by (Sparling and Bacmeister, 2001). One degree has been found adequate in other satellite validation studies (Yurganov et al., 2008; Kerzenmacher et al., 2012) and falls within the range of previous validation of the V6 MOPITT product, which used radii of 0.5° (against National Oceanic and Atmospheric Administration, NOAA, profiles) and 2° (against HIAPER Pole-to-Pole Observations, HIPPO; Deeter et al., 2014).

3.2 Averaging

Prior to validation, MOPITT retrievals that are co-located with FTS measurements are averaged and inversely weighted by the square of the relative retrieval measurement error. Thus, one MOPITT average is compared with several FTS measurements. There are several advantages of averaging the MOPITT data. Error-weighted averaging improves the signal-to-noise ratio and reduces the random uncertainty in the MOPITT data, while any systematic bias remains, allowing diagnosis of the MOPITT bias. Averaging will also reduce the co-location errors associated with non-coincidence of air masses, thereby reducing the sampling bias. Combining satellite measurements within the 1° radius criterion satisfies a compromise between reducing the effects of random retrieval noise and minimizing the spatial dilution of measurements through the use of a small radius. We use an approximation method during validation that also calculates error-weighted averaged MOPITT AK matrices and a priori profiles. This approximation method improves computational efficiency, reducing the number of comparisons at some stations from ∼ 70 000 to ∼ 3000. Instrument error is combined in quadrature. Depending upon the experiment, averages are restricted to include land-only scenes, water-only scenes, or specific detector element pixels.

The approximation method is evaluated in Appendix B. We find that for most stations, using mean AKs and a priori profiles produces equivalent results to point-wise application of AKs followed by averaging. Care must be taken when using validation results from three stations located in or close to mountainous regions that consequently experience large variability in surface pressure (Zugspitze, Jungfraujoch and Lauder), as mean bias may be incorrectly estimated by 1 to 3 percentage points (pp). However, we found no substantial effect of these stations on the all-station mean validation values; therefore, they are included in the reported all-station means.

Temporal averaging is not performed on the FTS measurements. In summary, the MOPITT daily spatial averages are compared separately with each FTS measurement in the same day.

3.3 Vertical re-gridding

Vertical grids between instruments are different in terms of resolution as well as the retrieved surface altitude. Specifically, FTS measurements are retrieved on a finer vertical grid than MOPITT. The FTS measurements must be re-gridded to the MOPITT vertical levels for two reasons: (1) measurements must describe the total column over the same altitude range in order to compare equivalent atmospheric amounts and (2) the FTS profile will be smoothed by the MOPITT AKs, which are reported on MOPITT vertical layers.

MOPITT CO profile values describe the average VMR within the layer above the reported level (Deeter et al., 2013). In contrast, FTS values are reported on layer midpoints and describe the average VMR within that layer (using SFIT) or are level values (when using PROFFIT). FTS profiles are re-gridded in a manner that is independent of the FTS profile definition, assuming hydrostatic equilibrium. We first interpolate the FTS profile in logP space to an ultrafine grid of 100 levels per MOPITT layer. The VMR values are then averaged over each set of 100 ultrafine levels to produce an average within each MOPITT layer. Resulting FTS averages are associated with levels, which is the same definition as for the MOPITT profile.

During re-gridding, two situations are accounted for: either the reported FTS surface pressure is larger than MOPITT's or the FTS surface pressure is smaller than MOPITT's. In the first case, if the FTS surface pressure is larger than MOPITT, any FTS layers below the MOPITT surface layer are not used when averaging the 100 ultrafine levels. In this case, if the MOPITT surface layer occurs at an altitude above 900 hPa, the MOPITT profile will have less than 10 vertical layers, as will the re-gridded FTS profile.

Alternatively, where the surface pressure for FTS is smaller than for MOPITT, a choice must be made on whether to remove some of the MOPITT profile to create a partial column or to fill in the FTS profile, in order to compare values over the same air mass range. These situations occur mainly for stations near highly varying terrain or at high altitude. We determine the effect of profile truncation and profile filling in Appendix C and conclude that the profile filling technique is the more suitable option, estimated to affect the bias calculations by generally less than 1 pp. Profile filling is achieved with a similar method to (Kerzenmacher et al., 2012), altered to use FTS a priori profiles rather than MOPITT a priori profiles. The FTS a priori profiles are re-gridded to the MOPITT vertical grid before determining a scaling factor, with a stipulation that surface levels of the FTS a priori and MOPITT profiles are within 20 hPa of each other (50 hPa at EUR, WOL, and AHS). Otherwise, comparisons are not made. Scaling factors determined for the first non-missing
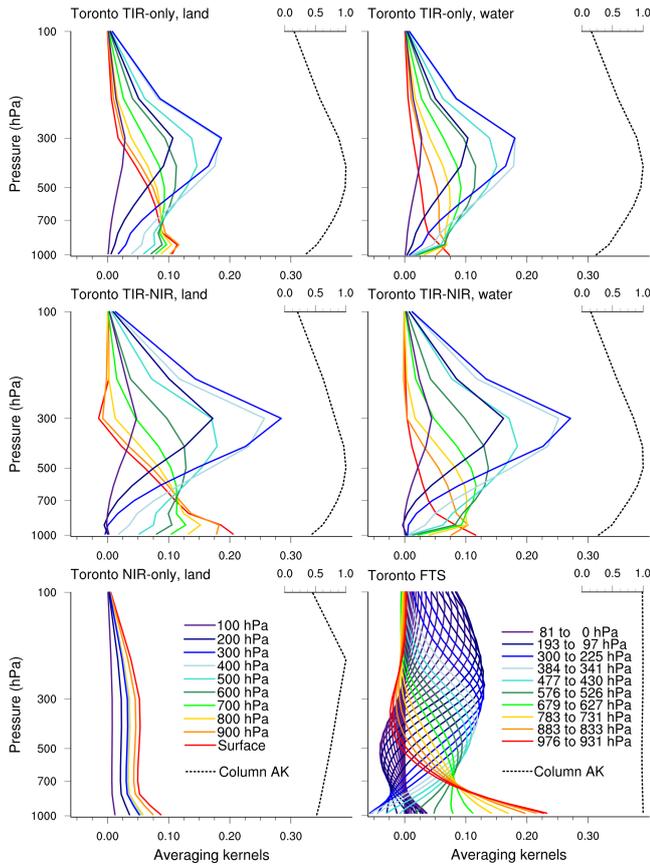
Figure 2. Mean AKs for MOPITT (top two rows and bottom left) and FTS (bottom right) at Toronto. Full averaging kernel matrices are shown with solid lines. The corresponding total column averaging kernel is shown with a dashed line. MOPITT column AKs are normalized.

FTS layer are used to fill missing layers by multiplying the associated a priori layers.

3.4 Averaging kernel smoothing

MOPITT AKs are used to smooth the re-gridded FTS profiles in order to account for sensitivity differences between instruments. The total column AKs of FTS are near unity over the altitude range covered by MOPITT (e.g., Toronto in Fig. 2, bottom right plot), indicating relatively uniform sensitivity to the true atmospheric state and little inclusion of the a priori information in the retrieved values. In contrast, the column AKs of MOPITT peak in the free troposphere and show overall less sensitivity than FTS, including more of the a priori information in the retrievals, particularly for the lower altitude levels. (Rodgers and Connor, 2003) show that when intercomparing instruments, if one instrument possesses a lower dependence on the a priori values and more information than the other, it can be used as a closer representation of the true atmospheric state. Therefore, we take the FTS retrievals to be "atmospheric truth" and smooth to MO-

PITT retrieval space. Specifically, the vertically re-gridded FTS profile is smoothed by MOPITT AK matrices.

The MOPITT AK matrices are applied following (Rodgers and Connor, 2003), modified for log(VMR):

$$\log_{10}(x_{\text{smooth},i}) = \log_{10}(x_{\text{ap},i}) + \sum_{j=1}^{n} A_{ij}$$
$$\left(\log_{10}(x_{\text{FTS},i}) - \log_{10}(x_{\text{ap},j})\right), \quad (1)$$

where n is the number of vertical layers, $\boldsymbol{x}_{\text{smooth}} = \{x_{\text{smooth},i} \mid i = 1 \ldots n\}$ is the required smoothed VMR profile, $\boldsymbol{x}_{\text{ap}} = \{x_{\text{ap},i} \mid i = 1 \ldots n\}$ is the MOPITT a priori VMR profile, $\boldsymbol{A} = \{A_{ij} \mid i = 1 \ldots n, j = 1 \ldots n\}$ is the MOPITT AK matrix, and $\boldsymbol{x}_{\text{FTS}} = \{x_{\text{FTS},i} \mid i = 1 \ldots n\}$ is the FTS re-gridded VMR profile. $\boldsymbol{A}$ has been calculated based on $\log_{10}(\text{VMR})$ and therefore must be applied to a profile of $\log_{10}(\text{VMR})$. Differences between AKs matrices for the three MOPITT products are visualized in Fig. 2. AKs are further discussed in Sect. 4.1 and 4.3.

The resulting smoothed profile of $\log_{10}(\text{VMR})$ is converted to a VMR profile. Equations (2) and (3) describe the relationship between the smoothed VMR values and the terms in Eq. (1).

$$x_{\text{smooth},i} = x_{\text{ap},i} \cdot 10^{\sum_{j=1}^{n} A_{ij}(\log_{10}(x_{\text{FTS},i}) - \log_{10}(x_{\text{ap},j}))} \quad (2)$$
$$= x_{\text{ap},i} \cdot \prod_{j=1}^{n} \left(\frac{x_{\text{FTS},i}}{x_{\text{ap},j}}\right)^{A_{ij}} \quad (3)$$

Smoothed FTS total column values (c_{smooth}) are calculated from $\boldsymbol{x}_{\text{smooth}}$ using pressure weighted integration with MOPITT retrieval pressure widths (Eq. 4). The smoothed FTS column value is calculated over the same altitude range as MOPITT and represents what MOPITT would have retrieved, assuming the FTS measurement describes the true atmosphere.

$$c_{\text{smooth}} = \alpha \sum_{i=1}^{n} \left(x_{\text{smooth},i}\right) \Delta p_i, \quad (4)$$

where α is the conversion factor between VMR and column amount and Δp_i is the pressure width of layer i.

The MOPITT retrieved column (c_{M}) is then validated against the smoothed FTS column, for example, by calculating the mean bias (Eq. 5).

$$\text{station mean bias} = \sum_{k=1}^{m} \frac{(c_{\text{M},k} - c_{\text{smooth},k})}{m}, \quad (5)$$

where m is the number of comparisons at an NDACC station.
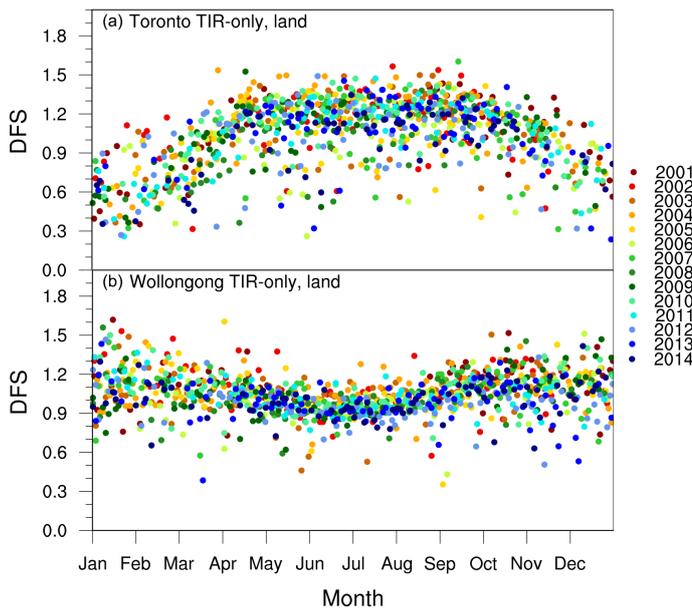
Figure 3. Seasonality of daily mean MOPITT DFS (TIR-only) within a 1° radius at the Northern Hemisphere station of Toronto (**a**) and Southern Hemisphere station of Wollongong (**b**).

4 Validation results

4.1 Information content

The information content of each instrument is described by degrees of freedom for signal (DFS). DFS are determined from the trace of the AK matrix, which is influenced by instrumental and geophysical parameters (Rodgers, 2000; Deeter et al., 2015). DFS are calculated from the AK matrices for each MOPITT and FTS retrieval.

Median DFS for each instrument at each station is recorded in Table 2. MOPITT median DFS are below 2 at all stations. The joint TIR–NIR product consistently retrieves more information than the TIR-only product, and the NIR-only product shows very low DFS, although some information is still present. In comparison, the FTS measurements retrieve more information than MOPITT at all stations. Median DFS for FTS is generally above 2 (except at La Réunion, Zugspitze, Bremen, and Ny-Ålesund). Higher information content in the ground-based measurements relative to MOPITT supports our choice to smooth the FTS measurements by the MOPITT AK.

4.2 Total column validation at each station

MOPITT is evaluated against smoothed FTS values using correlation and bias analysis. Validation results at each station are summarized in Table 3a, b, and c. Results are for daytime land scenes, except for three stations where MOPITT land scenes were sparse and consequently use water scenes (IZA, MLO, and LRN). Example correlation plots for

Table 2. Information content described by median DFS for different MOPITT products and FTS measurements at each station. DFS is calculated over approximately the same altitude range for each instrument from the surface to approximately 26 hPa. DFS are calculated for overlapping time periods between instrument; thus, the number of retrievals correspond to the #obs line in Tables 3a, b, c, and 4. DFS for the FTS are almost identical station-wise for all comparison ranges, so only the results for TIR-only overlap are shown.

Site	MOPITT TIR-only		MOPITT TIR–NIR		MOPITT NIR-only	FTS
	land	water	land	water		
EUR	0.66	0.65	1.18	0.96	0.37	2.38
NYA	0.62	0.89	1.07	1.23	0.17	1.65
THU	0.49	0.76	1.05	1.05	0.41	2.70
KIR	0.91	●	1.35	●	0.45	2.57
BRE	1.25	~	1.72	~	0.48	1.98
ZUG	1.07	●	1.50	●	0.32	1.77
JFJ	1.10	●	1.55	●	0.30	2.15
TAO	1.16	0.94	1.63	1.34	0.31	2.92
IZA	~	1.18	~	1.57	~	2.62
MLO	~	1.25	~	1.70	~	2.53
LRN	~	1.28	~	1.67	~	1.90
WOL	1.04	1.16	1.47	1.58	0.09	2.54
LAU	0.91	~	1.30	~	0.20	2.81
AHS	0.28	0.44	0.76	0.82	0.14	2.86
Mean	0.86	0.95	1.32	1.32	0.29	2.22

● No pixels of that surface type were available for analysis. ~ indicates where the number of pixels of this surface type is less than 60 % of the number of pixels of the other surface type.

each MOPITT product are shown for Toronto and Wollongong in Fig. 4. Comparison plots for all stations are found in Appendix A.

MOPITT is generally biased high relative to the FTS by a few percent. Overall, the TIR-only product performs the best, followed by the joint TIR–NIR and then NIR-only. Mean station biases are always less than 10 % for TIR-only, with an average bias across all stations of 2.4 %. Biases are generally less than 10 % for TIR–NIR (apart from at Ny-Ålesund, Bremen, and Lauder), producing an all-station mean bias of 5.1 %. NIR-only biases are less than 10 % (except at Ny-Ålesund and Lauder), with an all-station mean bias of 6.5 %. The standard deviation is always larger than bias, except for the NIR-only product at Lauder and the TIR–NIR product at Lauder and Bremen. Correlation values are generally the highest for the TIR-only product ($\bar{r}$: 0.83) compared to TIR–NIR ($\bar{r}$: 0.78) or NIR-only ($\bar{r}$: 0.77). Average correlation values found here are lower by about 0.1 compared with previous validation of the MOPITT version 6 products against NOAA column values, while bias is equivalent (Deeter et al., 2014).

Instrument sensitivity varies with season, which is reflected in the DFS seasonal variability. An example of the range of seasonal variability is shown by the TIR-only DFS

Table 3. (a) Site-wise validation results at Northern Hemisphere high-latitude stations. Mean bias (MOPITT–FTS) is expressed in units of 10^{17} molec cm^{-2} as well as percent relative to FTS values; units of drift are 10^{17} molec cm^{-2} yr^{-1} or % yr^{-1}, shown with standard error. Bold text indicates significant drift where $p < 0.01$, determined from a two-tailed t test. **(b)** Same as **(a)** but for Northern Hemisphere low- and mid-latitude sites. Results are for land pixels, except where the majority of pixels are water (w). **(c)** Same as Table **(b)** but for Southern Hemisphere sites.

(a) Version		EUR	NYA	THU	KIR		
TIR-only	bias	0.13	0.62	−0.51	0.04		
	% bias	0.70	3.40	−2.71	0.20		
	% SD	8.66	7.41	7.45	6.51		
	r	0.72	0.81	0.75	0.82		
	drift	**0.055 ± 0.021**	−0.015 ± 0.015	**0.044 ± 0.010**	−0.015 ± 0.015		
	drift (% yr^{-1})	**0.30 ± 0.11**	−0.08 ± 0.08	**0.23 ± 0.06**	−0.07 ± 0.07		
	# obs	1080	580	1703	692		
TIR–NIR	bias	0.58	2.37	0.69	0.80		
	% bias	3.06	12.86	3.56	3.99		
	% SD	9.47	17.94	13.46	7.97		
	r	0.88	0.61	0.64	0.80		
	drift	0.038 ± 0.043	**−0.135 ± 0.050**	**0.180 ± 0.025**	**−0.055 ± 0.021**		
	drift (% yr^{-1})	0.20 ± 0.23	**−0.73 ± 0.27**	**0.94 ± 0.13**	**−0.27 ± 0.10**		
	# obs	880	482	1395	643		
NIR-only	bias	0.83	2.09	1.05	1.36		
	% bias	4.36	11.19	5.48	7.08		
	% SD	7.37	17.05	10.50	8.86		
	r	0.92	0.59	0.77	0.78		
	drift	**0.123 ± 0.033**	**−0.225 ± 0.047**	**0.092 ± 0.019**	−0.055 ± 0.022		
	drift (% yr^{-1})	**0.65 ± 0.17**	**−1.21 ± 0.25**	**0.48 ± 0.10**	−0.28 ± 0.12		
	# obs	889	473	1392	643		
(b) Version		BRE	ZUG	JFJ	TAO	IZA (w)	MLO (w)
TIR-only	bias	1.83	−0.04	−0.41	1.55	0.12	−0.37
	% bias	8.56	−0.20	−2.20	7.16	0.67	−2.14
	% SD	9.26	6.47	7.72	9.33	6.20	5.88
	r	0.76	0.93	0.88	0.76	0.90	0.94
	drift	**−0.206 ± 0.034**	0.009 ± 0.005	**0.025 ± 0.009**	−0.042 ± 0.021	**−0.034 ± 0.009**	−0.024 ± 0.012
	drift (% yr^{-1})	**−0.96 ± 0.16**	0.05 ± 0.03	**0.14 ± 0.05**	−0.19 ± 0.09	**−0.19 ± 0.05**	−0.14 ± 0.07
	# obs	310	5156	1566	730	942	482
TIR–NIR	bias	2.49	0.29	−0.21	1.96	0.41	−0.33
	% bias	11.57	1.67	−1.10	9.09	2.24	−1.91
	% SD	11.16	8.55	10.36	10.74	6.39	6.83
	r	0.67	0.90	0.82	0.76	0.90	0.92
	drift	**−0.291 ± 0.073**	**−0.46 ± 0.008**	**−0.062 ± 0.018**	−0.009 ± 0.037	**−0.070 ± 0.016**	−0.080 ± 0.032
	drift (% yr^{-1})	**−1.35 ± 0.34**	**−0.27 ± 0.04**	**−0.33 ± 0.10**	−0.04 ± 0.17	**−0.39 ± 0.09**	−0.47 ± 0.18
	# obs	241	4197	1229	512	599	111
NIR-only	bias	1.39	0.94	0.72	1.07	–	–
	% bias	6.56	5.39	4.05	4.75	–	–
	% SD	9.53	12.65	9.56	6.39	–	–
	r	0.70	0.70	0.78	0.72	–	–
	drift	−0.064 ± 0.063	−0.018 ± 0.012	**−0.096 ± 0.016**	0.003 ± 0.023	–	–
	drift (% yr^{-1})	−0.30 ± 0.30	−0.10 ± 0.07	**−0.54 ± 0.09**	0.01 ± 0.10	–	–
	# obs	241	4274	1242	512	–	–
(c) Version		LRN (w)	WOL	LAU	AHS		
TIR−only	bias	0.53	0.80	0.72	−0.12		
	% bias	3.63	6.18	7.11	−1.30		
	% SD	6.57	9.86	7.16	4.49		
	r	0.93	0.86	0.93	0.96		
	drift	−0.016 ± 0.016	**0.020 ± 0.005**	**0.017 ± 0.005**	−0.003 ± 0.004		
	drift (% yr^{-1})	−0.11 ± 0.11	**0.15 ± 0.04**	**0.16 ± 0.05**	−0.04 ± 0.05		
	# obs	503	5729	924	463		
TIR–NIR	bias	0.58	1.19	1.16	0.57		
	% bias	3.94	9.07	11.24	6.65		
	% SD	7.64	12.58	9.10	14.07		
	r	0.93	0.81	0.93	0.79		
	drift	−0.055 ± 0.026	**−0.033 ± 0.009**	−0.009 ± 0.014	−0.033 ± 0.023		
	drift (% yr^{-1})	−0.37 ± 0.17	**−0.25 ± 0.07**	−0.09 ± 0.13	−0.39 ± 0.27		
	# obs	376	3994	569	291		
NIR-only	bias	–	0.43	1.10	0.84		
	% bias	–	3.14	10.30	9.49		
	% SD	–	4.19	8.26	11.89		
	r	–	0.93	0.86	0.76		
	drift	–	−0.007 ± 0.003	−0.006 ± 0.013	−0.013 ± 0.020		
	drift (% yr^{-1})	–	−0.05 ± 0.02	−0.06 ± 0.12	−0.14 ± 0.23		
	# obs	–	4002	570	290		

Note: (w) denotes sites where water pixels were more representative, and therefore there are no results for NIR retrievals, indicated by –.
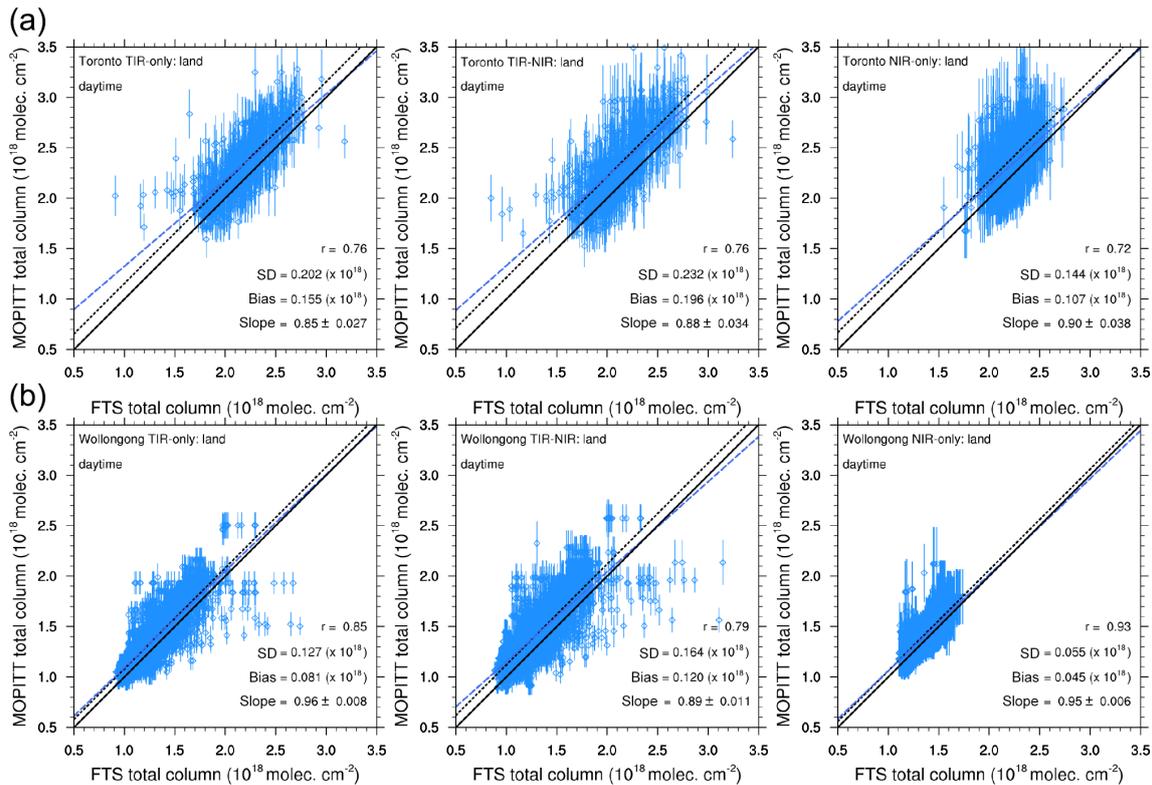
**Figure 4.** Example of the correlation plots at Toronto **(a)** and Wollongong **(b)**. Daytime, land-only pixels are shown for each retrieval type TIR-only, TIR–NIR, and NIR-only. The black dotted line indicates mean bias and the blue dashed line indicates correlation regression line. For reference, the 1 : 1 line is indicated in solid black. Error bars are the errors for MOPITT values, combined in quadrature from the pixels within a 1° radius of the station. Column amounts are molecules cm^{-2}.

Table 4. Validation results for MOPITT water pixels for V6 thermal and joint products at stations which had approximately equal representation by water and land pixels. Values for land validation are in Table 3a, b, and c. Mean bias is expressed as percent relative to FTS values. Bold text indicates significant drift where $p < 0.01$, determined from a two-tailed t test.

Version		EUR	NYA	THU	TAO	WOL	AHS
TIR-only water	% bias	0.30	4.11	−2.97	8.70	7.15	1.38
	% SD	8.22	6.87	8.38	12.37	11.15	5.60
	r	0.78	0.88	0.69	0.62	0.81	0.95
	drift (% yr^{-1})	**−0.87 ± 0.13**	0.02 ± 0.08	**0.30 ± 0.08**	−0.27 ± 0.16	**−0.22 ± 0.05**	0.05 ± 0.07
	# obs	700	589	765	519	4832	347
TIR–NIR water	% bias	2.82	4.80	−3.36	9.56	10.34	2.28
	% SD	12.77	8.17	10.65	13.63	14.50	9.93
	r	0.62	0.84	0.59	0.58	0.74	0.88
	drift (% yr^{-1})	**−1.07 ± 0.35**	−0.19 ± 0.12	**0.71 ± 0.14**	−0.14 ± 0.26	**−0.37 ± 0.09**	0.20 ± 0.22
	# obs	580	466	609	375	3356	229

time series at Toronto and Wollongong (Fig. 3). The question arises whether seasonal sensitivity differences are significant enough to affect validation results. We conducted seasonal validation at each station and found the maximum difference in station-wise bias between seasons was on average 4.5, 6.0, and 4.2 pp for TIR-only, TIR–NIR, and NIR-only respectively. Seasonal variation in bias is below the all-station aver-

age standard deviation for each product: 7.8 (TIR-only), 10.8 (TIR–NIR), and 9.7 % (NIR-only). We conclude that there is no significant seasonally dependent bias for MOPITT.

4.3 Surface-type specific validation

MOPITT classifies pixel surface type as land, water, or mixed. Different surface types have the potential to affect validation results by influencing MOPITT retrievals. Larger variability in surface height over land, combined with emissivity and albedo differences, results in greater geophysical noise relative to water scenes (Deeter et al., 2011). Additionally, thermal contrast between skin surface and the overlying air can affect MOPITT sensitivity to measuring CO. For instance, water scenes have lower thermal contrast, which is where skin surface and overlying air temperatures are similar. MOPITT has difficulty viewing the surface in low thermal contrast scenes and in these cases has better sensitivity to CO in the free troposphere. Consequently, water scenes tend to be sensitive to the free troposphere, while land scenes include more information from the lower troposphere (Worden et al., 2010).

AKs reflect the retrieval differences between surface types (Fig. 2). For example, when comparing the mean MOPITT AK matrices at Toronto, the TIR-only land AK shows increased sensitivity around 900 hPa relative to the water AK, as a result of improved thermal contrast. The TIR–NIR land AK shows even greater sensitivity at around 900 hPa relative to both the TIR-only land AK and the TIR–NIR water AK, due to the combination of improved thermal contrast with extra information from the NIR signal. While the TIR–NIR water product does not include reflected solar information, AKs are different between the TIR-only and TIR–NIR over water scenes, due to retrieval differences. Specifically, the joint product attributes less weight to the a priori profiles in the retrieval process, with a cost of higher variability (Deeter et al., 2011).

To assess the effect of different surface types in the MOPITT retrievals, validation is performed separately for land or water scenes. At some stations, MOPITT provides a significant number of pixels within the 1° radius for only one surface type: La Réunion, Mauna Loa, and Izaña are only represented by water pixels, while Bremen, Jungfraujoch, Zugspitze, Kiruna, and Lauder are only represented by land pixels. Consequently, comparison between land or water pixels is completed where stations are represented approximately equally by water and land surface types. At each station, the error-weighted average within a 1° radius is calculated with either all land or all water pixels. Mixed surface-type pixels are discarded.

Table 4 summarizes validation results over water scenes for TIR-only and joint TIR–NIR products. Land scene validation results were presented in Table 3a, b, and c. While the TIR–NIR product over water scenes does not include NIR information, differences arise compared to the TIR-only water scenes due to differences in the retrieval algorithm as discussed above. Validation statistics over water show a pattern consistent with validation over land, i.e., lower correlation, higher bias, and higher standard deviation occurs for the joint

Table 5. Validation results for each pixel over land scenes at Lauder. Bias and standard deviation (SD) units are 10^{17} molec cm^{-2}.

Version		Pixel 1	Pixel 2	Pixel 3	Pixel 4	Pixels 2–4	All
TIR-only	bias	0.58	0.70	0.89	0.52	0.74	0.72
	SD	1.22	0.99	0.76	0.84	0.76	0.73
	r	0.80	0.87	0.92	0.91	0.93	0.93
TIR–NIR	bias	1.05	1.34	1.36	0.83	1.18	1.16
	SD	1.60	1.11	1.00	1.06	1.01	0.94
	r	0.80	0.90	0.90	0.89	0.91	0.93
NIR-only	bias	1.71	1.18	0.80	0.72	0.89	1.10
	SD	1.91	1.06	0.86	1.10	0.70	0.89
	r	0.66	0.84	0.85	0.71	0.89	0.86

product compared to TIR-only. Overall, the choice of surface type has very little effect on validation statistics for the sites investigated here.

4.4 Pixel-wise validation

The MOPITT detectors are comprised of four detector elements, resulting in four pixels each with a nadir ground size of 22×22 km. Instrument-only noise is determined for each pixel from a periodic view of space. Pixel noise, combined with the response to geophysical variability, has been demonstrated to be highly variable between pixels (Deeter et al., 2015). We investigate the impact of pixel-specific variability on validation. At each station, the error-weighted average for each pixel is calculated within a 1° radius to be validated against FTS measurements. Analysis is for daytime-only and land-only retrievals (except for water-only at IZA, MLO, and LRN).

Validation results differ between pixels. Most noticeably, pixel 1 provides consistently poorer correlations and larger standard deviations than the other three pixels (e.g., summarized for Lauder in Table 5). To visualize results at all stations, correlation is plotted against bias in Fig. 5. Perfect validation occurs at the intersection of the zero bias and unity correlation lines. All stations generally produce similar results to Lauder, with pixel 1 showing the poorest correlation in all three products. Figure 5 indicates that for TIR–NIR and NIR-only, pixel 1 also consistently shows the highest bias. Pixel 3 generally shows the highest correlation for all three products. Pixel-wise validation is also consistent with Sect. 4.2, where TIR-only generally produces the higher correlations, lower bias, and smaller standard deviations compared to TIR–NIR and NIR-only.

The effect of the deficient pixel 1 on overall retrievals is assessed by comparing the all-pixel error-weighted average with an error-weighted average including only pixels 2 to 4. For retrieval types TIR-only and TIR–NIR the validation results for the all-pixel averages are almost equivalent to the pixel 2 to 4 average results (Fig. 5). Therefore, MOPITT error diagnostics for TIR-only and TIR–NIR adequately account for the poor performance of pixel 1. In comparison, for
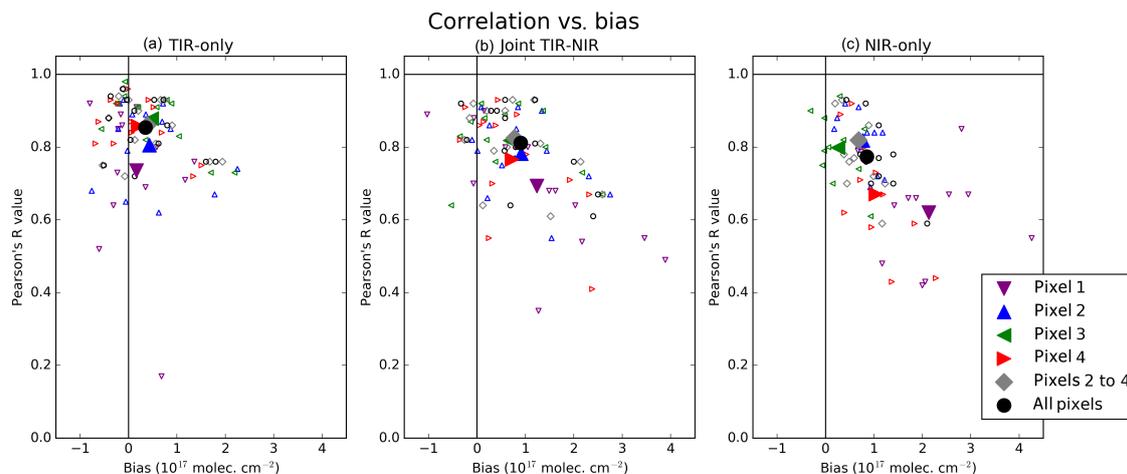
Figure 5. Pixel-wise validation results (correlation and bias) for each retrieval type at all FTS sites. Open markers represent results at each station and filled markers represent the mean value of all stations. The solid lines indicate zero bias and unity correlation. Columns are for the different MOPITT data products – **(a)** TIR-only, **(b)** TIR–NIR, and **(c)** NIR-only.

NIR-only, the all-pixel averages are systematically more biased than the pixel 2 to 4 averages. This suggests error characterization for the NIR-only product may need improvement.

Currently, the MOPITT level 3 product is a combination of pixels 1 and 2 (Deeter, 2013). Results here suggest the level 3 product may provide below-optimal representation of atmospheric CO due to pixel choice, particularly for the NIR-only retrievals. The current procedure of only maximizing DFS may not be the most optimal requirement for choosing pixels to be combined in the level 3 product.

Results found here may also be instructive for data assimilation, where weighted averages (or "superobservations" as described in Barré et al., 2015) are used to reduce satellite data to the model horizontal resolution. The poor performance of pixel 1 suggests that pixel 1 should be ignored when calculating the average. A more restrictive average would include the two best pixels for each dataset: pixels 3 and, 4 for TIR-only and TIR–NIR, and pixels 2 and 3, for NIR-only. The resulting average would include the satellite values that perform best against the FTS measurements.

5 Discussion and implications

5.1 Geographic dependence of validation

Retrievals from nadir-viewing satellite instruments that measure in the TIR are challenging over colder surfaces at higher latitudes, mainly due to low thermal contrast, resulting in a higher weighting to the a priori profiles and consequent lower information content. Information content of satellite retrievals is therefore dependent upon latitude. The latitudinal dependence of MOPITT DFS at these stations of interest is depicted in Fig. 6 (top row), which shows how DFS de-

creases moving closer to the poles in the TIR-only and TIR–NIR products. The relationship of DFS with bias and correlation is assessed through latitudinal dependence (Fig. 6).

The latitudinal dependence of MOPITT total column retrieval biases is consistent with (Deeter et al., 2014), who show V6 TIR-only biases relative to HIPPO are generally within $\pm 2 \times 10^{17}$ molec cm^{-2} (or approximately $\pm 10\%$). Results here show the latitudinal dependence is similarly bound for the TIR–NIR and NIR-only products (Fig. 6, middle row). The latitudinal resolution of NDACC stations is not as fine as for HIPPO measurements, particularly at the equator. Therefore, results here cannot confirm the negative bias in MOPITT found at the equator against HIPPO (Deeter et al., 2014). However, validation against FTS is complementary to HIPPO comparisons because FTS can validate MOPITT land and water scenes, compared to HIPPO measurements being taken only over water. The FTSs also provide longer time series, while HIPPO are a set of 5 $\sim$ 2-month campaigns taken between 2009 and 2011.

Supplementary material in (Jiang et al., 2015) suggested MOPITT bias may be related to DFS. We find that although DFS vary strongly with latitude for the TIR-only and TIR–NIR products, the MOPITT bias does not depend upon latitude. Similarly, we find no significant relationship between latitude and correlation values for these products. There is also no latitudinal dependence in the DFS, bias, or Pearson's R for the NIR-only product, reflecting that this product is not as affected by thermal contrast difficulties.

To help understand the driver of bias variability, we investigate the influence of altitude and find a larger range in bias at lower altitudes for the TIR-only and TIR–NIR products (Fig. 7). High biases in the TIR-only and TIR–NIR product (defined as $> 5\%$) all occur at low altitudes. High bias is most likely due to values from the single overpass time of
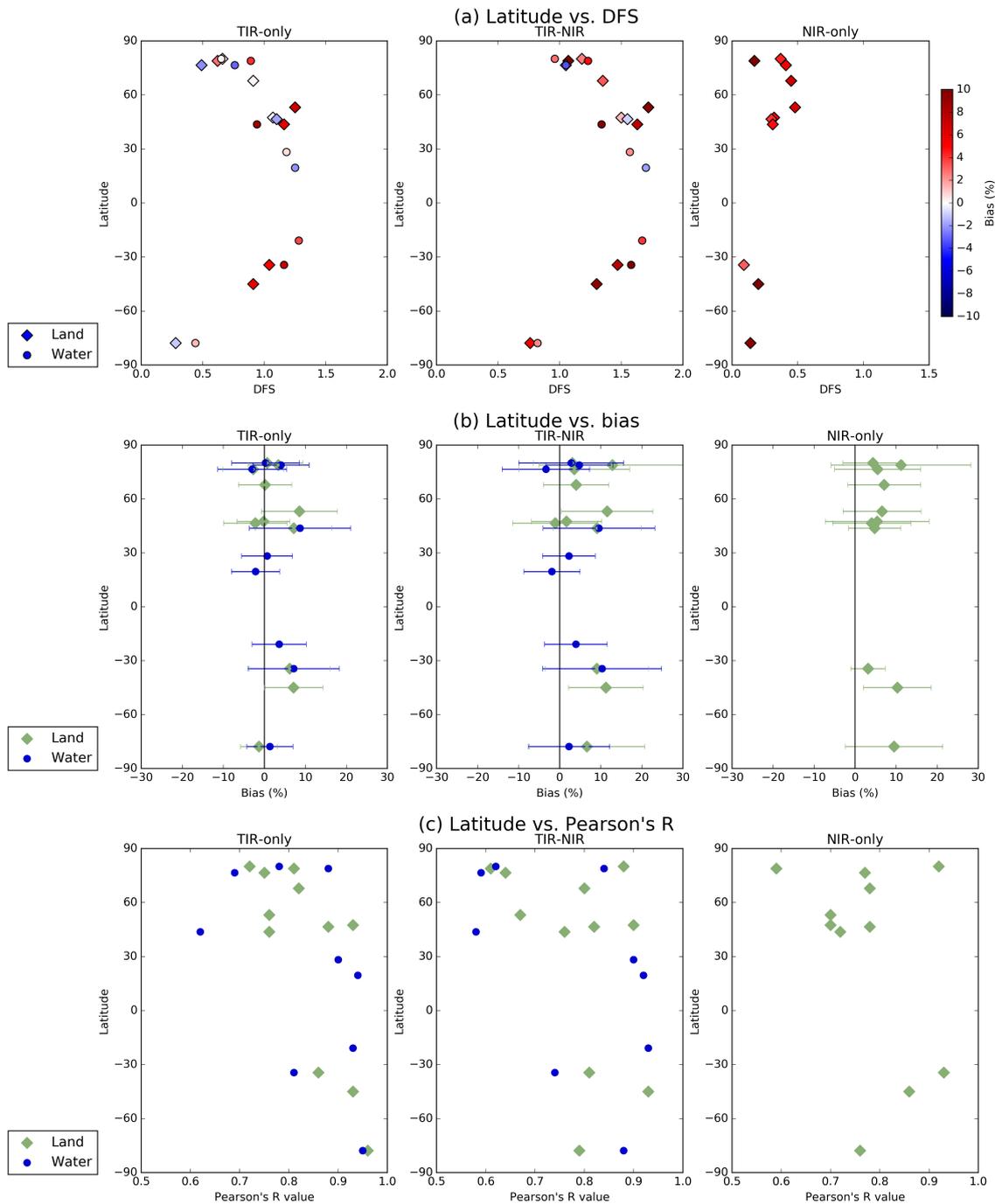
Figure 6. Latitudinal relationship with DFS **(a)**, relative bias ((MOPITT–FTS)/FTS, **b**), and Pearson's R correlation value **(c)**. Error bars in bias plots denote standard deviations. Columns are for the different MOPITT data products – column 1: TIR-only, column 2: TIR–NIR, and column 3: NIR-only.

MOPITT at 10:30 LT being compared with all daytime measurements from FTS at these stations. There is more variability throughout the day in the FTS column due to changes in lower tropospheric CO, which is not captured in the MOPITT measurements. For example, the FTS will capture diurnal variation due to greater atmospheric mixing through-

out the day, frontal systems bringing variable CO amounts, and/or rapid changes in nearby emissions and transport. Biases may be improved by temporally restricting comparisons closer to the 10:30 LT overpass. Further investigation would be necessary to determine the effect of temporal restriction at stations with high bias. The NIR-only product does not show
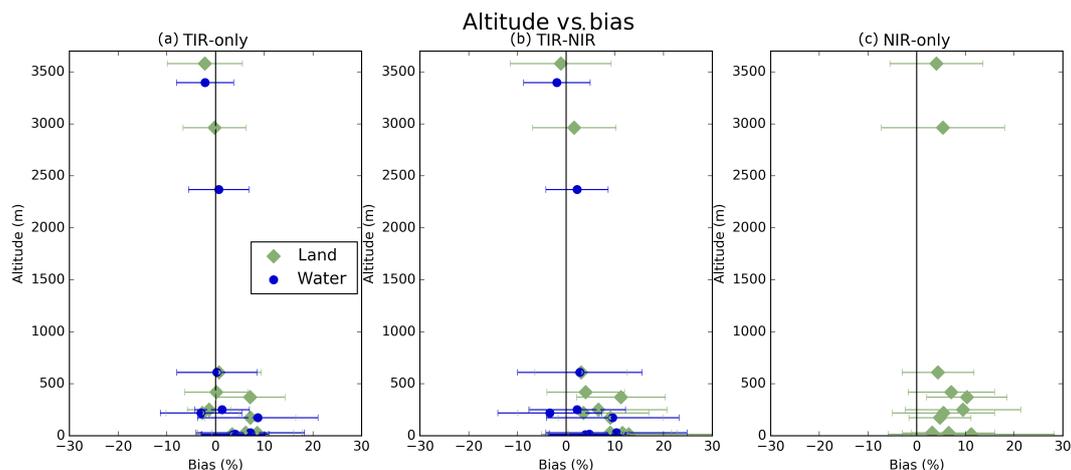
Figure 7. Altitude dependence of MOPITT relative bias ((MOPITT–FTS)/FTS). Error bars in bias plots denote standard deviations. Columns are for the different MOPITT data products – **(a)** TIR-only, **(b)** TIR–NIR, and **(c)** NIR-only.

bias dependence on altitude. Comparisons with NIR-only include a large amount of a priori information (mean DFS of 0.29), which masks the variability in FTS columns.

Satellite bias can introduce inaccuracies for data assimilation and inverse modeling studies, particularly at high latitudes (Hooghiemstra et al., 2012; Jiang et al., 2015; Gaubert et al., 2016). Significantly high emissions were attributed to high MOPITT bias in (Hooghiemstra et al., 2012), who suggest the need for satellite bias correction. Therefore, rather than restricting data to be assimilated to within ±40° (Jiang et al., 2015), we suggest that FTS could be used to either correct or account for MOPITT retrieval biases, particularly at high latitude stations, prior to data assimilation.

5.2 Temporal stability of MOPITT

Drift in MOPITT validation has been previously studied in relation to NOAA profiles (Deeter et al., 2014). However, there is limited evaluation of geographic variability in the drift as (Deeter et al., 2014) combined NOAA profiles at all locations for temporal analysis. The FTS datasets are ideal for assessing the global long-term stability in MOPITT retrievals, due to long records and spanning a range of latitudes.

Drift in validation is quantified by the slope of least-squares linear regression on the MOPITT–FTS biases against time. Bias drift is initially approximated as linear in time due to the unknown cause. Significant drift occurs when the p value from a two-tailed t test is less than 0.01. An example of the time dependence of MOPITT–FTS biases at all stations is shown for the TIR–NIR product in Fig. 8. Drift plots for other products are provided in Appendix A. Bias drift values are recorded in Table 3a–c.

Drift ranges between −0.96 and 0.30 (TIR-only), −1.4 and 0.94 (TIR–NIR), and −1.2 and 0.65 % yr^{-1} (NIR-only). TIR-only mean drift (−0.080 % yr^{-1}) is of opposite sign and lower in magnitude than that found by (Deeter et al., 2014),

who determined a drift of 0.003×10^{17} molec cm^{-2} yr^{-1} (approximately 0.15 % yr^{-1}). The TIR–NIR mean drift (−0.23 % yr^{-1}) is higher in magnitude and of opposite sign to the TIR–NIR drift found by (Deeter et al., 2014), approximately 0.15 % yr^{-1}. NIR-only mean drift (−0.14 % yr^{-1}) is equivalent to (Deeter et al., 2014), approximately −0.15 % yr^{-1}.

The geographical relationship of the bias drift is shown in Fig. 9. Drift in the Southern Hemisphere is small. In contrast, Northern Hemisphere drift is highly variable. Instrument degradation would be expected to produce consistent drift across stations. However, the variable drift implies the cause of drift is due to input parameters to the MOPITT retrieval process rather than instrument degradation.

Accounting for significant bias drift is challenging. At some stations MOPITT drift is large enough to obscure trend analysis, which has been shown to be on the order of −1 % yr^{-1} globally (Worden et al., 2013). A large drift at some stations is probably due to the influence of low time sampling and outliers in the FTS record (e.g., Bremen). Other stations with high bias drift occur at high latitudes in the Northern Hemisphere, where potential un-characterized surface errors could contribute to retrieval drift. (Ho et al., 2005) found a large standard deviation in a priori emissivity due to cloud detection uncertainties. Additionally, sea ice may not be correctly accounted for in the satellite retrievals because sea-ice scenes are retrieved with the same parameters as water, despite having different emissivity properties. Consequently, a trend in cloudiness or sea-ice extent could therefore produce a trend in MOPITT retrievals. As a result, we recommend avoiding the use of MOPITT retrievals above 60° N when assessing the temporal evolution of CO. Mean trends below 60° N are −0.12 for TIR-only, −0.33 for TIR–NIR, and −0.17 % yr^{-1} for NIR-only.
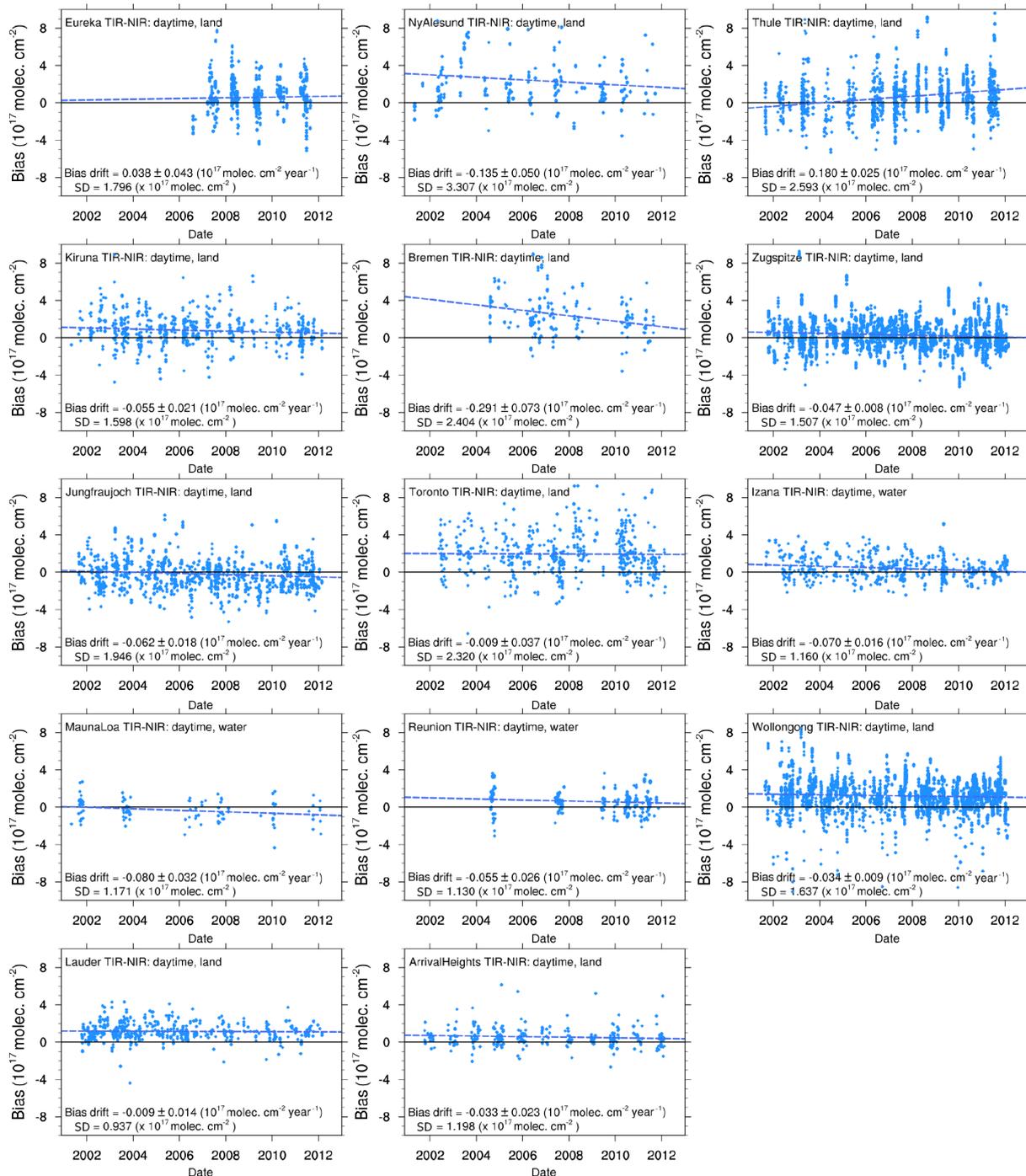
Figure 8. Temporal evolution of bias (MOPITT–FTS) at all stations for V6 joint TIR–NIR product. The blue dashed line indicates bias drift calculated from a linear least squares regression. Plots are ordered by latitude, from north to south.

6 Conclusions

The first systematic validation of MOPITT version 6 retrievals with ground-based FTS at 14 NDACC stations has demonstrated low bias of the MOPITT instrument (generally < 10 %) and has highlighted some important considerations

for using the satellite data in scientific analysis. While values have been calculated for an average of MOPITT values within a 1° radius, any systematic bias in MOPITT remains and is evaluated. MOPITT is generally biased high relative to FTS and bias was consistently higher for joint and NIR-only products than for the TIR-only product. Mean bias is 2.4 %
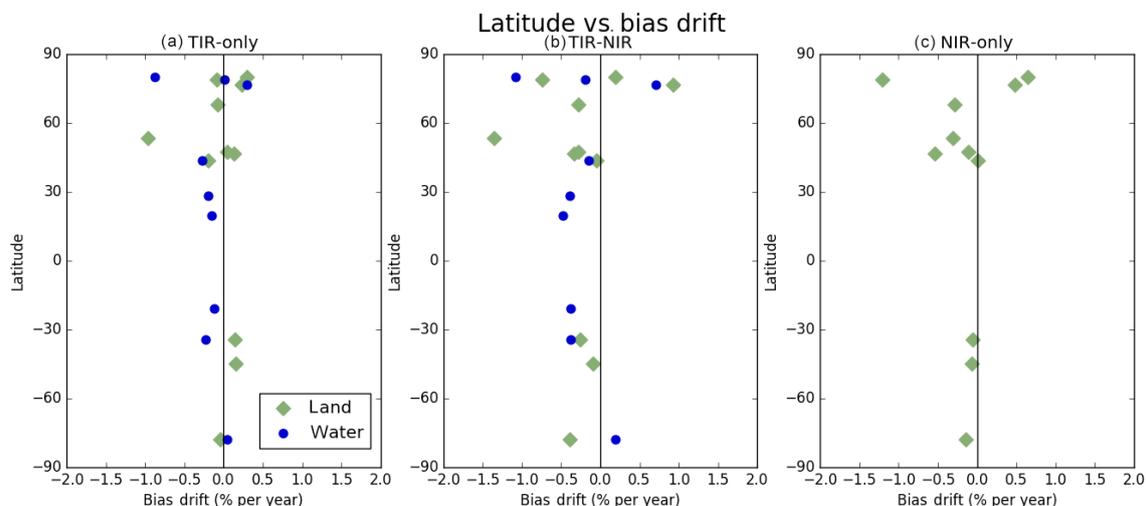
Figure 9. Geographic variability in bias drift. Columns are for the different MOPITT data products – **(a)** TIR-only, **(b)** TIR–NIR, and **(c)** NIR-only.

for TIR-only, 5.1 % for TIR–NIR, and 6.5 % for NIR-only. MOPITT retrieves with equivalent skill over land or water, and the highest amount of information is present in the joint TIR–NIR product as indicated by the largest DFS. Pixel-wise validation revealed the poor performance of pixel 1. Some applications that require data thinning techniques (for example data assimilation) may remove pixel 1 from weighted averages, as this pixel has the lowest correlation and most variability. The poor performance of pixel 1 also suggests that the processing of the level 3 product may need to be revised. We find no dependence of bias or correlation on latitude, suggesting no relationship to DFS. Variability in lower tropospheric CO influences MOPITT bias, which is probably due to sampling and sensitivity differences between instruments. The MOPITT bias found here may be used to account for satellite error prior to data assimilation, allowing for the inclusion of data over a wider spatial range than is currently used. The MOPITT long-term bias drift has been bound to within $\pm 0.5 \% \, \mathrm{yr}^{-1}$ or lower at almost all locations. Variable drift in the Northern Hemisphere implies an un-characterized retrieval parameter such as uncertainty in cloud detection or sea-ice representation. We recommend that trend analysis should not be performed above 60° N. Overall, this study extends the geographical and temporal analysis of MOPITT validation results.

Appendix A
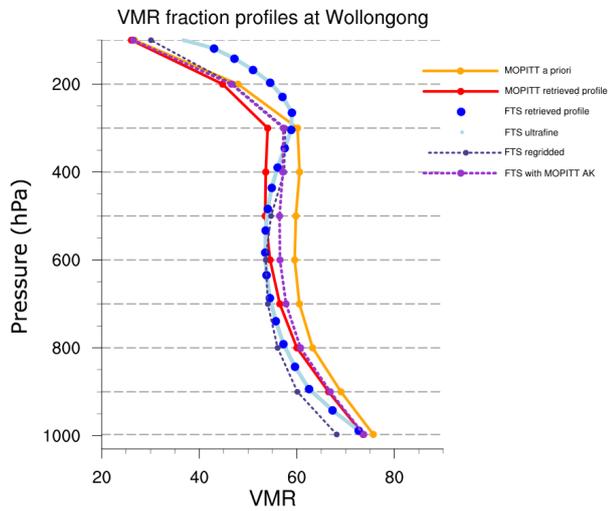
Figure A1. An example of the smoothing technique applying MO-PITT TIR-only AKs to an FTS profile (blue circles) at Wollongong. The FTS profile is interpolated in logP space to an ultrafine grid of 100 levels per MOPITT layer (light blue) and then averaged within each MOPITT layer (blue with dashed line) and smoothed by the MOPITT AK (purple with dashed line) to be commensurate with MOPITT values (red). Dashed grey horizontal lines indicate MO-PITT level edges.
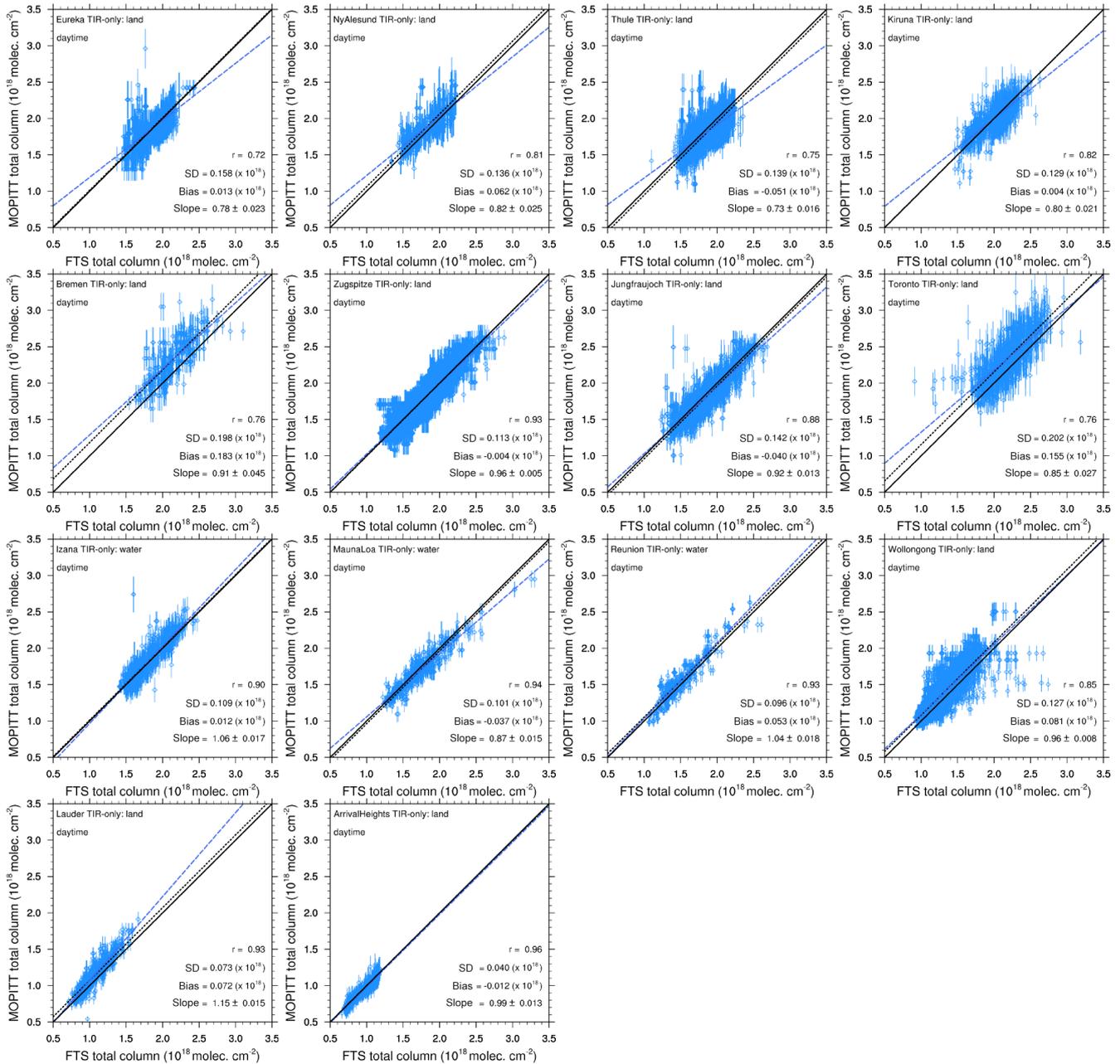
Figure A2. Correlation plot for V6 TIR-only product. The blue dashed line indicates correlation slope, the black dashed line indicates mean bias, and the solid black line is the 1:1 line for comparison. Error bars are the daily combined MOPITT error within a 1° radius.
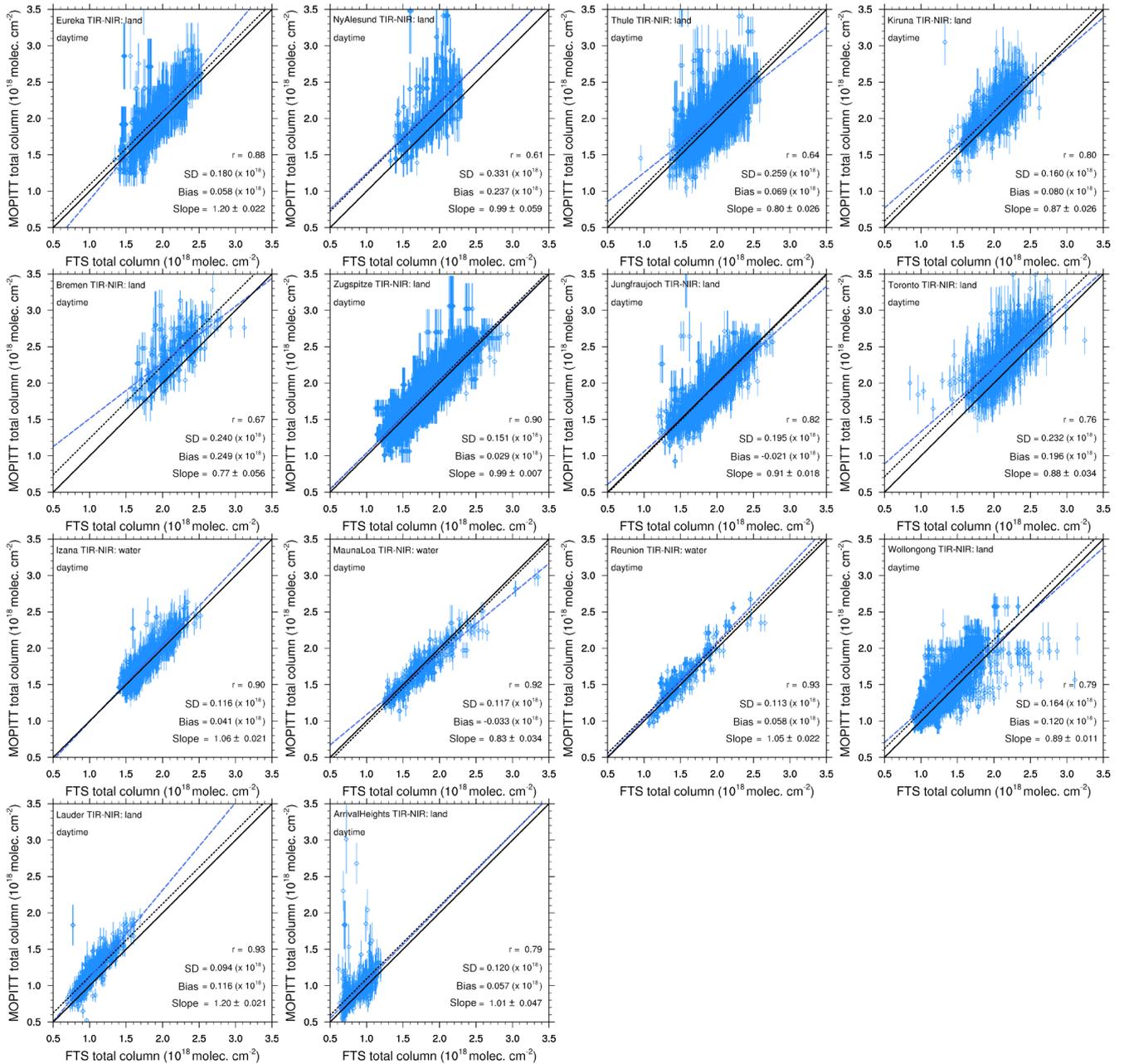
Figure A3. Correlation plot for V6 TIR–NIR product. The blue dashed line indicates correlation slope, the black dashed line indicates mean bias, and the solid black line is the 1 : 1 line for comparison. Error bars are the daily combined MOPITT error within a 1° radius.
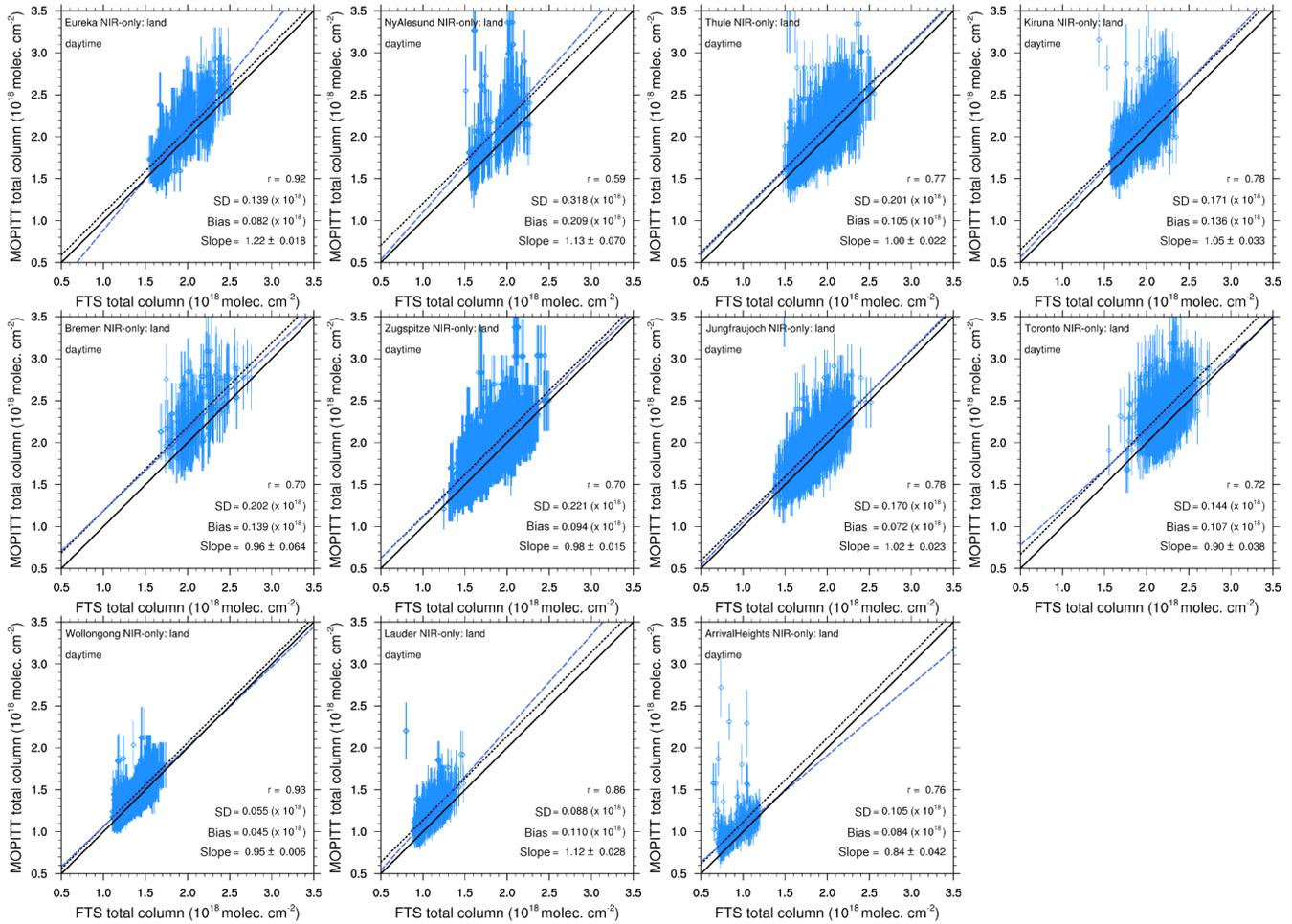
Figure A4. Correlation plot for V6 NIR-only product. The blue dashed line indicates correlation slope, the black dashed line indicates mean bias, and the solid black line is the 1 : 1 line for comparison. Error bars are the daily combined MOPITT error within a 1° radius.
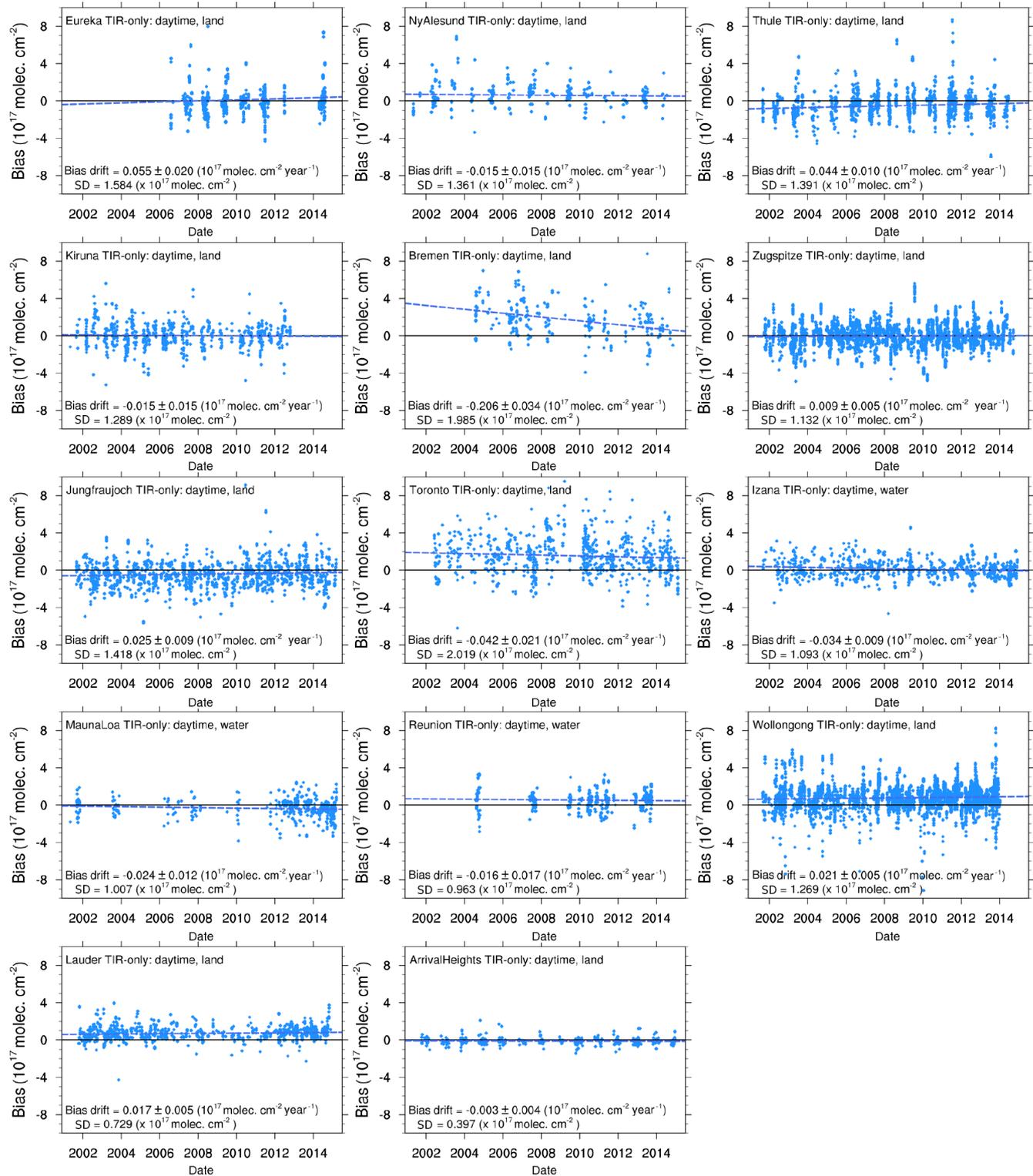
Figure A5. Temporal evolution of bias (MOPITT–FTS) at all stations for V6 TIR-only product. The blue dashed line indicates bias drift.
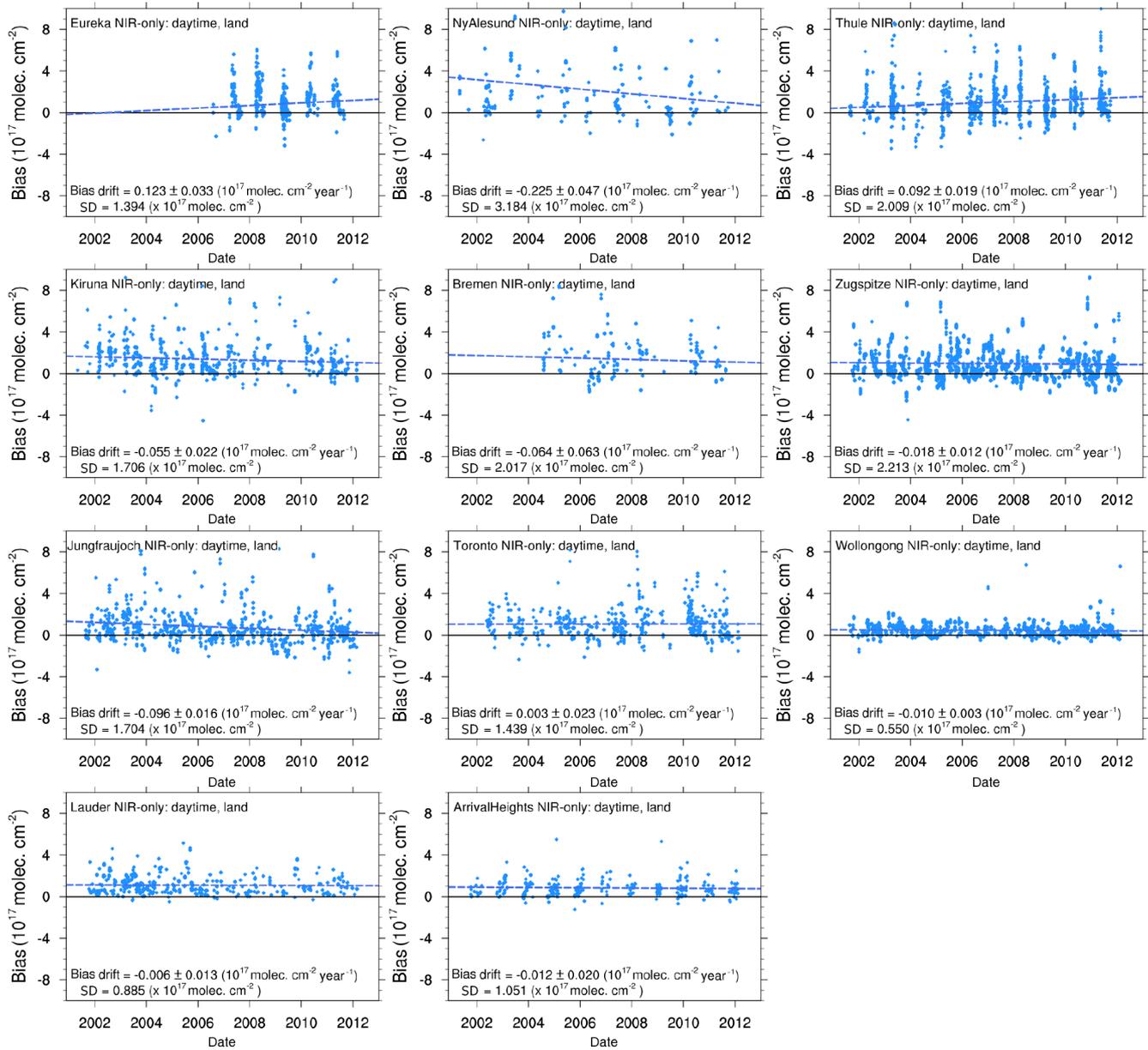
Figure A6. Temporal evolution of bias (MOPITT–FTS) at all stations for V6 NIR-only product. The blue dashed line indicates bias drift.

Appendix B: Evaluating the effect of daily averaging within a 1° radius

This section discusses the following three validation methods:

- Method 1: point-wise application of averaging kernels and a priori profiles.

- Method 2: point-wise application of AKs and a priori profiles followed by weighted averaging.

- Method 3: application of weighted average AKs and a priori profiles.

The point-wise application of AKs and a priori profiles (Method 1) is commonly used to compare two remote-sensing instruments (e.g., Clerbaux et al., 2008; Kerzenmacher et al., 2012; Martínez-Alonso et al., 2014). However, in the MOPITT case, Method 1 can show high variability due to random error. In this study, random error is reduced by daily averaging within a 1° radius of each station. Weighting the averages by MOPITT retrieval error adds more weight to the most reliable measurements. Any unknown systematic bias will remain and be determined in the validation against the FTS measurements. Averaging can therefore help produce a more robust calculation of the mean bias and long-term drift by minimizing the interference from random error and known retrieval errors.

Averaging may be applied after point-wise smoothing (Method 2). Our study aimed to complete multiple comparisons at the same location (e.g., land versus water, different detector elements, different MOPITT retrieval versions, and different seasons). This required a method that was faster than Method 2. The approximation method, where error-weighted daily mean MOPITT AKs and a priori profiles are applied (Method 3), was found to dramatically reduce calculation time. The use of mean AKs is justified because it is reasonable to assume low variability in our dataset – due to averages over small ranges in space (1°) and time (1 day). We further justify the use of mean AKs by empirically comparing results from Method 2 and Method 3. The MOPITT V6 TIR–NIR product is used for illustration.

Figures B1 and B2 provide a visual example of validation results for methods 1 to 3. Further validation results are provided in Table B1a, b, and c. The validation methodology is equivalent to that described in Sect. 3 – apart from whether, or when, daily averaging within a 1° radius is applied. These validation results show to what extent the choice of method affects results. From these results, it is clear that outliers and variance are reduced in the averaging methods (2 and 3) compared to the point-wise method (1). Our analysis is consistent with the supplementary material of (Martínez-Alonso et al., 2014). Correlation is also improved in methods 2 and 3 compared to Method 1, indicating a stronger relationship between the MOPITT and FTS datasets when weighted averages are used and supporting our choice to reduce random error.

A direct comparison of validation results between methods 2 and 3 is shown in Table . A negative (positive) bias means Method 2 produced larger (smaller) magnitude results than Method 3. Results show that Method 3 is equivalent to Method 2 at 11 out of 14 stations, as determined by Welch's t test on the mean bias to a significance level of 0.01. The three stations that lie outside this significance level (Zugspitze, Jungfraujoch, and Lauder) are close to mountainous regions which may not conform to the assumption of low variability within 1°, specifically due to variable surface pressures. Care must be taken when interpreting the validation values at these stations. However, for two stations (Zugspitze and Jungfraujoch) bias is very low – below 3 % (Table B1b and c). In practice, the choice of method in these two cases does not impact conclusions, i.e., that these two stations have some of the lowest bias.

We have also tested the effect of using mean AKs (Method 3) on the all-station mean values (bias, correlation, drift) and found no substantial differences to Method 2. We conclude that using error-weighted daily mean MOPITT AKs and a priori profiles within a 1° radius (Method 3) is a valid methodology and produces results that are at most stations consistent with the point-wise application of AKs and a priori profiles followed by averaging (Method 2). Therefore, we can rely on the resulting biases and drift values presented in the main text.
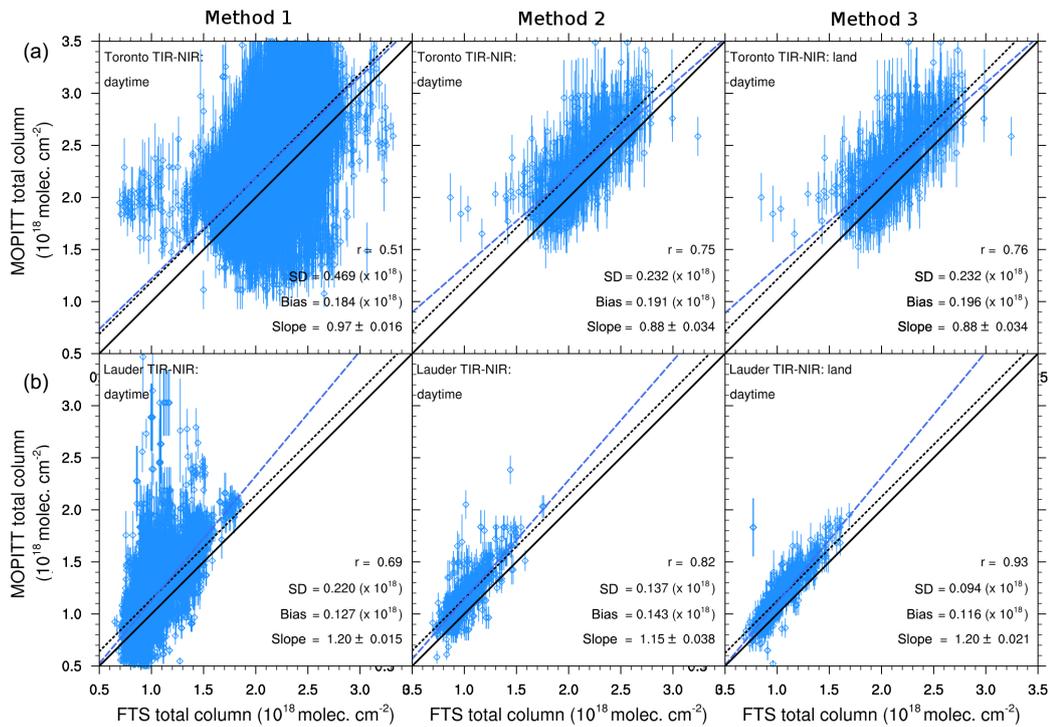
Figure B1. Example of correlation plots for validation methods 1 to 3 for V6 TIR–NIR product at Toronto (**a**) and Lauder (**b**). Methods are arranged in columns. The blue dashed line indicates the correlation slope, the black dashed line indicates mean bias, and the solid black line is the 1 : 1 line for comparison. Error bars are the MOPITT error.
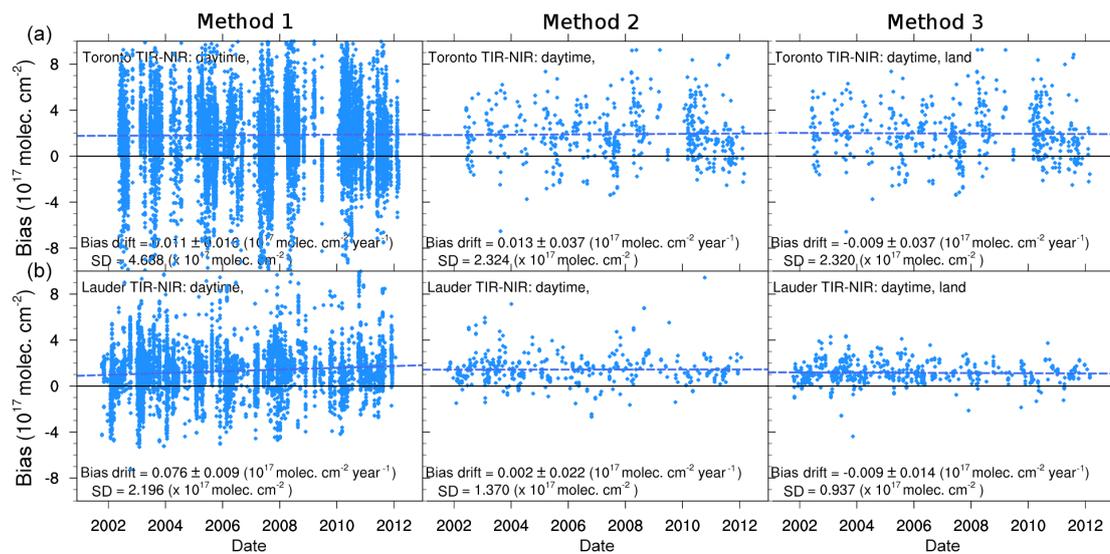

Figure B2. Example of the temporal evolution of bias (MOPITT–FTS) for validation methods 1 to 3 for the V6 TIR–NIR product at Toronto (**a**) and Lauder (**b**). Methods are arranged in columns. The blue dashed line indicates bias drift.

Table B1. Validation results and mean smooth FTS total column from the **(a)** point-wise method (Method 1), **(b)** point-wise followed by daily error-weighted averaging (Method 2), and **(c)** error-weighted mean AKs and a priori profiles (Method 3). Validation values in **(c)** are the same as reported in the main text, Table 3a, b, and c: they are reproduced here for clarity. Total column (tcol) is in $\times 10^{18}$ molec cm^{-2}.

Station	tcol	r	bias (%)	drift (% yr^{-1})	SD (%)	#obs
(a) Method 1: point-wise application						
EUR	1.99	0.63	4.72	0.42	18.08	7780
NYA	1.88	0.47	9.71	−0.51	22.84	2948
THU	2.01	0.52	2.18	0.50	17.16	6568
KIR	2.00	0.67	3.83	−0.35	12.74	7601
BRE	2.16	0.60	10.32	−0.99	13.96	5612
ZUG	1.70	0.74	3.72	−0.22	15.57	65742
JFJ	1.80	0.74	0.08	−0.28	15.09	17813
TAO	2.18	0.51	8.42	0.05	21.47	10776
IZA	1.80	0.72	2.48	−0.32	12.36	18226
MLO	1.72	0.83	−2.76	−0.50	10.93	2529
LRN	1.51	0.88	4.67	−0.48	10.73	12747
WOL	1.31	0.67	8.13	0.01	18.76	86202
LAU	1.01	0.69	12.55	0.75	21.67	6748
AHS	0.826	0.38	7.03	−0.39	31.28	2297
(b) Method 2: point-wise application and then averaging						
EUR	1.89	0.79	4.16	0.76	12.60	950
NYA	1.84	0.61	12.78	−0.71	17.82	482
THU	1.93	0.60	3.71	0.92	15.14	1455
KIR	2.00	0.80	4.04	−0.27	7.95	643
BRE	2.16	0.67	11.62	−1.22	11.00	249
ZUG	1.76	0.89	2.97	−0.33	8.85	4355
JFJ	1.88	0.81	0.43	−0.34	10.39	1243
TAO	2.16	0.75	8.84	0.06	10.75	512
IZA	1.82	0.90	2.22	−0.38	6.38	599
MLO	1.71	0.92	−1.95	−0.47	6.86	111
LRN	1.48	0.93	3.92	−0.37	7.64	376
WOL	1.31	0.80	9.62	−0.52	13.49	3618
LAU	1.03	0.82	13.98	0.02	13.35	425
AHS	0.865	0.57	9.85	−0.36	28.02	248
(c) Method 3: application of mean AKs						
EUR	1.90	0.88	3.06	0.20	9.47	880
NYA	1.84	0.61	12.86	−0.73	17.94	482
THU	1.93	0.64	3.56	0.94	13.46	1395
KIR	2.00	0.80	3.99	−0.27	7.97	643
BRE	2.15	0.67	11.57	−1.35	11.16	241
ZUG	1.76	0.90	1.67	−0.27	8.55	4197
JFJ	1.88	0.82	−1.10	−0.33	10.36	1229
TAO	2.16	0.76	9.09	−0.04	10.74	512
IZA	1.82	0.90	2.24	−0.39	6.39	599
MLO	1.71	0.92	−1.91	−0.47	6.83	111
LRN	1.48	0.93	3.94	−0.37	7.64	376
WOL	1.31	0.81	9.07	−0.25	12.58	3994
LAU	1.03	0.93	11.24	−0.09	9.10	569
AHS	0.852	0.79	6.65	−0.39	14.07	291

Table B2. Differences between validation results from the point-wise application of AKs and a priori profiles followed by averaging (Method 2) and mean AKs (Method 3). Differences are calculated between absolute values (|Method 3|–|Method 2|).

| Station | Difference between Method 2 and Method 3. | | | | | |
	Δ tcol[1] %	Δr	Δ bias pp	Δ drift pp yr^{-1}	t test[2]	p value[2]
EUR	0.1	0.09	−1.1	−0.56	2.10	0.036
NYA	−0.1	0	0.08	0.02	−0.05	0.962
THU	−0.1	0.04	−0.15	0.02	0.28	0.779
KIR	0	0	−0.05	0	0.11	0.911
BRE	−0.2	0	−0.05	0.13	0.09	0.926
ZUG	0.1	0.01	**−1.3**	−0.06	**6.92**	**4.75×10^{-12}**
JFJ	−0.1	0.01	**0.67***	−0.01	**3.67**	**2.47×10^{-04}**
TAO	−0.1	0.01	0.25	−0.02*	−0.34	0.730
IZA	0	0	0.02	0.01	−0.06	0.952
MLO	0	0	−0.04	0	−0.04	0.965
LRN	0	0	0.02	0	−0.04	0.971
WOL	−0.4	0.01	−0.55	−0.27	1.75	0.081
LAU	0.4	0.11	**−2.74**	0.07*	**3.50**	**4.99×10^{-04}**
AHS	−1.5	0.22	−3.2	0.03	1.69	0.093

[1] Relative to Method 3 values. [2] Welch's T test and corresponding p value are calculated on mean bias values. * Change in sign between methods. Bold indicates where mean bias values are significantly different according to a Welch's t test at the $p < 0.01$ level.

Appendix C: Evaluation of the validation methodology when altitude differences exist between instruments

As discussed in Sect. 3.3, the FTS profile must be re-gridded to the satellite retrieval grid before smoothing by the MO-PITT AK. When the FTS surface pressure is smaller than the MOPITT surface pressure, a choice must be made as to how to deal with the missing FTS levels in the re-gridded profile. Here, we explore two main methodologies to deal with missing levels.

The first option is a "profile truncation" method that ensures the missing levels of the re-gridded FTS profile remain missing. Consequently, the smoothing equation will introduce errors as there will be zero contribution from the lowest level onto values in the levels above. The a priori layers are also masked where FTS layers are missing. This is in contrary to how the MOPITT instrument has retrieved information and will introduce bias on every level. The smoothed FTS result in this case is compared with a MOPITT partial column that is calculated over the same levels as the re-gridded FTS, i.e., subtracting the missing levels.

The second option is to fill in the missing levels of the re-gridded FTS profile. This will also introduce error during smoothing because we have assumed a true value below the ground level of the FTS. Previously, studies have achieved profile filling by determining a scaling factor between non-missing FTS levels and the satellite a priori layers (e.g., Kerzenmacher et al., 2012). Instead, we use the FTS a priori profiles in determining scaling factors in order to avoid artificially moving the FTS "truth" toward satellite measurements. Seeing as the FTS a priori profiles originate from the global model WACCM, we call this the "WACCM scaling"

method. WACCM profiles are re-gridded to the MOPITT vertical grid prior to determining a scaling factor.

We assess the impact of the two methods on validation results by testing with FTS profiles where values exists at all levels of the re-gridded profile. The TIR–NIR product is used for illustration. Original validation results are compared with those from either removing or replacing (filling) one lowest level. Eureka, Zugspitze, Jungfraujoch, Izaña, Mauna Loa, and La Réunion experience FTS surface pressures that are smaller than MOPITT at all times and are therefore not included in the comparison.

We present the mean smooth FTS total column and MO-PITT relative bias values in Table C1. Values in Table C1 are slightly different than those found in the main text because they do not include instances where altitude differences occur.

The differences between the profile truncation or WACCM scaling methods and the values from the original profile validation are shown in Table C2. A negative bias means the truncation or filling methodology produced larger magnitude results than the original profile results, and vice versa. Table C2 shows that although there are large differences in total column between the original profile and profile truncation methods (over 20 % at Bremen), the bias only changes within a few percentage points. Bias is generally improved for the truncation results compared to the original profile.

WACCM scaling produces differences in total column of generally less than 0.6 %, and mean bias is altered generally less than 0.6 pp. Bremen is the only station that does not agree with the other results, and this may be due to over 20 % of the total column being contributed by the lowest level.

While the profile truncation method would not significantly affect the validation results, the profile filling method using WACCM scaling is superior and is therefore the method chosen in the main text. Table C3 shows that the TIR–NIR results from WACCM scaling are consistent across the TIR-only and NIR-only products.

Table C1. Evaluation of replacement or removal of the lowest level (V6 TIR–NIR). MOPITT relative bias is (MOPITT–FTS)/FTS.

Station	Original profile		Profile truncation		WACCM scaling	
	mean FTS tcol (smooth) $\times 10^{18}$ molec cm^{-2}	MOPITT relative bias %	mean FTS tcol (smooth) $\times 10^{18}$ molec cm^{-2}	MOPITT relative bias %	mean FTS tcol (smooth) $\times 10^{18}$ molec cm^{-2}	MOPITT relative bias %
NYA	1.85	13.61	1.61	13.71	1.85	13.69
THU	1.90	3.89	1.69	2.87	1.90	3.95
KIR	2.00	4.22	1.86	3.00	2.00	4.16
BRE	2.16	11.43	1.70	9.72	2.12	13.40
TAO	2.17	9.11	1.78	6.85	2.16	9.54
WOL	1.31	9.00	1.16	8.29	1.31	9.52
LAU	1.03	10.99	0.977	10.48	1.03	10.94
AHS	0.797	6.67	0.712	3.84	0.797	6.63

Table C2. Difference between original profile validation and truncation or profile filling techniques (V6 TIR–NIR).

Station	Profile truncation		WACCM scaling	
	Δ tcol[1] %	Δ bias pp	Δ tcol[1] %	Δ bias pp
NYA	−13.1	−0.10	−0.1	−0.08
THU	−10.8	1.02	−0.1	−0.06
KIR	−7.2	1.22	0.1	0.06
BRE	−21.4	1.71	−1.7	−1.97
TAO	−17.7	2.26	−0.4	−0.43
WOL	−11.3	0.71	−0.5	−0.52
LAU	−4.9	0.51	0.1	0.05
AHS	−10.6	2.83	0.0	0.04

Differences are calculated as |original profile| − |X method|, where
X is the profile truncation or WACCM scaling. [1] Relative to
original profile values.

Table C3. Difference between original profile validation and profile filling techniques (V6 TIR-only and NIR-only).

Station	TIR-only		NIR-only	
	Δ tcol[1] %	Δ bias pp	Δ tcol[1] %	Δ bias pp
NYA	−0.0	−0.03	−0.0	−0.04
THU	0.0	−0.01	−0.1	−0.07
KIR	0.0	0.03	0.1	0.06
BRE	−1.3	−1.46	−1.1	−1.17
TAO	−0.4	−0.39	−0.2	−0.24
WOL	−0.3	−0.37	−0.1	−0.09
LAU	0.1	0.08	0.1	0.06
AHS	−0.0	0.02	0.1	0.07

Differences are calculated as |original profile| − |X method|, where
X is the WACCM scaling. [1] Relative to original profile values.

References

Bacsik, Z., Mink, J., and Keresztury, G.: FTIR spectroscopy of the atmosphere. I. Principles and methods, Appl. Spectrosc. Rev., 39, 295–363, 2004.

Barré, J., Gaubert, B., Arellano, A. F. J., Worden, H. M., Edwards, D. P., Deeter, M. N., Anderson, J. L., Raeder, K., Collins, N., Tilmes, S., Francis, G., Clerbaux, C., Emmons, L. K., Pfister, G. G., Coheur, P.-F., and Hurtmans, D.: Assessing the impacts of assimilating IASI and MOPITT CO retrievals using CESM-CAM-chem and DART, J. Geophys. Res.-Atmos., 120, 10501–10529, doi:10.1002/2015JD023467, 2015.

Batchelor, R. L., Strong, K., Lindenmaier, R., Mittermeier, R. L., Fast, H., Drummond, J. R., Fogal, P. F., Batchelor, R. L., Strong, K., Lindenmaier, R., Mittermeier, R. L., Fast, H., Drummond, J. R., and Fogal, P. F.: A New Bruker IFS 125HR FTIR Spectrometer for the Polar Environment Atmospheric Research Laboratory at Eureka, Nunavut, Canada: Measurements and Comparison with the Existing Bomem DA8 Spectrometer, J. Atmos. Ocean. Tech., 26, 1328–1340, doi:10.1175/2009JTECHA1215.1, 2009.

Blumenstock, T., Kopp, G., Hase, F., Hochschild, G., Mikuteit, S., Raffalski, U., and Ruhnke, R.: Observation of unusual chlorine activation by ground-based infrared andmicrowave spectroscopy in the late Arctic winter 2000/01, Atmos. Chem. Phys., 108, 897–905, doi:10.5194/acp-6-897-2006, 2006.

Clerbaux, C., George, M., Turquety, S., Walker, K. A., Barret, B., Bernath, P., Boone, C., Borsdorff, T., Cammas, J. P., Catoire, V., Coffey, M., Coheur, P.-F., Deeter, M., De Mazière, M., Drummond, J., Duchatelet, P., Dupuy, E., de Zafra, R., Eddounia, F., Edwards, D. P., Emmons, L., Funke, B., Gille, J., Griffith, D. W. T., Hannigan, J., Hase, F., Höpfner, M., Jones, N., Kagawa, A., Kasai, Y., Kramer, I., Le Flochmoën, E., Livesey, N. J., López-Puertas, M., Luo, M., Mahieu, E., Murtagh, D., Nédélec, P., Pazmino, A., Pumphrey, H., Ricaud, P., Rinsland, C. P., Robert, C., Schneider, M., Senten, C., Stiller, G., Strandberg, A., Strong, K., Sussmann, R., Thouret, V., Urban, J., and Wiacek, A.: CO measurements from the ACE-FTS satellite instrument: data analysis and validation using ground-based, airborne and spaceborne observations, Atmos. Chem. Phys., 8, 2569–2594, doi:10.5194/acp-8-2569-2008, 2008.

David, S. J., Beaton, S. A., Anderberg, M. H., and Murcray, F. J.: Determination of total ozone over Mauna Loa using very high resolution infrared solar spectra, Geophys. Res. Lett., 20, 2055–2058, doi:10.1029/93GL02470, 1993.

de Laat, A. T. J., Gloudemans, A. M. S., Schrijver, H., Aben, I., Nagahama, Y., Suzuki, K., Mahieu, E., Jones, N. B., Paton-Walsh, C., Deutscher, N. M., Griffith, D. W. T., De Mazière, M., Mittermeier, R. L., Fast, H., Notholt, J., Palm, M., Hawat, T., Blumenstock, T., Hase, F., Schneider, M., Rinsland, C., Dzhola, A. V., Grechko, E. I., Poberovskii, A. M., Makarova, M. V., Mellqvist, J., Strandberg, A., Sussmann, R., Borsdorff, T., and Rettinger, M.: Validation of five years (2003–2007) of SCIAMACHY CO total column measurements using ground-based spectrometer ob-

servations, Atmos. Meas. Tech., 3, 1457–1471, doi:10.5194/amt-3-1457-2010, 2010.

de Laat, A. T. J., Aben, I., Deeter, M., Nédélec, P., Eskes, H., Attié, J.-L., Ricaud, P., Abida, R., El Amraoui, L., and Landgraf, J.: Validation of nine years of MOPITT V5 NIR using MOZAIC/IAGOS measurements: biases and long-term stability, Atmos. Meas. Tech., 7, 3783–3799, doi:10.5194/amt-7-3783-2014, 2014.

Deeter, M. N.: MOPITT (Measurements of Pollution in the Troposphere) Version 5 Product, User's guide, MOPITT Algorithm Development Team, Atmospheric Chemistry Division, National Center for Atmospheric Research, Boulder, CO, USA, 2011.

Deeter, M. N.: MOPITT (Measurements of Pollution in the Troposphere) Version 6 Product, User's guide, MOPITT Algorithm Development Team, Atmospheric Chemistry Division, National Center for Atmospheric Research, Boulder, CO, USA, 2013.

Deeter, M. N., Emmons, L. K., Francis, G. L., Edwards, D. P., Gille, J. C., Warner, J. X., Khattatov, B., Ziskin, D., Lamarque, J.-F., Ho, S. P., Yudin, V., Attie, J. L., Packman, D., Chen, J., Mao, D., and Drummond, J. R.: Operational carbon monoxide retrieval algorithm and selected results for the MOPITT instrument, J. Geophys. Res., 108, 4399, doi:10.1029/2002JD003186, 2003.

Deeter, M. N., Emmons, L. K., Francis, G. L., Edwards, D. P., Gille, J. C., Warner, J. X., Khattatov, B., Ziskin, D., Lamarque, J.-F., Ho, S. P., Yudin, V., Attie, J. L., Packman, D., Chen, J., Mao, D., Drummond, J. R., Novelli, P., and Sachse, G.: Evaluation of operational radiances for the Measurements of Pollution in the Troposphere (MOPITT) instrument CO thermal band channels, J. Geophys. Res., 109, D03308, doi:10.1029/2003JD003970, 2004.

Deeter, M. N., Edwards, D. P., Gille, J. C., and Drummond, J. R.: CO retrievals based on MOPITT near-infrared observations, J. Geophys. Res., 114, D04303, doi:10.1029/2008JD010872, 2009.

Deeter, M. N., Edwards, D. P., Gille, J. C., Emmons, L. K., Francis, G., Ho, S. P., Mao, D., Masters, D., Worden, H., and Drummond, J. R.and Novelli, P. C.: The MOPITT version 4 CO product: Algorithm enhancements, validation, and long-term stability, J. Geophys. Res.-Atmos., 115, D07306, doi:10.1029/2009JD013005, 2010.

Deeter, M. N., Worden, H. M., Gille, J. C., Edwards, D. P., Mao, D., and Drummond, J. R.: MOPITT multispectral CO retrievals: Origins and effects of geophysical radiance errors, J. Geophys. Res.-Atmos., 116, 1–10, doi:10.1029/2011JD015703, 2011.

Deeter, M. N., Martínez-Alonso, S., Edwards, D. P., Emmons, L. K., Gille, J. C., Worden, H. M., Pittman, J. V., Daube, B. C., and Wofsy, S. C.: Validation of MOPITT Version 5 thermal-infrared, near-infrared, and multispectral carbon monoxide profile retrievals for 2000-2011, J. Geophys. Res.-Atmos., 118, 6710–6725, doi:10.1002/jgrd.50272, 2013.

Deeter, M. N., Martínez-Alonso, S., Edwards, D. P., Emmons, L. K., Gille, J. C., Worden, H. M., Sweeney, C., Pittman, J. V., Daube, B. C., and Wofsy, S. C.: The MOPITT Version 6 product: algorithm enhancements and validation, Atmos. Meas. Tech., 7, 3623–3632, doi:10.5194/amt-7-3623-2014, 2014.

Deeter, M. N., Edwards, D. P., Gille, J. C., and Worden, H. M.: Information content of MOPITT CO profile retrievals: Temporal and geographical variability, J. Geophys. Res.-Atmos., 120, 12723–12738, doi:10.1002/2015JD024024, 2015.

Drummond, J. R.: Measurements of pollution in the troposphere (MOPITT), Use of EOS for Studies of Atmospheric Physics, 77–101, 1992.

Drummond, J. R. and Mand, G. S.: The Measurements of Pollution in the Troposphere (MOPITT) Instrument: Overall Performance and Calibration Requirements, J. Atmos. Ocean. Tech., 13, 314–320, doi:10.1175/1520-0426(1996)013<0314:TMOPIT>2.0.CO;2, 1996.

Drummond, J. R., Zou, J., Nichitiu, F., Kar, J., Deschambaut, R., and Hackett, J.: A review of 9-year performance and operation of the {MOPITT} instrument, Adv. Space Res., 45, 760–774, doi:10.1016/j.asr.2009.11.019, 2010.

Duncan, B. N., Logan, J. A., Bey, I., Megretskaia, I. A., Yantosca, R. M., Novelli, P. C., Jones, N. B., and Rinsland, C. P.: Global budget of CO, 1988–1997: Source estimates and validation with a global model, J. Geophys. Res.-Atmos., 112, D22301, doi:10.1029/2007JD008459, 2007.

Edwards, D. P., Halvorson, C. M., and Gille, J.: Radiative transfer modeling for the EOS Terra satellite Measurements of Pollution in the Troposphere (MOPITT) instrument, J. Geophys. Res.-Atmos., 104, 16755–16775, 1999.

Edwards, D. P., Emmons, L. K., Gille, J. C., Chu, A., Attié, J. L., Giglio, L., Wood, S. W., Haywood, J., Deeter, M. N., Massie, S. T., Ziskin, D. C., and Drummond, J. R.: Satellite Observed Pollution From Southern Hemisphere Biomass Burning, J. Geophys. Res.-Atmos., 111, D14312, doi:10.1029/2005JD006655, 2006.

Emmons, L. K., Deeter, M. N., Gille, J. C., Edwards, D. P., Attié, J.-L., Warner, J., Ziskin, D., Francis, G., Khattatov, B., Yudin, V., Lamarque, J.-F., Ho, S.-P., Mao, D., Chen, J. S., Drummond, J., Novelli, P., Sachse, G., Coffey, M. T., Hannigan, J. W., Gerbig, C., Kawakami, S., Kondo, Y., Takegawa, N., Schlager, H., Baehr, J., and Ziereis, H.: Validation of Measurements of Pollution in the Troposphere (MOPITT) CO retrievals with aircraft in situ profiles, J. Geophys. Res.-Atmos., 109, D03309, doi:10.1029/2003JD004101, 2004.

Emmons, L. K., Edwards, D. P., Deeter, M. N., Gille, J. C., Campos, T., Nédélec, P., Novelli, P., and Sachse, G.: Measurements of Pollution In The Troposphere (MOPITT) validation through 2006, Atmos. Chem. Phys., 9, 1795–1803, doi:10.5194/acp-9-1795-2009, 2009.

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, doi:10.1029/2006JD007485, 2007.

Gaubert, B., Arellano, A. F., Barré, J., Worden, H. M., Emmons, L. K., Tilmes, S., Buchholz, R. R., Vitt, F., Raeder, K., Collins, N., Anderson, J. L., Wiedinmyer, C., Martinez Alonso, S., Edwards, D. P., Andreae, M. O., Hannigan, J. W., Petri, C., Strong, K., and Jones, N.: Toward a chemical reanalysis in a coupled chemistry-climate model: An evaluation of MOPITT CO assimilation and its impact on tropospheric composition, J. Geophys. Res.-Atmos., 121, 7310–7343, doi:10.1002/2016JD024863, 2016.

George, M., Clerbaux, C., Hurtmans, D., Turquety, S., Coheur, P.-F., Pommier, M., Hadji-Lazaro, J., Edwards, D. P., Worden, H., Luo, M., Rinsland, C., and McMillan, W.: Carbon monoxide distributions from the IASI/METOP mission: evaluation with other

space-borne remote sensors, Atmos. Chem. Phys., 9, 8317–8330, doi:10.5194/acp-9-8317-2009, 2009.

Hannigan, J., Coffey, M. T., and Goldman, A.: Semiautonomous FTS Observation System for Remote Sensing of Stratospheric and Tropospheric Gases, J. Atmos. Ocean. Technol., 26, 1814–1828, doi:10.1175/2009JTECHA1230.1, 2009.

Hase, F.: Inversion von Spurengasprofilen aus hochaufgelösten bodengebundenen FTIR-Messungen in Absorption, Dissertation, Wissenschaftliche Berichte Forschungszentrum Karlsruhe, Forschungszentrum Karlsruhe, Karlsruhe, Germany, 2000.

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. Spectrosc. Ra., 87, 24–52, 2004.

Ho, S.-P., Edwards, D. P., Gille, J. C., Chen, J., Ziskin, D., Francis, G. L., Deeter, M. N., and Drummond, J. R.: Estimates of 4.7 µm surface emissivity and their impact on the retrieval of tropospheric carbon monoxide by Measurements of Pollution in the Troposphere (MOPITT), J. Geophys. Res., 110, D21308, doi:10.1029/2005JD005946, 2005.

Hooghiemstra, P. B., Krol, M. C., Bergamaschi, P., de Laat, A. T. J., van der Werf, G. R., Novelli, P. C., Deeter, M. N., Aben, I., and Röckmann, T.: Comparing optimized CO emission estimates using MOPITT or NOAA surface network observations, J. Geophys. Res.-Atmos., 117, D06309, doi:10.1029/2011JD017043, 2012.

Jiang, Z., Jones, D. B. A., Worden, H. M., and Henze, D. K.: Sensitivity of top-down CO source estimates to the modeled vertical structure in atmospheric CO, Atmos. Chem. Phys., 15, 1521–1537, doi:10.5194/acp-15-1521-2015, 2015.

Kerzenmacher, T., Dils, B., Kumps, N., Blumenstock, T., Clerbaux, C., Coheur, P.-F., Demoulin, P., García, O., George, M., Griffith, D. W. T., Hase, F., Hadji-Lazaro, J., Hurtmans, D., Jones, N., Mahieu, E., Notholt, J., Paton-Walsh, C., Raffalski, U., Ridder, T., Schneider, M., Servais, C., and De Mazière, M.: Validation of IASI FORLI carbon monoxide retrievals using FTIR data from NDACC, Atmos. Meas. Tech., 5, 2751–2761, doi:10.5194/amt-5-2751-2012, 2012.

Lamarque, J. F., Emmons, L. K., Hess, P. G., Kinnison, D. E., Tilmes, S., Vitt, F., Heald, C. L., Holland, E. A., Lauritzen, P. H., Neu, J., Orlando, J. J., Rasch, P. J., and Tyndall, G. K.: CAM-chem: description and evaluation of interactive atmospheric chemistry in the Community Earth System Model, Geosci. Model Dev., 5, 369–411, doi:10.5194/gmd-5-369-2012, 2012.

Mahieu, E., Zander, R., Delbouille, L., Demoulin, P., Roland, G., and Servais, C.: Observed trends in total vertical column abundances of atmospheric gases from IR solar spectra recorded at the Jungfraujoch, J. Atmos. Chem., 28, 227–243, 1997.

Martínez-Alonso, S., Deeter, M. N., Worden, H. M., Gille, J. C., Emmons, L. K., Pan, L. L., Park, M., Manney, G. L., Bernath, P. F., Boone, C. D., Walker, K. A., Kolonjari, F., Wofsy, S. C., Pittman, J., and Daube, B. C.: Comparison of upper tropospheric carbon monoxide from MOPITT, ACE-FTS, and HIPPO-QCLS, J. Geophys. Res.-Atmos., 119, 14144–14164, doi:10.1002/2014JD022397, 2014.

Morgenstern, O., Zeng, G., Wood, S. W. W., Robinson, J., Smale, D., Paton-Walsh, C., Jones, N. B., and Griffith, D. W. T.: Long-range correlations in Fourier transform infrared, satellite, and modeled CO in the Southern Hemisphere, J. Geophys. Res., 117, D11301, doi:10.1029/2012JD017639, 2012.

MOPITT Science Team: MOPITT/Terra Level 2 CO vertical profiles derived from Thermal Infrared Radiances, version 6, Hampton, VA, USA:NASA Atmospheric Science Data Center (ASDC), doi:10.5067/TERRA/MOPITT/DATA202, 2013.

Notholt, J., Neuber, R., Schrems, O., and von Clarmann, T.: Stratospheric trace gas concentrations in the Arctic polar night derived by FTIR-spectroscopy with the moon as IR light source, Geophys. Res. Lett., 20, 2059–2062, 1993.

Pan, L., Gille, J. C., Edwards, D. P., Bailey, P. L., and Rodgers, C. D.: Retrieval of carbon monoxide for the MOPITT instrument, J. Geophys. Res., 103, 32277–32290, 1998.

Paton-Walsh, C., Jones, N. B., Wilson, S. R., Haverd, V., Meier, A., Griffith, D. W. T., and Rinsland, C. P.: Measurements of trace gas emissions from Australian forest fires and correlations with coincident measurements of aerosol optical depth, J. Geophys. Res., 110, D24305, doi:10.1029/2005JD006202, 2005.

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.

Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M. G., Schubert, S. D., Takacs, L., Kim, G.-K., Bloom, S., Chen, J., Collins, D., Conaty, A., da Silva, A., Gu, W., Joiner, J., Koster, R. D., Lucchesi, R., Molod, A., Owens, T., Pawson, S., Pegion, P., Redder, C. R., Reichle, R., Robertson, F. R., Ruddick, A. G., Sienkiewicz, M., and Woollen, J.: MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications, J. Climate, 24, 3624–3648, doi:10.1175/JCLI-D-11-00015.1, 2011.

Rinsland, C. P., Goldman, A., Hannigan, J. W., Wood, S. W., Chiou, L. S., and Mahieu, E.: Long-term trends of tropospheric carbon monoxide and hydrogen cyanide from analysis of high resolution infrared solar spectra, J. Quant. Spectrosc. Ra., 104, 40–51, doi:10.1016/j.jqsrt.2006.08.008, 2007.

Rodgers, C. D.: Inverse Methods for Atmospheric Sounding, Theory and Practice, World Scientific Publishing Co. Ptd. Ltd, New Jersey, 2000.

Rodgers, C. D. and Connor, B. J.: Intercomparison of remote sounding instruments, J. Geophys. Res.-Atmos., 108, 4116, doi:10.1029/2002JD002299, 2003.

Rothman, L. S., Gordon, I., Barbe, A., Benner, D. C., Bernath, P. F., Birk, M., Boudon, V., Brown, L. R., Campargue, A., Champion, J.-P., Chance, K., Coudert, L. H., Dana, V., Devi, V. M., Fally, S., Flaud, J.-M., Gamache, R. R., Goldman, I. A., Jacquemart, D., Kleiner, I., Lacome, N., Lafferty, W. J., Mandin, J. Y., Massie, S. T., Mikhailenko, S. N., Miller, C. E., Moazzen-Ahmadi, N., Naumenko, O. V., Nikitin, A. V., J., O., Perevalov, V. I., Perrin, A., Predoi-Cross, A., Rinsland, C. P., Rotger, M., Šimečková, M., Smith, M. A. H., Sung, K., Tashkun, S., Tennyson, J., Toth, R., Vandaele, A., and Vander Auwera, J.: The HITRAN 2008 molecular spectroscopic database, J. Quant. Spectrosc. Ra., 110, 533–572, 2009.

Schneider, M., Blumenstock, T., Chipperfield, M. P., Hase, F., Kouker, W., Reddmann, T., Ruhnke, R., Cuevas, E., and Fischer, H.: Subtropical trace gas profiles determined by ground-based FTIR spectroscopy at Izaña (28° N, 16° W): Five-year record, error analysis, and comparison with 3-D CTMs, Atmos. Chem. Phys., 5, 153–167, doi:10.5194/acp-5-153-2005, 2005.

Senten, C., De Mazière, M., Dils, B., Hermans, C., Kruglanski, M., Neefs, E., Scolas, F., Vandaele, A. C., Vanhaelewyn, G., Vigouroux, C., Carleer, M., Coheur, P. F., Fally, S., Barret, B., Baray, J. L., Delmas, R., Leveau, J., Metzger, J. M., Mahieu, E., Boone, C., Walker, K. A., Bernath, P. F., and Strong, K.: Technical Note: New ground-based FTIR measurements at Ile de La Réunion: observations, error analysis, and comparisons with independent data, Atmos. Chem. Phys., 8, 3483–3508, doi:10.5194/acp-8-3483-2008, 2008.

Sparling, L. C. and Bacmeister, J. T.: Scale dependence of tracer microstructure: PDFs, intermittency and the dissipation scale, Geophys. Res. Lett., 28, 2823–2826, doi:10.1029/2000GL012781, 2001.

Sussmann, R. and Buchwitz, M.: Initial validation of ENVISAT/SCIAMACHY columnar CO by FTIR profile retrievals at the Ground-Truthing Station Zugspitze, Atmos. Chem. Phys., 5, 1497–1503, doi:10.5194/acp-5-1497-2005, 2005.

Sussmann, R. and Schäfer, K.: Infrared spectroscopy of tropospheric trace gases: combined analysis of horizontal and vertical column abundances, Appl. Opt., 36, 735–741, doi:10.1364/AO.36.000735, 1997.

Velazco, V., Wood, S. W., Sinnhuber, M., Kramer, I., Jones, N. B., Kasai, Y., Notholt, J., Warneke, T., Blumenstock, T., Hase, F., Murcray, F. J., and Schrems, O.: Annual variation of strato-mesospheric carbon monoxide measured by ground-based Fourier transform infrared spectrometry, Atmos. Chem. Phys., 7, 1305–1312, doi:10.5194/acp-7-1305-2007, 2007.

Warner, J., Gille, J., Edwards, D. P., Ziskin, D., Smith, M., Bailey, P., and Rokke, L.: Cloud detection and clearing for the Earth observing system Terra satellite Measurements of Pollution in the Troposphere (MOPITT) experiment, Appl. Opt., 40, 1269–1284, 2001.

Wiacek, A., Taylor, J. R., Strong, K., Saari, R., and Kerzenmacher, T. E.: Ground-based solar absorption FTIR spectroscopy: Characterization of retrievals and first results from a novel optical design instrument at a new NDACC complementary station, J. Atmos. Oceann Technol., 24, 432–448, doi:10.1175/JTECH1962.1, 2007.

Worden, H. M., Deeter, M. N., Edwards, D. P., Gille, J. C., Drummond, J. R., and Nédélec, P.: Observations of near-surface carbon monoxide from space using MOPITT multispectral retrievals, J. Geophys. Res., 115, D18314, doi:10.1029/2010JD014242, 2010.

Worden, H. M., Deeter, M. N., Frankenberg, C., George, M., Nichitiu, F., Worden, J., Aben, I., Bowman, K. W., Clerbaux, C., Coheur, P. F., de Laat, A. T. J., Detweiler, R., Drummond, J. R., Edwards, D. P., Gille, J. C., Hurtmans, D., Luo, M., Martínez-Alonso, S., Massie, S., Pfister, G., and Warner, J. X.: Decadal record of satellite carbon monoxide observations, Atmos. Chem. Phys., 13, 837–850, doi:10.5194/acp-13-837-2013, 2013.

Yurganov, L. N., McMillan, W. W., Dzhola, A. V., Grechko, E. I., Jones, N. B., and van der Werf, G. R.: Global AIRS and MOPITT CO measurements: Validation, comparison, and links to biomass burning variations and carbon cycle, J. Geophys. Res., 113, D09301, doi:10.1029/2007JD009229, 2008.

Zeng, G., Wood, S., Morgenstern, O., Jones, N. B., Robinson, J., and Smale, D.: Trends and variations in CO, C2H6, and HCN in the Southern Hemisphere point to the declining anthropogenic emissions of CO and C2H6, Atmos. Chem. Phys., 12, 7543–7555, doi:10.5194/acp-12-7543-2012, 2012.