



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

Rebecca R. Buchholz¹, Merritt N. Deeter¹, Helen M. Worden¹, John Gille¹, David P. Edwards¹, James W. Hannigan¹, Nicholas B. Jones², Clare Paton-Walsh², David W. T. Griffith², Dan Smale³, John Robinson³, Kimberly Strong⁴, Stephanie Conway⁴, Ralf Sussmann⁵, Frank Hase⁶, Thomas Blumenstock⁶, Emmanuel Mahieu⁷, and Bavo Langerock⁸

¹Atmospheric Chemistry Observations & Modeling, National Center for Atmospheric Research, Boulder, Colorado, USA

²Centre for Atmospheric Chemistry, University of Wollongong, Australia

³National Institute of Water and Atmospheric Research, New Zealand

⁴Department of Physics, University of Toronto, Canada

⁵Karlsruhe Institute of Technology, IMK-IFU, Garmisch-Partenkirchen, Germany

⁶Karlsruhe Institute of Technology, IMK-ASF, Karlsruhe, Germany

⁷Institute of Astrophysics and Geophysics, University of Liège, Belgium

⁸Royal Belgian Institute for Space Aeronomy (BIRA-IASB), Belgium

Correspondence to: R. 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. 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 2.8% for TIR-only, 5.4% for TIR-NIR and 7.0% for NIR-only. The TIR-NIR and NIR-only products consistently produce greater 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, and rather is influenced by the proximity to rapidly changing atmospheric CO. MOPITT bias drift has been bound geographically to within $\pm 0.5\%$ 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 two months makes CO an ideal atmospheric constituent to track atmospheric physical and chemical processes over a range of spatial scales (Edwards et al., 2006; Duncan et al., 2007).

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 analyse 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., 2014, 2010; Emmons et al., 2009, 2004). Validation and intercomparison was 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. Also, 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 pre-date 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 three 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 smoothing of FTS measurements with MOPITT averaging kernels. 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 $\sim 22:30$ local time. MOPITT is nadir-viewing with a ground resolution of 22×22 km. Each of the four indium antimonide 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 ($4.7 \mu\text{m}$) is measured from terrestrial thermal infrared radiation (TIR) (Drummond, 1992; Drummond et al., 2010; Drummond and Mand, 1996; Deeter et al., 2003). More recently, incorporating the overtone band ($2.3 \mu\text{m}$) from reflected solar radiation in the near infrared (NIR) provides multi-spectral 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), joint thermal- and near infrared (TIR-NIR), and near infrared only (NIR-only). MOPITT products are publicly available through several repositories linked via <http://terra.nasa.gov/about/terra-instruments/mopitt>.

The retrieval process for MOPITT follows maximum likelihood optimal estimation (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 one degree 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 a priori climatology; 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 ten 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 $\log_{10}(\text{VMR})$, which means averaging kernel (AK) values are associated with $\log_{10}(\text{VMR})$ (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 pre-date the launch of MOPITT. The spectrometers used are 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 signal-to-noise ratio in the CO spectral region. The simultaneous use of spectral microwindows during trace gas retrieval further improves signal to noise 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 are used for this validation study. Retrieval code may be SFIT4 (10 stations), SFIT2 (2 stations; Pougatchev et al., 1995), or PROFFIT (2 stations; Hase, 2000). SFIT and PROFFIT retrieval codes were inter-compared 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). 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 mid-point 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 analyse 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 nonoperational in May 2001, as a result of cooler failure. The period before 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–February 2012. In contrast, TIR-only are validated between August 2001 and 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 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 $\sim 10:30$ a.m. (local time) are considered. While the MOPITT overpass also occurs at $\sim 10:30$ p.m., the daytime-only MOPITT measurements are used in order to include enhanced information 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 one degree 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, inversely weighted by the square of relative retrieval measurement error. Thus, one MOPITT average is compared with several FTS measurements. There are several advantages of averaging the MOPITT data. Combining satellite measurements within the one degree radius criterion satisfies a compromise between reducing the effects of random retrieval noise, and minimizing spatial dilution of measurements through using a small radius. 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. Additionally, averaging improves computational efficiency, reducing the number of comparisons at some stations from $\sim 40,000$ to $\sim 5,000$. Error-weighted averaging is also performed for the corresponding MOPITT AK matrices and a priori.

5 The instrument error is combined in quadrature. Depending upon the experiment, averages are restricted to include land-only scenes, water-only scenes, or specific detector elements pixels.

A comparison between the validation of averaged and raw MOPITT values against HIPPO and ACE-FTS measurements is available in the supplementary material of Martínez-Alonso et al. (2014). These authors concluded that the two methods produced equivalent results, and found that averaged MOPITT values produced bias of $<1.2\%$ against HIPPO and $<0.8\%$

10 against ACE-FTS, over raw MOPITT values. Therefore, averaging MOPITT values here could overestimate MOPITT biases by approximately 1%.

Temporal averaging is not performed on the FTS measurements, which enables a qualitative assessment of the influence of diurnal changes CO. In summary, the MOPITT spatial averages are compared separately with each FTS measurement in the same day.

15 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; (2) the FTS profile will be smoothed by the MOPITT AKs, which are reported on

20 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 mid-points 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

25 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, 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. This process is visualized in

30 Fig. 2 (a). 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 FTS surface pressure is smaller than the MOPITT surface pressure, the FTS values are not extrapolated outside the FTS surface pressure. These situations occur mainly for stations located near highly varying terrain or at high altitude. One possible method is to replace the missing lower values of the interpolated FTS profile with a scaled version of



the MOPITT a priori, such as in Kerzenmacher et al. (2012). However, seeing as land exists below the altitude of the FTS station, we choose an alternative method that uses the lowest level of the re-gridded FTS profile to define the lower bound of the comparison altitude range. In these cases, a new MOPITT column is calculated from a truncated profile to compare with the re-gridded FTS values. Figure 2 (b) shows the schematic of this process.

5 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 and Wollongong in Fig. 3; column 4), indicating relatively uniform sensitivity to the true atmospheric state and little inclusion of the a priori 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 in the retrievals, particularly for the lower altitude levels. Rodgers and Connor (2003) show that when intercomparing instruments, if one instrument possesses less dependence on the a priori 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 MOPITT retrieval space. Specifically, the vertically re-gridded FTS profile is smoothed by MOPITT AK matrices.

15 The MOPITT AK matrices are applied following Rodgers and Connor (2003), modified for $\log(\text{VMR})$:

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

where n is the number of vertical layers, $\mathbf{x}_{smooth} = \{x_{smooth,i} \mid i = 1 \dots n\}$ is the required smoothed VMR profile; $\mathbf{x}_{ap} = \{x_{ap,i} \mid i = 1 \dots n\}$ is the MOPITT a priori VMR profile; $\mathbf{A} = \{A_{i,j} \mid i = 1 \dots n, j = 1 \dots n\}$ is the MOPITT AK matrix; and $\mathbf{x}_{FTS} = \{x_{FTS,i} \mid i = 1 \dots n\}$ is the FTS re-gridded VMR profile. The \mathbf{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. 4. AKs are further discussed in Sect. 4.1, 4.2 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.

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

$$= x_{ap,i} \cdot \prod_{j=1}^n \left(\frac{x_{FTS,j}}{x_{ap,j}} \right)^{A_{ij}} \quad (3)$$

Smoothed FTS total column values (c_{smooth}) are calculated from \mathbf{x}_{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, had the FTS measurement described the true atmosphere.



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

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

MOPITT retrieved column (c_M) is then validated against the smoothed FTS column, for example by calculating the mean bias (Eq. 5).

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

where m is the number of comparisons at an NDACC station.

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
10 the trace of the AK matrix, which is influenced by instrumental and geophysical parameters (Rodgers, 2000; Deeter et al., 2015). DFS are provided for each MOPITT retrieval and are calculated for FTS measurements from the AK matrices.

The theoretical maximum DFS of each instrument may be higher than reported values because AKs depend on both the
retrieval methods and the choice of a priori, which are different between instruments. However, we aim to validate the op-
erational MOPITT products rather than perform instrument intercomparison, which means a comparison of retrieved DFS
15 between instruments is indicative of the information content differences between retrieved values.

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 and Ny-Ålesund).
20 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 Tables 3a, 3b and 3c. Results are for daytime land scenes, except for three stations where MOPITT land scenes
25 were sparse, and consequently use water scenes (IZA, MLO & LRN). Example correlation plots for each MOPITT product are
shown for Toronto and Wollongong in Fig. 5. Comparison plots for all stations are found in the supplementary material.

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 (\overline{bias}) of 2.8%. Biases are generally less than 10% for TIR-NIR (apart from at Ny-Ålesund and Bremen) producing \overline{bias} of 5.4%. NIR-only biases are less than 10% (except at Ny-Ålesund) with a \overline{bias} of 7.0%. Standard deviation is always larger than bias, except for the TIR-NIR product at Lauder and the NIR-only product at Lauder and Arrival Heights. Correlation values are generally the highest for the TIR-only product (\bar{r} : 0.85) compared to TIR-NIR (\bar{r} : 0.80) or NIR-only
5 (\bar{r} : 0.77). Average correlation values found here are lower by about 0.09 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 column AK seasonal variability. An example of the range of AK variability is shown by the normalized MOPITT column AKs 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
10 each station and found the maximum difference in station-wise bias between seasons was on average 4.8%, 4.5% and 3.2% for TIR-only, TIR-NIR and NIR-only respectively. Seasonal variation in bias is below the all-station average standard deviation for each product: 7.6% (TIR-only), 10.3% (TIR-NIR) and 9.4% (NIR-only). We conclude there is no significant seasonally dependent bias for MOPITT.

4.3 Surface-type specific validation

15 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). Also, thermal contrast between skin surface and the overlying air can affect MOPITT sensitivity to measuring CO. For instance, water scenes have lower thermal contrast, where skin surface and overlying air temperatures are similar. MOPITT has difficulty viewing the surface in
20 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. 4). 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
25 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 in the retrieval process, with a cost of higher variability (Deeter et al., 2011).

30 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 degree 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 one degree 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 Tables 3a, 3b and 3c. 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 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 one degree 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. 6. 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 6 indicates that for TIR-NIR and NIR-only, pixel 1 also consistently shows the highest bias. Pixel 3 generally shows 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 deviation 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 2–4 average results (Fig. 6). Therefore, MOPITT error diagnostics for TIR-only and TIR-NIR adequately account for the poor performance of pixel 1. In comparison for NIR-only, the all-pixel averages are systematically more biased than the pixel 2–4 averages. This suggests error characterization for the NIR-only product may need improving.

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 to first remove pixel 1 from the average. A more restrictive average would include the two best pixels for each dataset: pixels 3 and 4 for TIR-only and pixels 2 and 3 for TIR-NIR and 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

Satellite retrievals over colder surfaces at higher latitudes are challenging mainly due to low thermal contrast, resulting in a higher weighting to the a priori 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. 7 (top row), which shows how DFS decreases 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. 7).

The latitudinal dependence of MOPITT total column retrieval biases are 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. 7, 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 complimentary to HIPPO comparisons because FTS can validate MOPITT land and water scenes, compared to HIPPO measurements being taken only over water. The FTS also provide longer timeseries, while HIPPO are a set of 5 ~two-month campaigns taken between 2009–2011.

Supplementary material in Jiang et al. (2015) suggested MOPITT bias may be related to DFS. We find that although DFS vary strongly with latitude, the MOPITT bias does not depend upon latitude. 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.

In contrast, correlations are weakly determined by latitude, which suggests that DFS are related to correlation values (Fig. 7, bottom row). Two situations occur, with high correlations at high DFS values (TIR-NIR and the Northern Hemisphere in TIR-only), and high correlations at low DFS (Southern Hemisphere TIR-only). There are two mechanisms in the validation process that may be responsible for a relationship between DFS and Pearson's R. First, when smoothing the FTS comparison value by the MOPITT AK and a priori, at locations where DFS are lower, more a priori is included. In this case, strong correlations are expected for comparisons at low DFS. This effect is evident in the TIR-only Pearson's R versus latitude results for the Southern Hemisphere. A second influence is that improved correlation can be expected at higher DFS since more of the true atmospheric state is observed. This effect is shown in the Pearson's R versus latitude plot for TIR-NIR and the Northern Hemisphere of TIR-only.

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. 8). High biases in 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 MOPITT at 10:30 a.m.



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 throughout 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 a.m. 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 bias dependence on altitude. Comparisons with NIR-only include a large amount of a priori (mean DFS of 0.32), 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 $\pm 40^\circ$ (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. 9. Drift plots for other products are provided in supplementary material. Bias drift values are recorded in Tables 3a–3c.

Drift ranges between -0.91 and $0.33\% \text{ yr}^{-1}$ (TIR-only), -1.2 and $0.87\% \text{ yr}^{-1}$ (TIR–NIR) and -1.1 and $0.59\% \text{ yr}^{-1}$ (NIR-only). TIR mean drift ($0.003\% \text{ yr}^{-1}$) is lower than Deeter et al. (2014) who determined a drift of $0.003 \times 10^{17} \text{ molec. cm}^{-2} \text{ yr}^{-1}$ (approximately $0.15\% \text{ yr}^{-1}$). The TIR–NIR mean drift ($-0.14\% \text{ yr}^{-1}$) is about equal magnitude, but opposite sign to Deeter et al. (2014) (approximately $0.15\% \text{ yr}^{-1}$). NIR-only mean drift ($-0.17\% \text{ yr}^{-1}$) is equivalent to Deeter et al. (2014) (approximately $-0.15\% \text{ yr}^{-1}$).

The geographical relationship of the bias drift is shown in Fig. 10. 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 $-1\% \text{ yr}^{-1}$ 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 (Mauna Loa, Bremen). Other stations with high bias



occur at high latitudes in the Northern Hemisphere, where potential uncharacterized surface errors contribute to retrieval drift. Ho et al. (2005) found 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.

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 2.8% for TIR-only, 5.4% for TIR-NIR and 7.0% for NIR-only. MOPITT retrieves with equivalent skill over land or water, although the most information is present in the joint TIR-NIR land product as indicated by 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 processing of the level 3 product may need to be revised. We find no dependence of bias on latitude, suggesting no relationship to DFS. In contrast, latitude-dependent information content had a weak relationship to correlation results. Variability in lower tropospheric CO influences MOPITT bias, which is probably due to sampling/sensitivity differences between instruments. MOPITT bias found here may be used to account for satellite error prior to data assimilation, allowing for inclusion of data over a wider spatial range than is currently used. The MOPITT long-term bias drift has been bound to within $\pm 0.5\%$ yr⁻¹ or lower at almost all locations. Variable drift in the Northern Hemisphere implies an uncharacterized retrieval parameter such as uncertainty in cloud detection or sea-ice representation. We recommend trend analysis should not be performed above 60°N. Overall, this study extends the geographical and temporal analysis of MOPITT validation results.

Acknowledgements. The National Center for Atmospheric Research (NCAR) is sponsored by the National Science Foundation (NSF). Any opinions, findings and conclusions or recommendations expressed in the publication are those of the author(s) and do not necessarily reflect the views of the NSF. The NCAR MOPITT project is supported by the National Aeronautics and Space Administration (NASA) Earth Observing System (EOS) Program. The MOPITT team also acknowledges support from the Canadian Space Agency (CSA), the Natural Sciences and Engineering Research Council (NSERC) and Environment Canada, along with the contributions of COMDEV (the prime contractor) and ABB BOMEM. MOPITT data sets used in this study are publicly available at <http://reverb.echo.nasa.gov> and at <https://eosweb.larc.nasa.gov/datapool>. We thank NCAR internal reviewers Benjamin Gaubert and Ivan Ortega for their valuable suggestions.

The FTS data used in this publication were obtained as part of the Network for the Detection of Atmospheric Composition Change (NDACC), publicly available from <http://www.ndacc.org>. We acknowledge the numerous colleagues who contribute to FTS data acquisition



and the many agencies that support the facilities, retrievals and data analysis. **Eureka:** The Polar Environment Atmospheric Research Laboratory (PEARL) is implemented by the Canadian Network for the Detection of Atmospheric Change (CANDAC), led by James R. Drummond and in part by the Canadian Arctic ACE/OSIRIS Validation Campaigns, led by Kaley A. Walker, Support is provided by AIF/NSRIT, CFI, CFCAS, CSA, EC, GOC-IPY, NSERC, NSTP, OIT, PCSP and ORF. Logistical and operational support is provided by PEARL Site Manager Pierre Fogal, CANDAC operators, and the EC Weather Station. **Thule and Mauna Loa:** The NCAR operated FTS are supported by NASA. The Thule project is also supported by the NSF Office of Polar Programs (OPP) and the Danish Meteorological Institute. **Kiruna:** Operational support is provided by Uwe Raffalski and Peter Völger. **Jungfraujoch:** Supported by the F.R.S.-FNRS and the Fédération Wallonie-Bruxelles; GAW-CH program of MeteoSwiss, and the International Foundation High Altitude Research Stations Jungfraujoch and Gornergrat (HFSJG, Bern). **Toronto:** FTS at the University of Toronto Atmospheric Observatory (TAO) is supported by CFCAS, ABB Bomem, CFI, CSA, EC, NSERC, ORDCE, PREA, and the University of Toronto. Data analysis at Toronto (and Eureka) is supported by the CAFTON project, funded by the CSA's FAST Program. **Izaña:** We thank O. E. García and E. Sepúlveda, who are contracted by the Meteorological State Agency of Spain (AEMET). **La Réunion:** Supported by the Université de La Réunion and CNRS (LACy-UMR8105 and UMS3365). **Wollongong:** supported by the Australian Research Council for many years, most recently by grants DP160101598 and LE0668470. **Lauder and Arrival Heights:** The NIWA MIR-FTS programme is core-funded through New Zealand's Ministry of Business, Innovation and Employment. We thank Antarctica New Zealand and the Scott Base staff for providing logistical and technical support at Arrival Heights.



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, 10,501–10,529, doi:10.1002/2015JD023467, <http://dx.doi.org/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 and microwave 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, <http://www.atmos-chem-phys.net/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, <http://doi.wiley.com/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 observations, *Atmos. Meas. Tech.*, 3, 1457–1471, doi:10.5194/amt-3-1457-2010, <http://www.atmos-meas-tech.net/3/1457/2010/amt-3-1457-2010.html>, 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, <http://www.atmos-meas-tech.net/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, <http://doi.wiley.com/10.1029/2002JD003186>, 2003.
- 5 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, D03 308, doi:10.1029/2003JD003970, <http://doi.wiley.com/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, D04 303, doi:10.1029/2008JD010872, <http://doi.wiley.com/10.1029/2008JD010872>, 2009.
- 10 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, D07 306, 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.
- 15 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, <http://doi.wiley.com/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, <http://www.atmos-meas-tech.net/7/3623/2014/amt-7-3623-2014.html>, 2014.
- 20 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.*, doi:10.1002/2015JD024024, <http://doi.wiley.com/10.1002/2015JD024024>, 2015.
- Drummond, J. R.: Measurements of pollution in the troposphere (MOPITT), *Use of EOS for Studies of Atmospheric Physics*, pp. 77–101, 25 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, *Advances in Space Research*, 45, 760–774, doi:10.1016/j.asr.2009.11.019, 2010.
- 30 Duncan, B. N., Logan, J. A., Bey, I., Megretskaja, 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, *Journal of Geophysical Research: Atmospheres*, 112, doi:10.1029/2007JD008459, <http://dx.doi.org/10.1029/2007JD008459>, d22301, 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, 16 755–16 775, 1999.
- 35 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, D14 312, 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, doi:10.1029/2003JD004101, <http://dx.doi.org/10.1029/2003JD004101>, d03309, 2004.
- 5 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, <http://www.atmos-chem-phys.net/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, D09 301, doi:10.1029/2006JD007485, 2007.
- 10 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.*, doi:10.1002/2016JD024863, <http://dx.doi.org/10.1002/2016JD024863>, 2016JD024863, 2016.
- 15 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. Oceanic Technol.*, 26, 1814–1828, doi:10.1175/2009JTECHA1230.1, 2009.
- 20 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, D21 308, doi:10.1029/2005JD005946, <http://doi.wiley.com/10.1029/2005JD005946>, 2005.
- 25 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, doi:10.1029/2011JD017043, <http://dx.doi.org/10.1029/2011JD017043>, d06309, 2012.
- 30 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, <http://www.atmos-chem-phys.net/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, <http://www.atmos-meas-tech.net/5/2751/2012/amt-5-2751-2012.html>, 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.
- 5 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.*, doi:10.1002/2014JD022397, <http://doi.wiley.com/10.1002/2014JD022397>, 2014.
- 10 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, D11 301, doi:10.1029/2012JD017639, 2012.
- 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.
- 15 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, 32 277–32 290, 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, D24 305, doi:10.1029/2005JD006202, 2005.
- 20 Pougatchev, N. S., Connor, B. J., and Rinsland, C. P.: Infrared measurements of the ozone vertical distribution above Kitt Peak, *J. Geophys. Res.*, 100, 16 689–16 697, 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
- 25 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,
- 30 2000.
- Rodgers, C. D. and Connor, B. J.: Intercomparison of remote sounding instruments, *J. Geophys. Res.-Atmos.*, 108, 4116, 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.,
- 35 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., Reddman, 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, <http://www.atmos-chem-phys.net/8/3483/2008/acp-8-3483-2008.html>, 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, <http://dx.doi.org/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. Oceanic Tech.*, 24, 432–448, doi:10.1175/JTRCH1962.1, 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, D18 314, doi:10.1029/2010JD014242, <http://doi.wiley.com/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., Martinez-Alonso, S., Massie, S., Pfister, G., and Warner, J. X.: Decadal record of satellite carbon monoxide observations, *Atmos. Chem. Phys.*, 13, 837–850, 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, D09 301, doi:10.1029/2007JD009229, <http://doi.wiley.com/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.



Tables

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. (masl)	Lat.	Lon.	Observation Period	Instrument †	Code ‡	Site reference
<i>Northern Hemisphere</i>							
Eureka, Canada (EUR)	610	80.05° N	82.42° W	2006–2013	Br125	SFIT4	Batchelor et al. (2009)
Ny-Ålesund, Norway (NYA)	15	78.92° N	11.93° E	1992–2013	Br120	SFIT4	Notholt et al. (1993)
Thule, Greenland (THU)	225	76.53° N	68.74° W	1999–2013	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)
Jungfrauoch, 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	2002–2013	BoDA8	SFIT4	Wiacek et al. (2007)
Izaña, Spain (IZA)	2367	28.30° N	16.48° W	1999–2005	Br120M	PRO	Schneider et al. (2005)
				2005–2012	Br125		
Mauna Loa, USA (MLO)	3397	19.54° N	155.58° W	1995–2007	Br120M	SFIT4	David et al. (1993)
				2007–2012	Br125		
<i>Southern Hemisphere</i>							
La Réunion ^a (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)
				2008–2014	Br125		
Lauder, New Zealand (LAU)	370	45.04° S	169.68° E	2001–2014	Br120	SFIT4	Morgenstern et al. (2012)
Arrival Heights (AHS)	250	77.82° S	166.65° E	1997–2014	Br120M	SFIT4	Zeng et al. (2012)

† BoDA8 = Bomem DA8, Br120 = Bruker 120HR, Br120M = Bruker 120M, Br125 = Bruker 125 HR

‡ 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 Jungfrauoch is v3.91, and at Zugspitze V3.90i; PRO = PROFFIT96.

^a Data investigated is only from the St Denis station, and not from Maito.



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, surface to ~ 26 hPa.

Site	MOPITT		MOPITT		MOPITT	FTS
	TIR-only		TIR-NIR		NIR-only	
	land	water	land	water		
EUR	0.63	0.66	1.20	0.95	0.49	2.53
NYA	0.69	0.91	1.11	1.23	0.18	1.73
THU	0.48	0.76	1.05	1.06	0.41	2.70
KIR	0.90	~	1.35	~	0.42	2.60
BRE	1.16	~	1.69	~	0.53	1.99
ZUG	1.03	~	1.50	~	0.33	1.77
JFJ	1.05	~	1.52	~	0.31	2.15
TAO	1.12	0.94	1.60	1.32	0.34	2.92
IZA	~	1.19	~	1.58	~	2.62
MLO	~	1.27	~	1.66	~	2.46
LRN	~	1.26	~	1.64	~	1.90
WOL	1.04	1.15	1.46	1.55	0.10	2.55
LAU	0.91	~	1.30	~	0.20	2.80
AHS	0.30	0.48	0.77	0.81	0.16	3.44
mean	0.85	0.96	1.32	1.31	0.32	2.44

~ indicates where not enough pixels of that surface-type were available for analysis.



Table 3a. Site-wise validation results for land pixels 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 MOPITT values; units of drift are 10^{17} molec. cm^{-2} yr^{-1} or $\% \text{yr}^{-1}$, shown with standard error. Bold text indicates significant drift where $p < 0.01$, determined from a two-tailed t-test.

Version		EUR	NYA	THU	KIR
TIR-only	bias	0.10	0.80	-0.38	0.20
	% bias	0.59	4.23	-2.19	1.00
	% SD	10.54	7.09	7.85	6.39
	r	0.61	0.85	0.78	0.84
	drift	0.054±0.022	-0.013±0.015	0.045±0.010	-0.016±0.015
	drift ($\% \text{yr}^{-1}$)	0.33±0.14	-0.07±0.08	0.25±0.06	-0.08±0.07
	# obs	1139	580	1720	695
TIR-NIR	bias	0.57	2.52	0.68	0.93
	% bias	3.36	12.15	3.58	4.50
	% SD	14.73	15.85	13.19	7.65
	r	0.67	0.66	0.66	0.83
	drift	-0.013±0.059	-0.125±0.050	0.166±0.024	-0.060±0.021
	drift ($\% \text{yr}^{-1}$)	-0.08±0.34	-0.60±0.24	0.87±0.12	-0.29±0.10
	# obs	944	482	1461	646
NIR-only	bias	0.92	2.26	1.21	1.52
	% bias	5.27	10.91	6.17	7.41
	% SD	8.08	15.34	10.03	8.32
	r	0.91	0.65	0.78	0.83
	drift	0.102±0.032	-0.223±0.047	0.083±0.019	-0.063±0.022
	drift ($\% \text{yr}^{-1}$)	0.59±0.19	-1.08±0.23	0.43±0.10	-0.31±0.11
	# obs	948	473	1463	646



Table 3b. Same as Table 3a but for Northern Hemisphere, low- and mid-latitude sites. Results are for land pixels, except where the majority of pixels are water (w).

Version		BRE	ZUG	JFJ	TAO	IZA (w)	MLO (w)
TIR-only	bias	1.72	-0.11	0.15	1.72	-0.13	0.04
	% bias	7.88	-0.77	1.32	7.37	-0.96	0.37
	% SD	8.52	6.67	9.54	8.64	6.98	5.92
	r	0.86	0.89	0.75	0.78	0.87	0.95
	drift	-0.200±0.032	0.003±0.004	0.028±0.007	-0.026±0.021	-0.006±0.007	0.060±0.011
	drift (% yr ⁻¹)	-0.91±0.15	0.02±0.03	0.24±0.06	-0.11±0.09	-0.04±0.06	0.54±0.10
	# obs	324	5646	1600	717	942	177
TIR-NIR	bias	2.34	0.004	0.22	2.07	-0.08	-0.14
	% bias	10.26	0.03	1.88	8.78	-0.59	-1.35
	% SD	10.25	8.31	12.34	9.90	5.89	6.68
	r	0.80	0.87	0.66	0.77	0.91	0.93
	drift	-0.283±0.069	-0.009±0.006	0.064±0.015	-0.009±0.037	0.007±0.011	0.102±0.017
	drift (% yr ⁻¹)	-1.24±0.30	-0.06±0.04	0.55±0.11	-0.04±0.16	0.05±0.08	0.95±0.16
	# obs	255	4558	1257	512	599	109
NIR-only	bias	1.33	0.78	0.68	1.19	~	~
	% bias	6.23	5.17	5.57	5.04	~	~
	% SD	8.95	13.31	10.36	6.09	~	~
	r	0.84	0.59	0.62	0.77	~	~
	drift	-0.062±0.058	-0.034±0.010	-0.088±0.011	0.006±0.023	~	~
	drift (% yr ⁻¹)	-0.29±0.27	-0.23±0.07	-0.72±0.09	0.02±0.10	~	~
	# obs	255	4559	1259	512	~	~

(w) denotes sites where water pixels were more representative and therefore there are no results for NIR retrievals, indicated by ~.



Table 3c. Same as Table 3b but for Southern Hemisphere sites.

Version		LRN (w)	WOL	LAU	AHS
TIR-only	bias	0.58	0.86	0.84	0.14
	% bias	4.29	6.41	7.71	1.57
	% SD	6.45	9.40	6.74	5.21
	r	0.94	0.88	0.93	0.95
	drift	-0.009±0.015	0.022±0.005	0.017±0.005	0.007±0.005
	drift (% yr ⁻¹)	-0.06±0.11	0.16±0.04	0.15±0.05	0.07±0.06
	# obs	503	5729	925	473
TIR-NIR	bias	0.63	1.21	1.27	0.77
	% bias	4.57	8.72	11.09	8.40
	% SD	7.10	11.68	8.09	13.15
	r	0.94	0.83	0.93	0.73
	drift	-0.045±0.023	-0.023±0.009	-0.009±0.014	-0.036±0.023
	drift (% yr ⁻¹)	-0.33±0.16	-0.16±0.06	-0.08±0.12	-0.40±0.25
	# obs	376	3991	570	297
NIR-only	bias	~	0.49	1.21	1.06
	% bias	~	3.60	10.23	10.84
	% SD	~	4.13	7.44	10.82
	r	~	0.95	0.85	0.70
	drift	~	-0.006±0.003	-0.005±0.013	-0.019±0.020
	drift (% yr ⁻¹)	~	-0.05±0.02	-0.04±0.11	-0.20±0.20
	# obs	~	4002	570	296

(w) denotes sites where water pixels were more representative and therefore there are no results for NIR retrievals, indicated by ~.



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 is in Tables 3a, 3b and 3c. Mean bias is expressed as percent relative to MOPITT 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	-0.18	4.85	-3.24	8.64	6.29	4.21
	% SD	9.06	6.81	8.38	13.14	9.46	5.78
	r	0.70	0.92	0.75	0.79	0.86	0.92
	drift (% yr ⁻¹)	-0.94±0.14	0.01±0.07	0.50±0.07	-0.23±0.15	-0.12±0.04	0.03±0.07
	# obs	856	620	1703	608	4832	347
TIR-NIR water	% bias	1.83	5.41	-4.04	9.24	7.28	5.33
	% SD	13.34	8.00	10.80	14.12	11.68	9.58
	r	0.49	0.89	0.68	0.74	0.85	0.84
	drift (% yr ⁻¹)	-1.20±0.35	-0.24±0.12	0.86±0.10	-0.01±0.24	0.05±0.06	-0.11±0.21
	# obs	740	491	1471	445	3356	229



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	0.70	0.81	1.00	0.64	0.87	0.85
	SD	1.22	0.99	0.76	0.84	0.76	0.74
	r	0.80	0.88	0.91	0.90	0.93	0.93
TIR-NIR	bias	1.15	1.47	1.48	0.95	1.31	1.27
	SD	1.54	1.08	1.00	1.06	0.99	0.93
	r	0.81	0.90	0.90	0.89	0.91	0.93
NIR-only	bias	1.84	1.30	0.90	0.83	1.01	1.21
	SD	1.93	1.04	0.86	1.09	0.70	0.88
	r	0.64	0.83	0.85	0.71	0.89	0.86



Figures

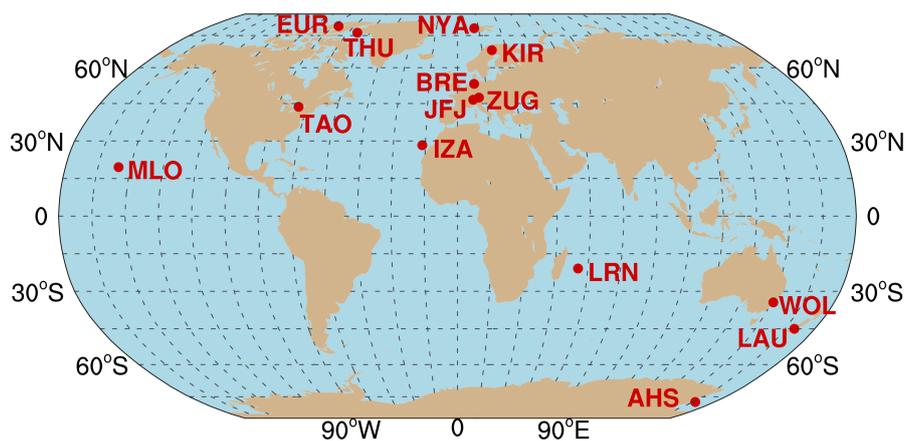


Figure 1. Location of the 14 NDACC ground-based remote-sensing FTS sites used in this study. Three-letter acronyms correspond to the information in Table 1.

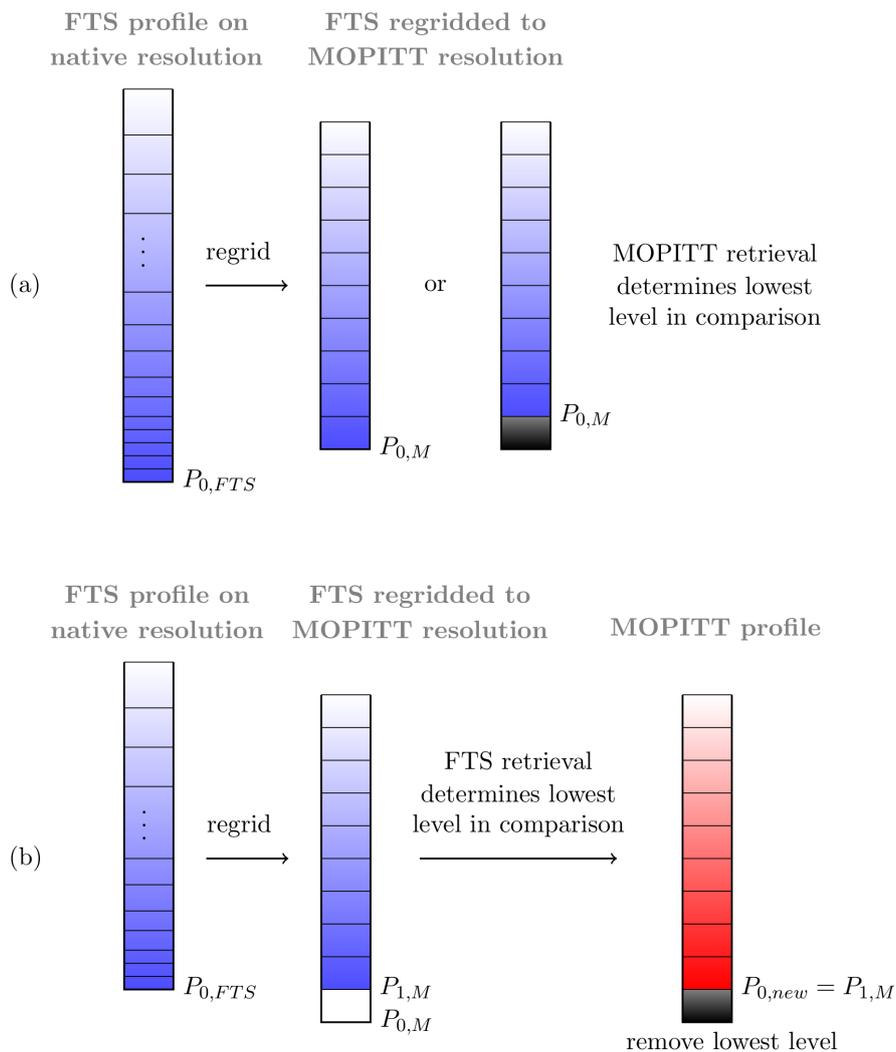


Figure 2. Interpolating FTS altitude levels to match MOPITT levels in the case of (a) FTS surface pressure being greater than MOPITT surface pressure; and (b) FTS surface pressure being less than MOPITT surface pressure.

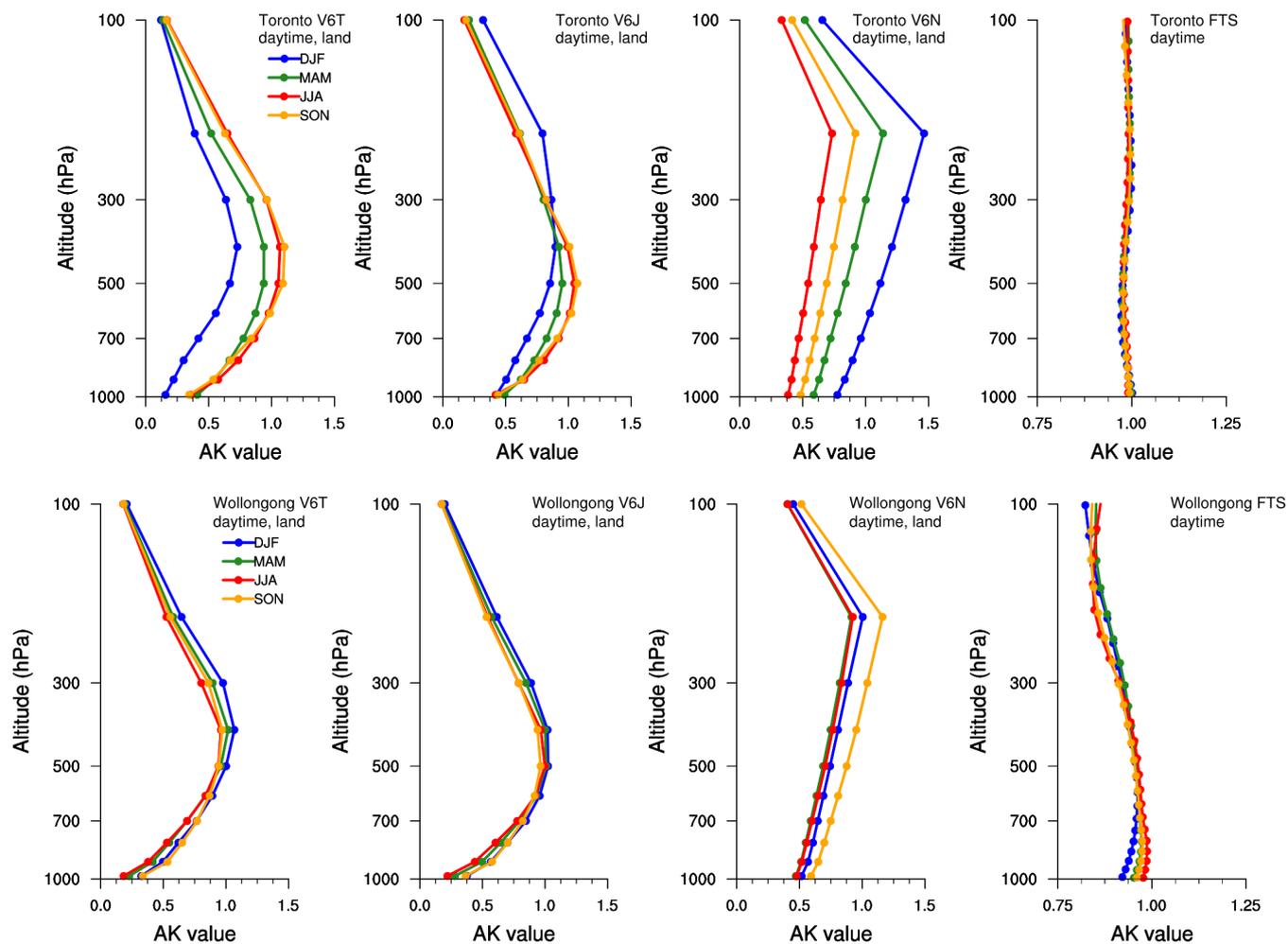


Figure 3. Seasonal mean MOPITT and FTS column AKs at Northern Hemisphere station Toronto (top row) and Southern Hemisphere station Wollongong (bottom row). Columns correspond to data products: column 1 = TIR-only; column 2 = TIR-NIR; column 3 = NIR-only; and column 4 = FTS. The MOPITT AKs have been normalised by the dataset mean. Note the x-axis scale for the FTS AKs is zoomed.

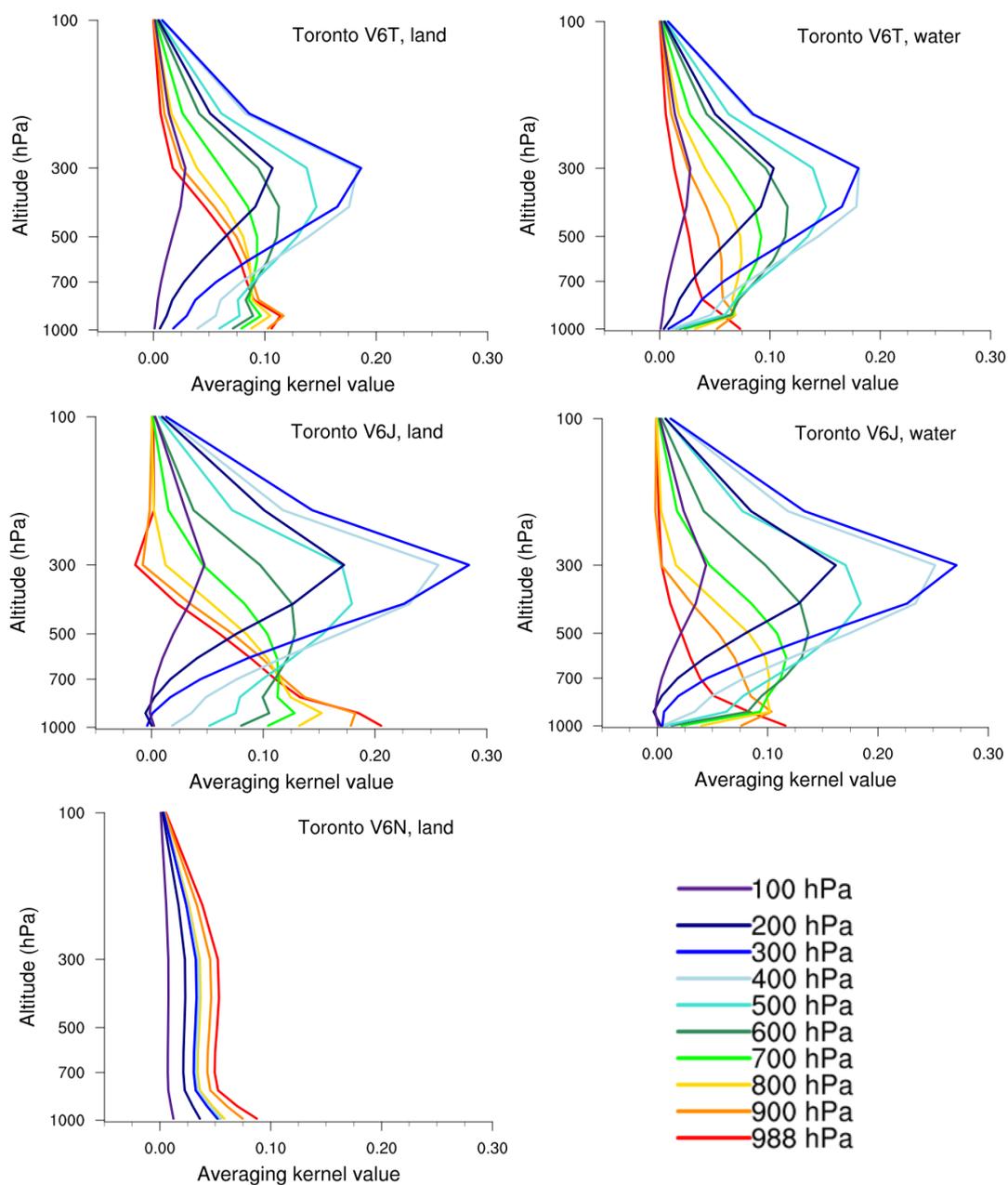


Figure 4. Mean MOPITT AKs at Toronto.

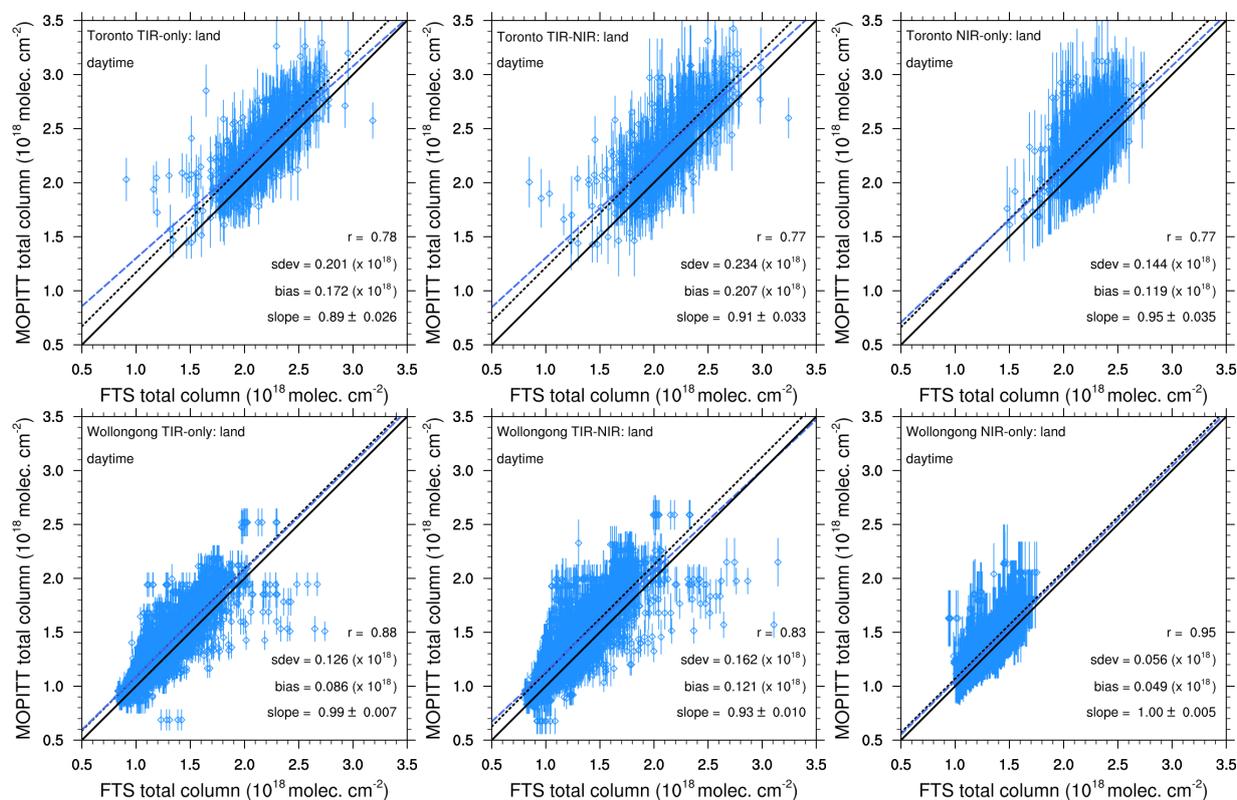


Figure 5. Example of the correlation plots at Toronto (top row) and Wollongong (bottom row). Daytime, land-only pixels are shown for each retrieval type TIR-only, TIR-NIR and NIR-only. Black dotted line indicates mean bias and blue dashed line indicates correlation regression line. For reference, the 1:1 line is indicated in solid black. Error bars are the error for MOPITT values, combined in quadrature from the pixels within a 1 deg radius of the station. Column amounts are molecules/cm².

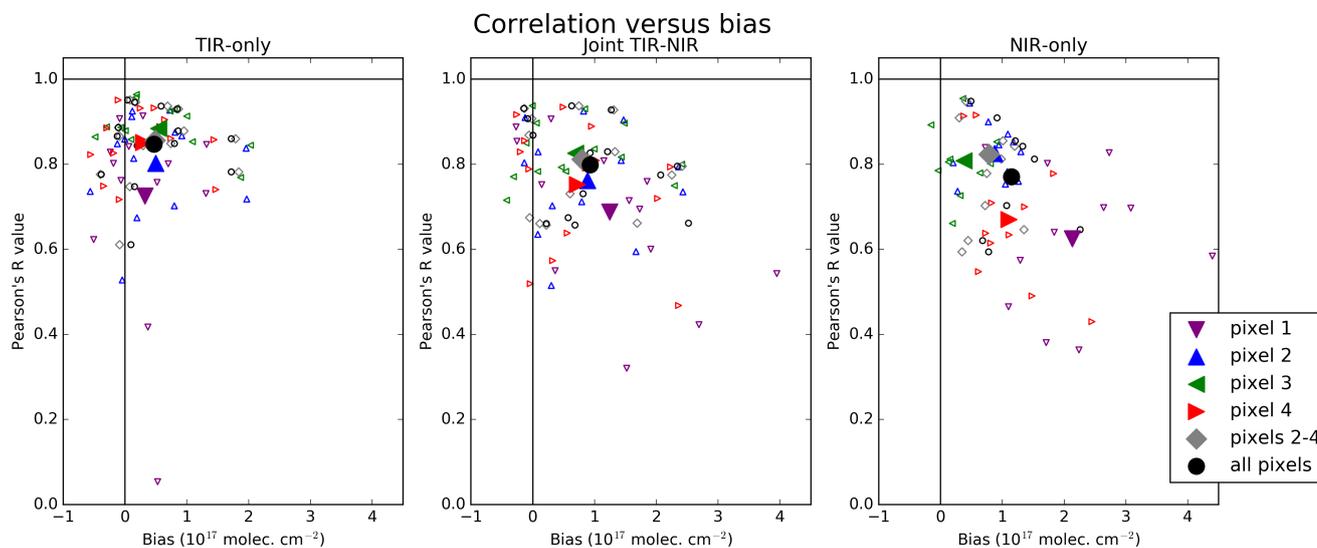


Figure 6. Pixel-wise validation results (correlation and bias) for each retrieval type at all FTS sites. Open markers represent results at each station and filled markers represent the mean value of all stations. Solid lines indicate zero bias and unity correlation. Columns are for the different MOPITT data products - column 1: TIR-only, column 2: TIR-NIR, and column 3: NIR-only.

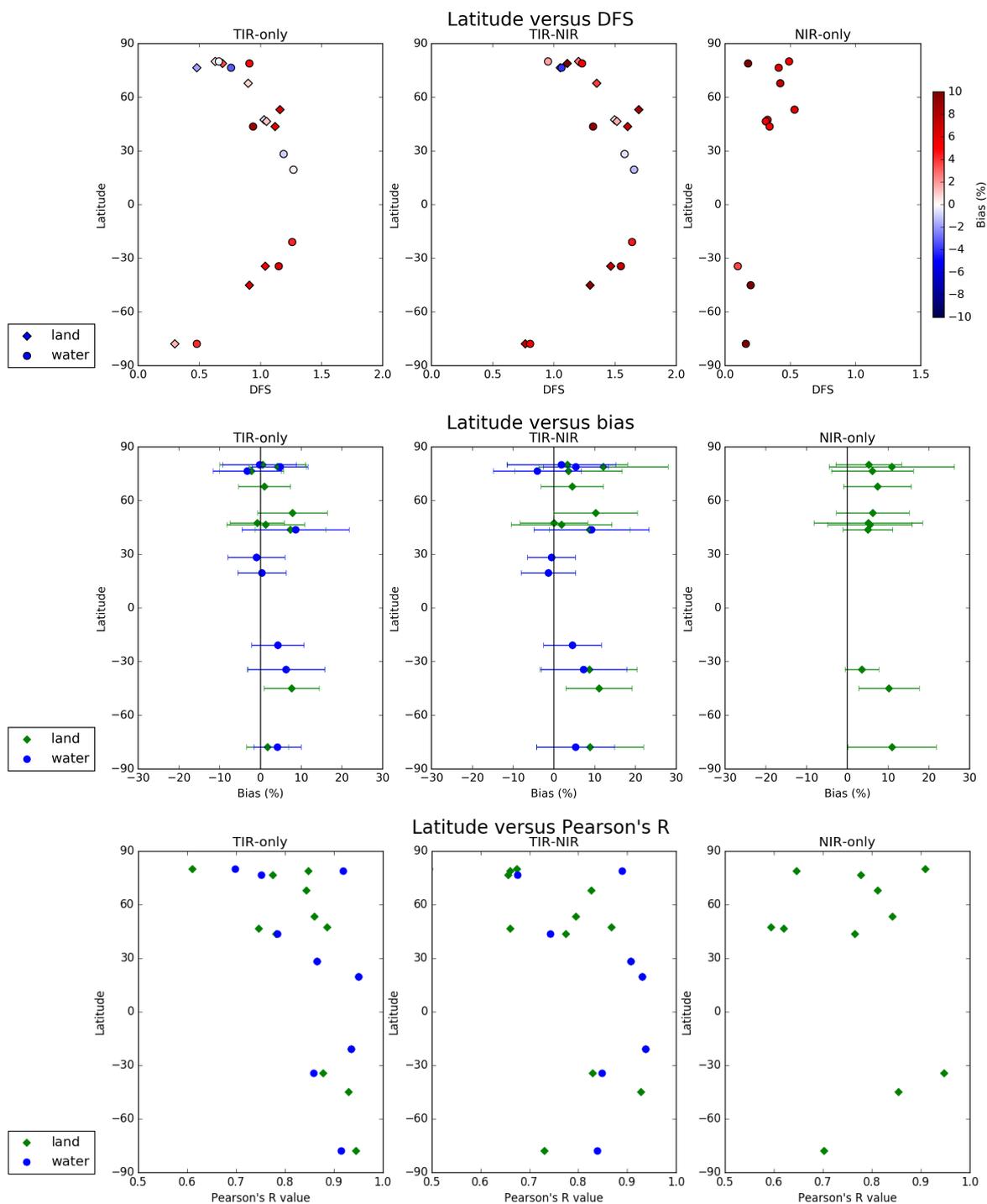


Figure 7. Latitudinal relationship with DFS (top row), bias (second row), and Pearson's R correlation value (bottom row). Error bars in bias plot denote standard deviations. Columns are for the different MOPITT data products - column 1: TIR-only, column 2: TIR-NIR, and column 3: NIR-only.

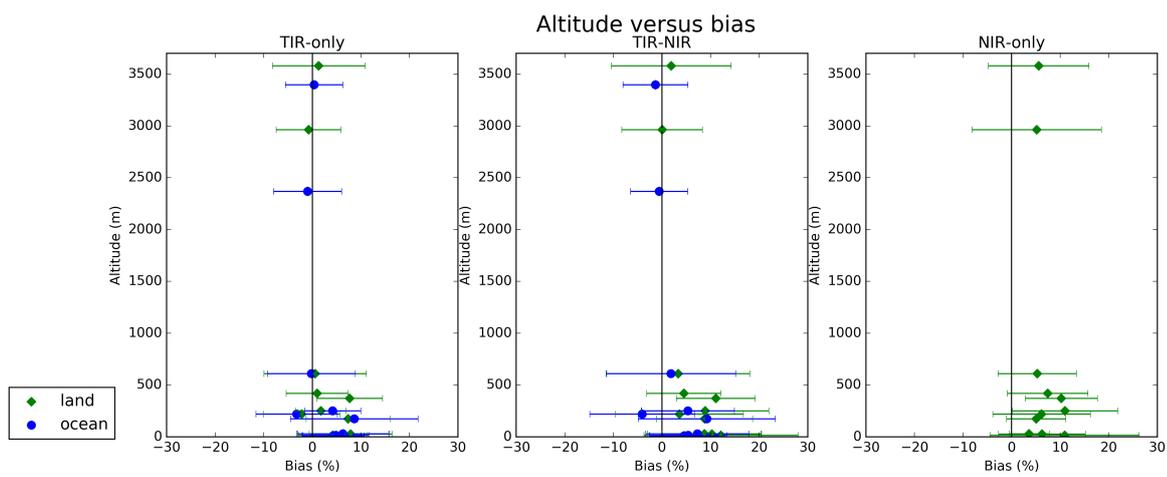


Figure 8. Altitude dependence of MOPITT retrieval bias. Error bars in bias plot denote standard deviations. Columns are for the different MOPITT data products - column 1: TIR-only, column 2: TIR-NIR, and column 3: NIR-only.

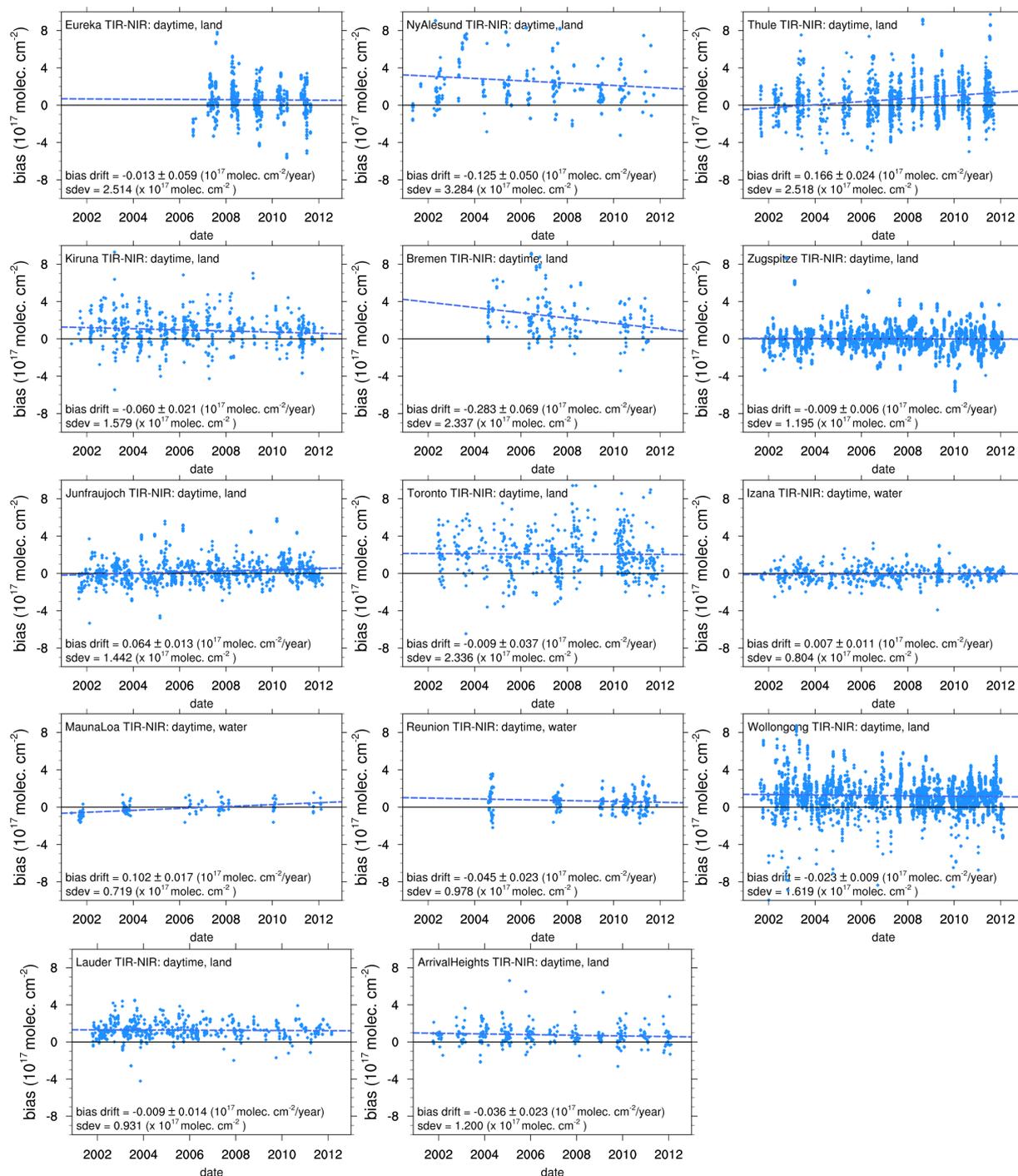


Figure 9. Temporal evolution of bias (MOPITT-FTS) at all stations for V6 joint TIR-NIR product. Blue dashed line indicates bias drift calculated from a linear least squares regression. Plots are ordered by latitude, from north to south.

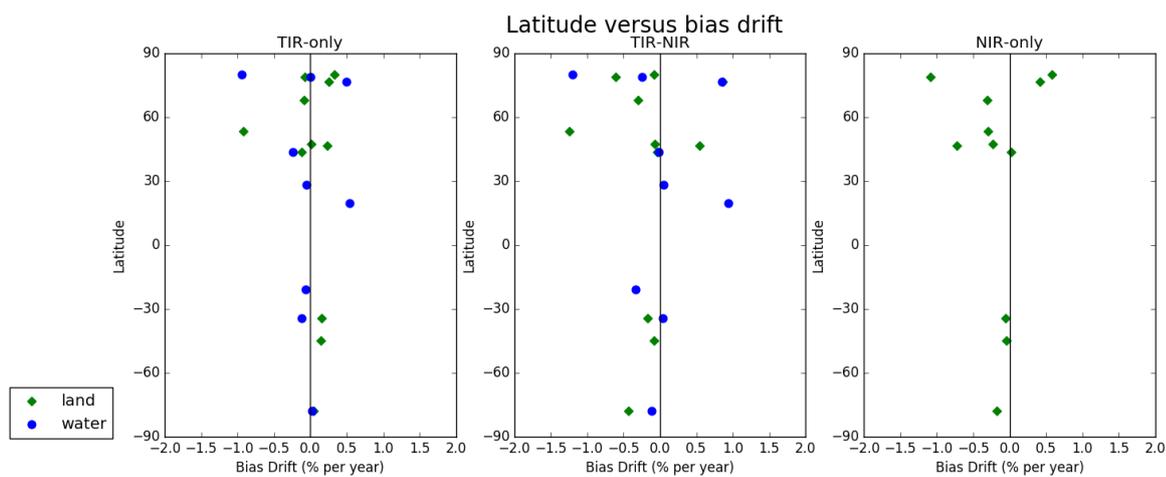


Figure 10. Geographic variability in bias drift. Columns are for the different MOPITT data products - column 1: TIR-only, column 2: TIR-NIR, and column 3: NIR-only.



Appendix A: Supplementary Figures

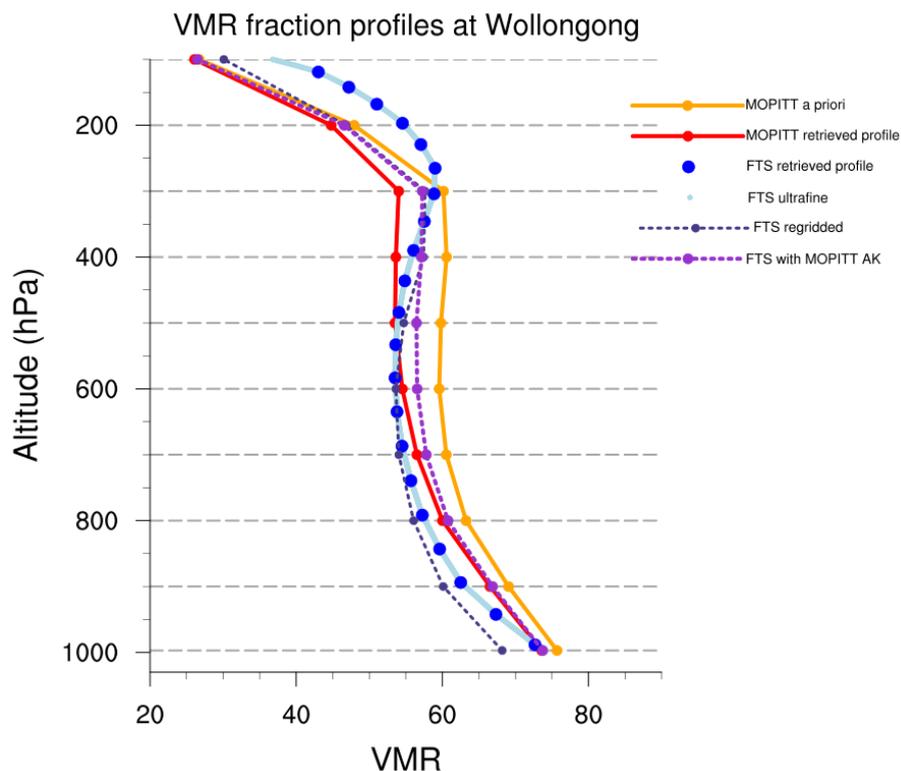


Figure 11. An example of the smoothing technique applying MOPITT TIR-only AKs to a FTS profile (blue circles) at Wollongong. The FTS profile is interpolated in logP space to an ultrafine grid of 100 levels per MOPITT layer (light blue), then averaged within each MOPITT layer (steel blue) and smoothed by the MOPITT AK (purple) to be commensurate with MOPITT values (red circles). Dashed grey lines indicate MOPITT level edges.

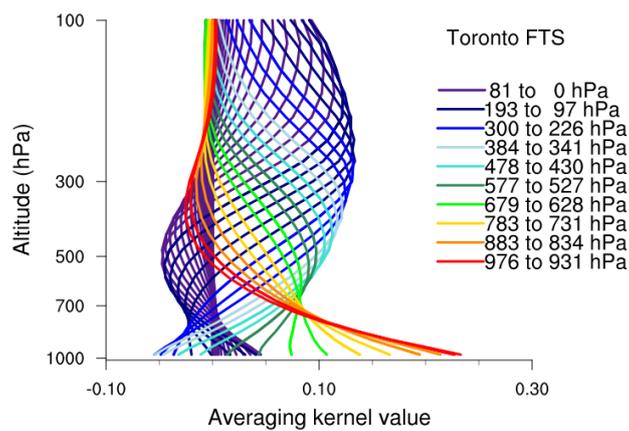


Figure 12. Mean FTS AKs at Toronto. Colors approximately correspond to MOPITT layers.

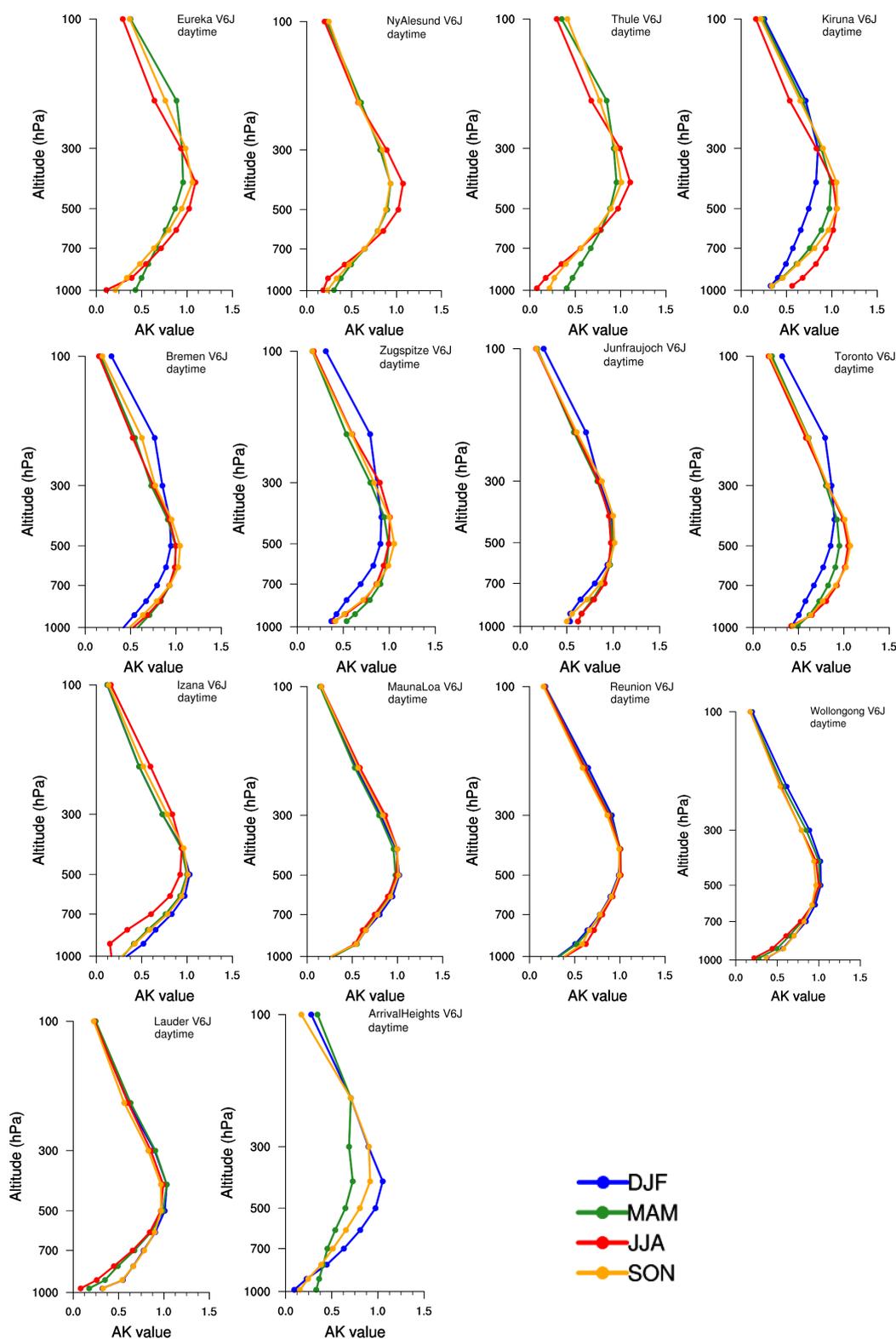


Figure 13. Mean MOPITT seasonal column AKs at all stations for the TIR-NIR product. Stations are ordered by latitude.

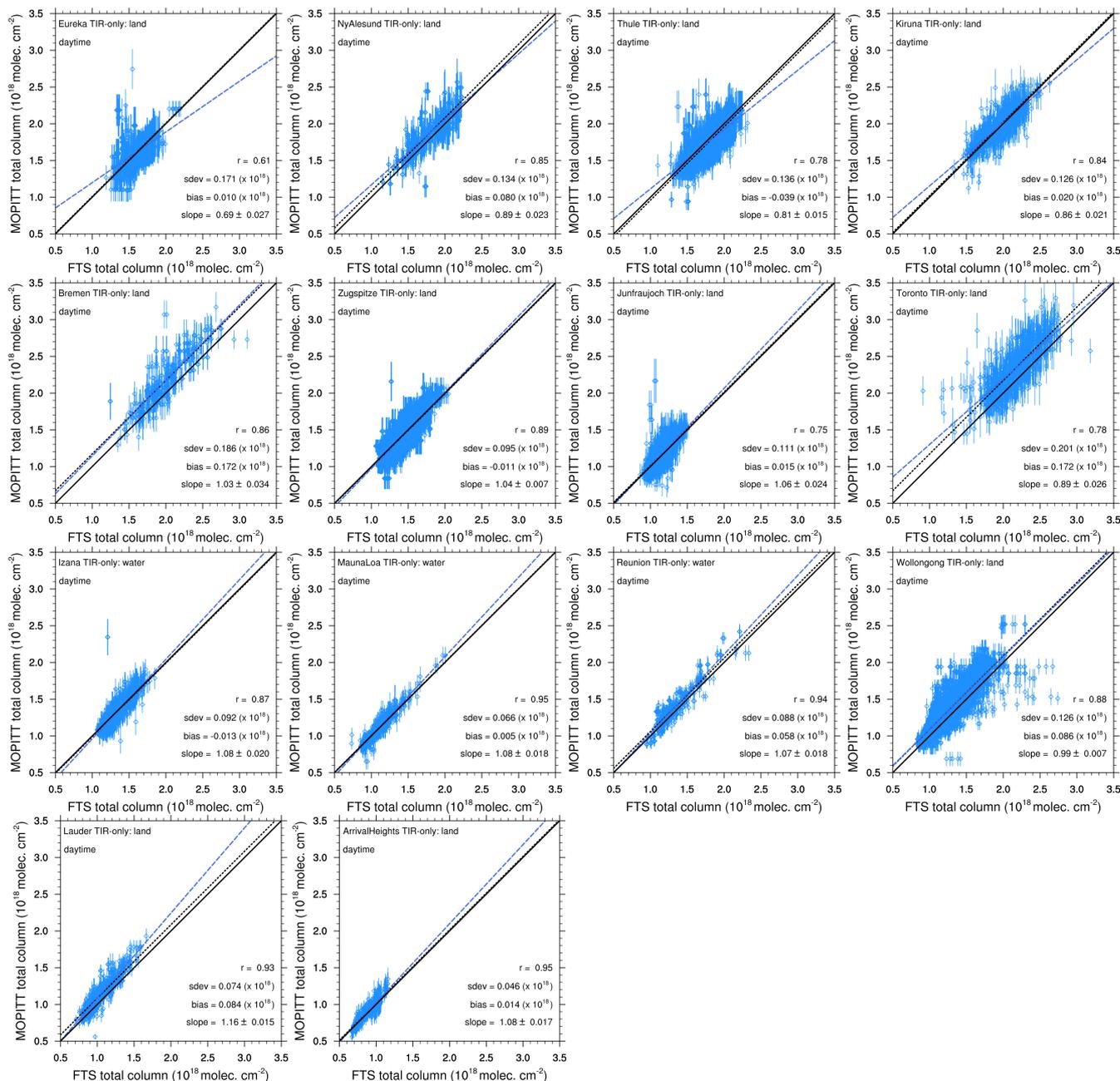


Figure 14. Correlation plot for V6 TIR-only product. Blue dashed line indicates correlation slope, black dashed line indicates mean bias, solid black line is 1:1 line for comparison. Error bars are the daily combined MOPITT error within 1° radius.

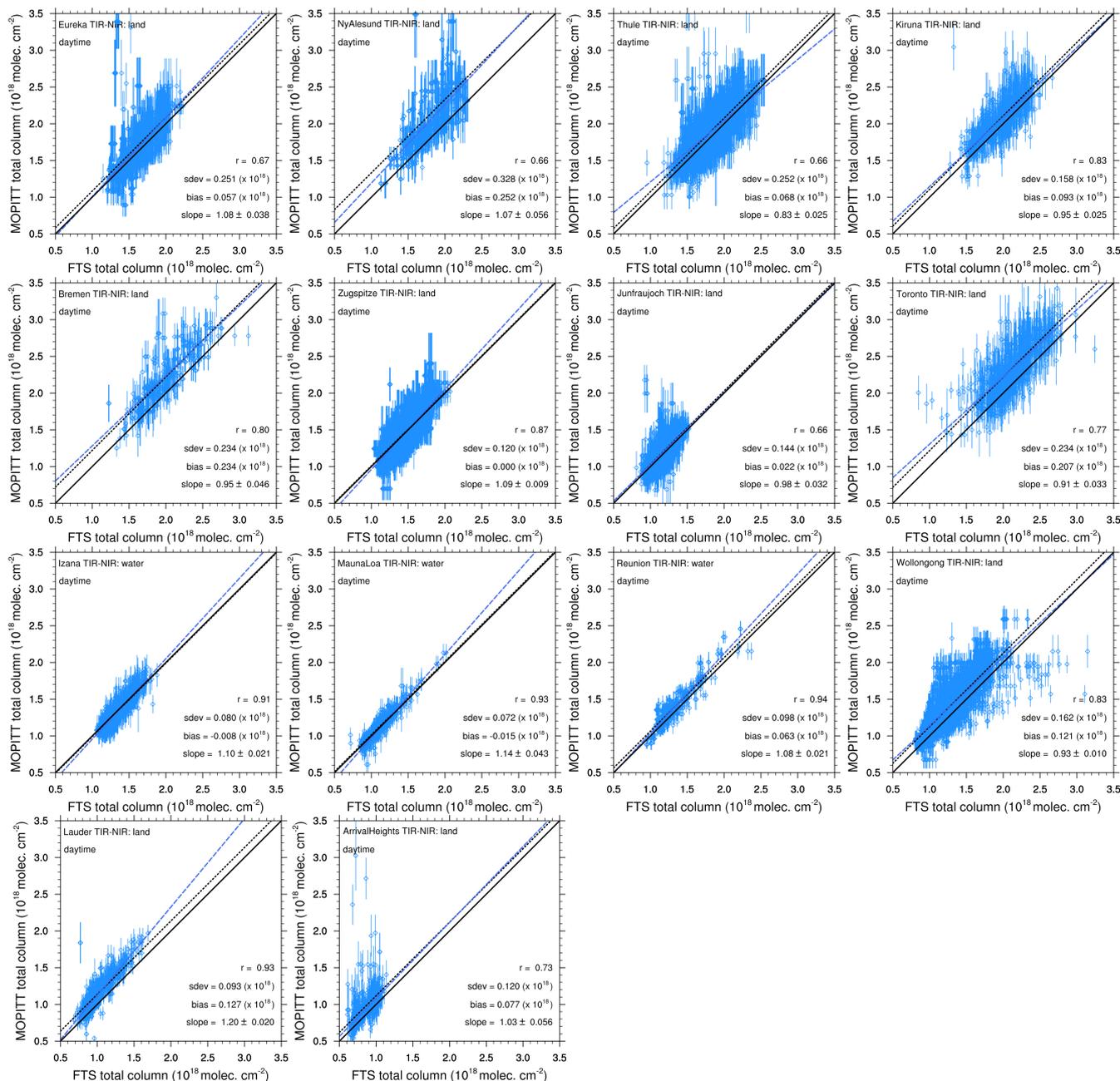


Figure 15. Correlation plot for V6 TIR-NIR product. Blue dashed line indicates correlation slope, black dashed line indicates mean bias, solid black line is 1:1 line for comparison. Error bars are the daily combined MOPITT error within 1° radius.

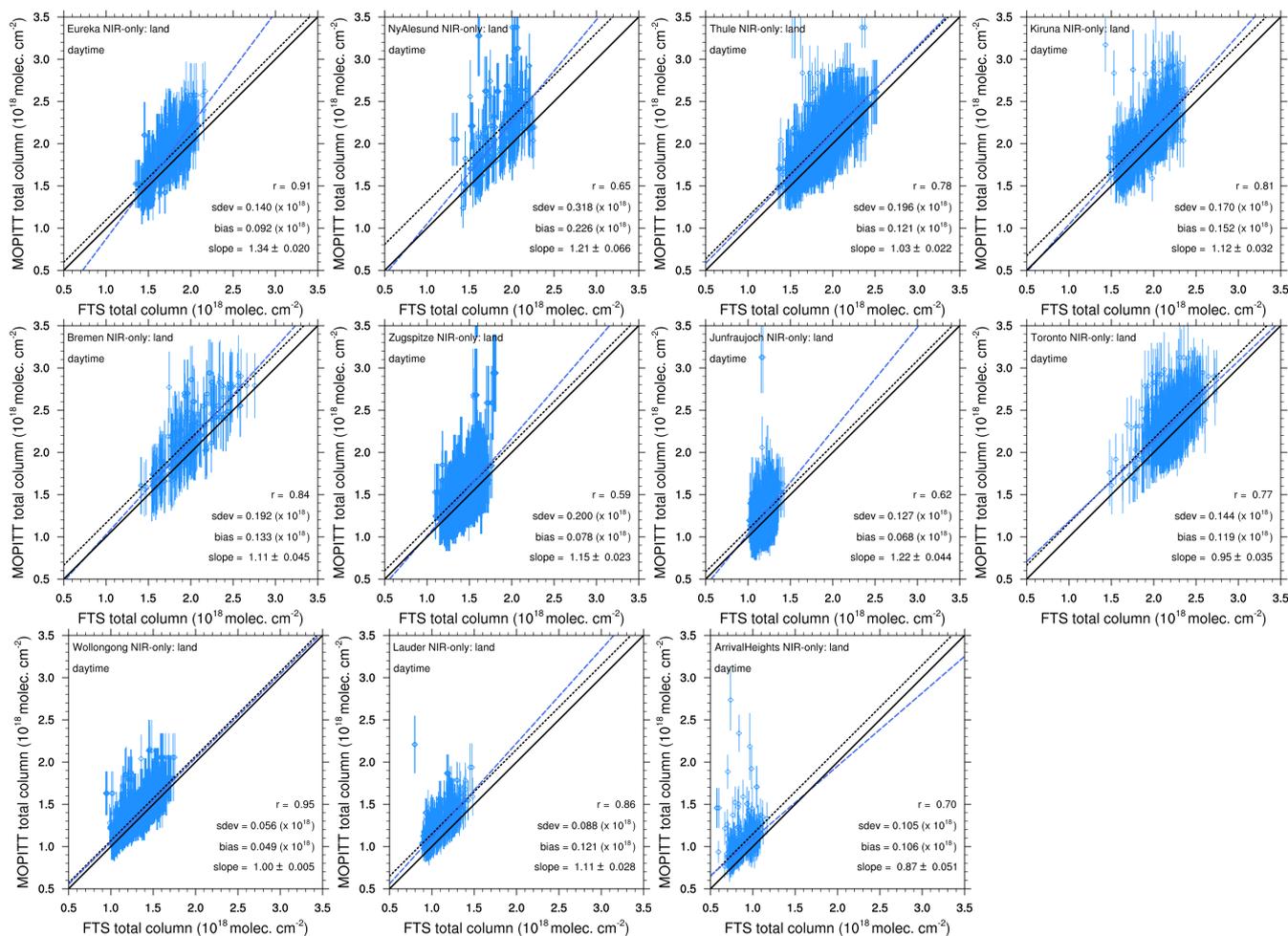


Figure 16. Correlation plot for V6 NIR-only product. Blue dashed line indicates correlation slope, black dashed line indicates mean bias, solid black line is 1:1 line for comparison. Error bars are the daily combined MOPITT error within 1° radius.

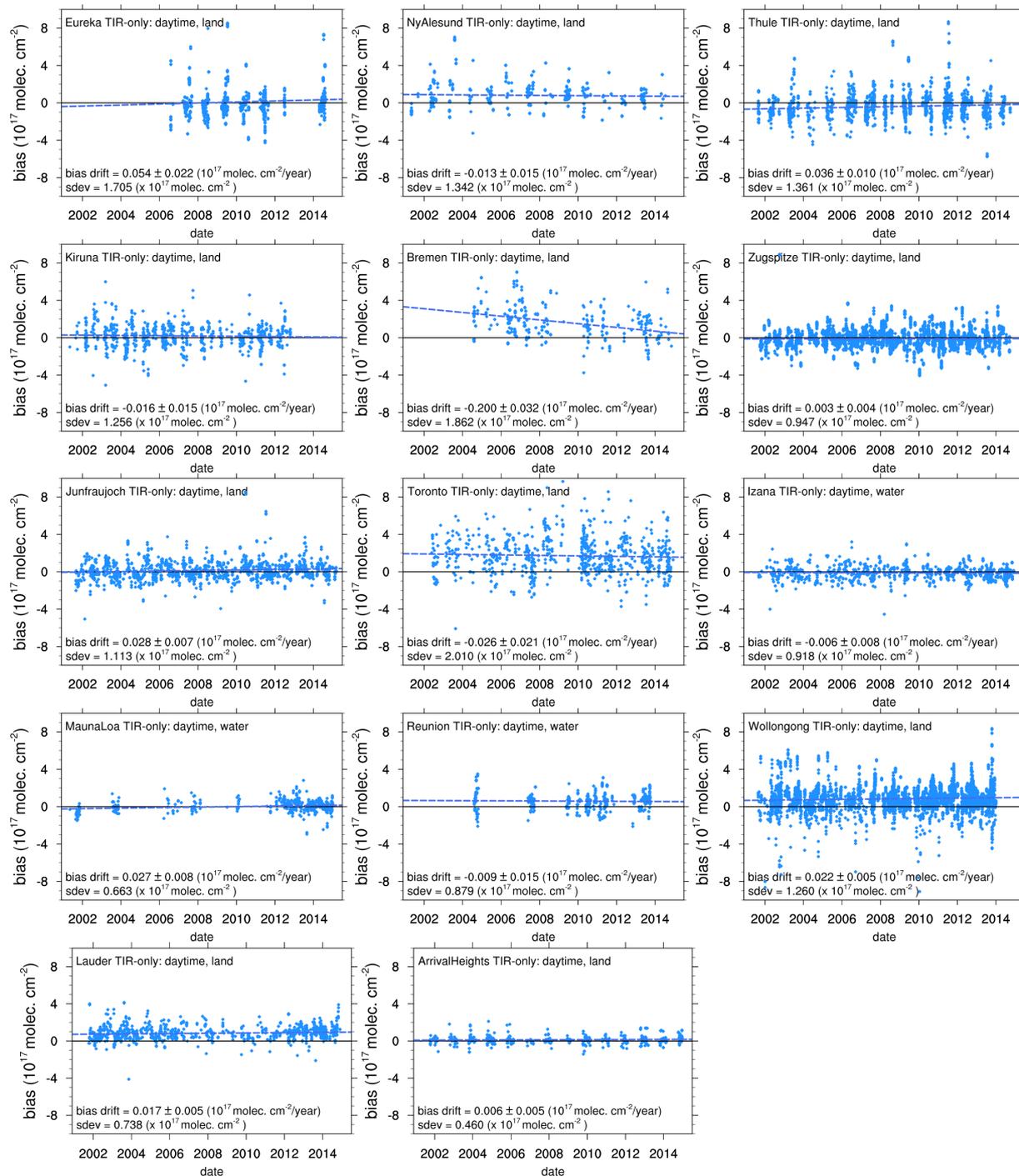


Figure 17. Temporal evolution of bias (MOPITT - FTS) at all stations for V6 TIR-only product. Blue dashed line indicates bias drift.

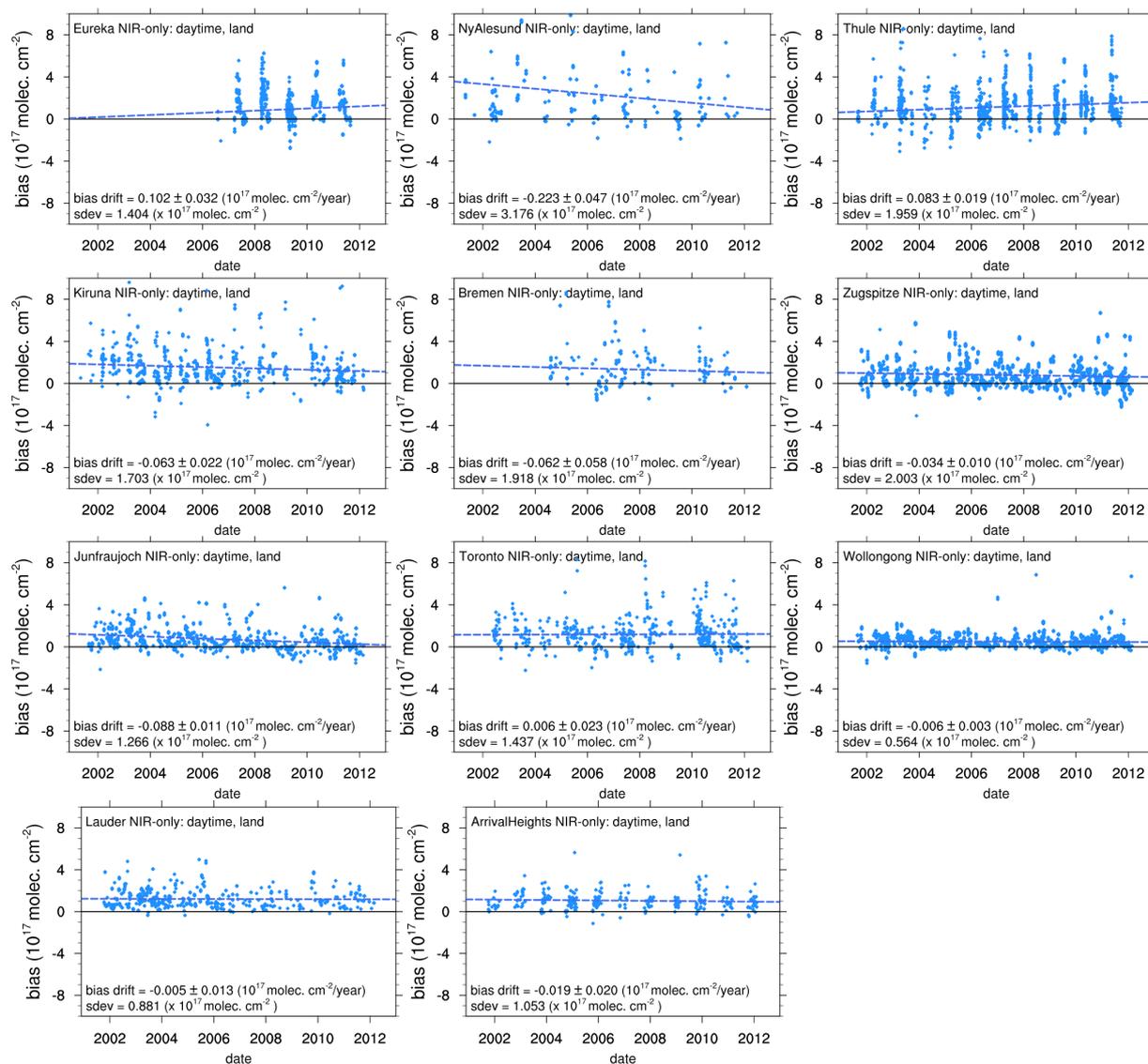


Figure 18. Temporal evolution of bias (MOPITT - FTS) at all stations for V6 NIR-only product. Blue dashed line indicates bias drift.