Remote Sensing of Environment 124 (2012) 210–216

Contents lists available at SciVerse ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Biases in long-term NO_2 averages inferred from satellite observations due to cloud selection criteria

Jeffrey A. Geddes [a], Jennifer G. Murphy [a,*], Jason M. O'Brien [b], Edward A. Celarier [c]

[a] Department of Chemistry, University of Toronto, Toronto, ON, Canada
[b] Air Quality Research Division, Environment Canada, Toronto, ON, Canada
[c] Universities Space Research Association and Code 614, NASA Goddard Space Flight Center, Greenbelt, MD, United States

ARTICLE INFO

Article history:
Received 7 December 2011
Received in revised form 7 May 2012
Accepted 10 May 2012
Available online 12 June 2012

Keywords:
Nitrogen dioxide, NO_2
Ground level pollution
Satellite
Ozone Monitoring Instrument, OMI
Cloud fraction
Remote sensing
Cloud bias
Toronto
Air quality
Vertical column density

ABSTRACT

Retrievals of atmospheric trace gas column densities from space are compromised by the presence of clouds, requiring most studies to exclude observations with significant cloud fractions in the instrument's field of view. Using NO_2 observations at three ground stations representing urban, suburban, and rural environments, and tropospheric vertical column densities measured by the Ozone Monitoring Instrument (OMI) over each site, we show that the observations from space represent monthly averaged ground-level pollutant conditions well ($R = 0.86$) under relatively cloud-free conditions. However, by analyzing the ground-level data and applying the OMI cloud fraction as a filter, we show there is a significant bias in long-term averaged NO_2 as a result of removing the data during cloudy conditions. For the ground-based sites considered in this study, excluding observations on days when OMI-derived cloud fractions were greater than 0.2 causes 12:00–14:00 mean summer mixing ratios to be underestimated by 12% ± 6%, 20% ± 7%, and 40% ± 10% on average (± 1 standard deviation) at the urban, suburban, and rural sites respectively. This bias was investigated in particular at the rural site, a region where pollutant transport is the main source of NO_2, and where long-term observations of NO_y were also available. Evidence of changing photochemical conditions and a correlation between clear skies and the transport of cleaner air masses play key roles in explaining the bias. The magnitude of a bias is expected to vary from site to site depending on meteorology and proximity to NO_x sources, and decreases when longer averaging times of ground station data (e.g. 24-h) are used for the comparison.

1. Introduction

Satellite observations of trace gas species and aerosol in the atmosphere have become a valuable tool in atmospheric chemistry research. For example, they are already being used to provide top-down emission budgets for NO_2 (Lamsal et al., 2008), SO_2 (Lee et al., 2011), and non-methane hydrocarbons (Stavrakou et al., 2009), in addition to evaluating global chemical models (Jaeglé et al., 2011; Myriokefalitakis et al., 2008), and estimating long-term pollutant exposure for epidemiological studies (van Donkelaar et al., 2010). Retrievals of tropospheric pollutant column densities have the potential to expand the spatial coverage available from current ground-based measurements of species whose vertical profile is dominated by concentrations within the atmospheric boundary layer. The history and application of satellite observations for observing tropospheric and ground-level air quality have been recently reviewed in Fishman et al. (2008), and Martin (2008).

One of the critical weaknesses of satellite observations is their inability to accurately retrieve trace gas column densities under cloudy conditions. The presence of clouds can increase instrument sensitivity to trace gasses above the clouds due to light scattering, and/or decrease its sensitivity to trace gasses below the clouds due to shielding (Stammes et al., 2008). As a result, most analyses use satellite retrievals obtained only under conditions when the cloud fraction (CF) determined by the instrument is lower than 0.2–0.5, significantly reducing the temporal coverage of the observations. For some pollutants that are involved in photochemical reactions (e.g. NO_2), and heterogeneous cloud processes (e.g. SO_2), and for locations where pollutant transport is correlated with meteorological conditions, this selection criterion could introduce biases in satellite-derived climatologies.

To the best of our knowledge, this potential bias has not yet been investigated in detail, and is rarely acknowledged in studies that apply such cloud screening. Boersma et al. (2009) observed that monthly mean ground-level NO_2 at some sites in Israel calculated by including only measurements coinciding with $CF < 0.5$ is lower than the true monthly mean, but the reasons for and implications of this bias were not explored. Noguchi et al. (2009) also report a bias in summer NO_2 medians in the Tokyo region for $CF < 0.2$, but argue that the influence

* Corresponding author. Tel.: +1 416 946 0260.
E-mail address: jmurphy@chem.utoronto.ca (J.G. Murphy).

0034-4257/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.rse.2012.05.008

is not significant. Likewise, Gupta and Christopher (2008) observed a minimal bias in ground-based $PM_{2.5}$ in Southeast U.S. when comparing true monthly averages to averages calculated only with cloud-free days as determined by MODIS. However, the magnitude and influence of possible biases could depend on location and the pollutant of interest.

Nitrogen dioxide (NO_2) is a trace gas of particular importance in ozone chemistry, particulate matter formation, and acidification, and whose vertical column densities have been retrieved by remote sensing instruments on board several satellites, beginning with GOME-1 (1995–2003), followed by SCIAMACHY (2002–) and OMI (2004–). OMI, the Ozone Monitoring Instrument on board NASA's Aura satellite, dramatically improves the resolution (13 km × 24 km at nadir, larger at non-nadir viewing) over previous instruments and provides near daily global coverage (Levelt et al., 2006), making it possible to acquire observations of urban-scale pollution in the Earth's atmosphere (Beirle et al., 2011).

Tropospheric vertical column densities (VCDs) of NO_2 retrieved from OMI have previously been validated by aircraft profile measurements, and column densities retrieved from ground-based DOAS instruments sensitive to the troposphere (Bucsela et al., 2008; Kramer et al., 2008; Wenig et al., 2008). One particular challenge to this validation approach is the sensitivity of non-satellite based methods to the heterogeneity of air pollution within the satellite footprint. The location of point and mobile sources, and the influence of wind conditions and atmospheric chemistry can strongly impact the spatial distribution of pollutant concentrations. OMI-retrieved tropospheric column densities have also been related to observations from ground-based in situ monitors (Boersma et al., 2009; Kramer et al., 2008; Lamsal et al., 2008; Zhou et al., 2009). These comparisons are additionally problematic due to the location of ground-based instruments near the surface in the lower boundary layer, where pollutant concentrations are often enhanced by near-field ground-level emissions. Furthermore, all the comparisons cited above focused on observations when the cloud fraction was less than 0.2–0.5 due to the retrieval uncertainty during cloudy conditions.

The purpose of this study is to investigate the potential bias in satellite-retrieved pollutant climatologies that could be introduced by focusing on relatively cloud-free conditions. Ground-level observations of NO_2 at three stations across an urban to rural gradient are first compared with tropospheric VCDs under cloud-free conditions (CF<0.2) at different spatial and temporal scales. Possible biases in long-term averages arising from the exclusion of cloudy days are investigated by applying the OMI cloud fraction to ground-based observations. Photochemical and meteorological conditions that are associated with cloudiness are explored in order to explain the biases observed.

2. Methods

2.1. Tropospheric NO_2 VCDs and cloud fraction

The Ozone Monitoring Instrument on board the NASA EOS-Aura satellite is a UV–VIS spectrometer capable of measuring backscattered radiation over 270–500 nm with approximately 0.5 nm resolution. Pixel size (along track by across track) is 13 km by 24 km at nadir viewing. The Aura satellite is in sun-synchronous orbit passing the equator at 13:45 local time, and provides daily global coverage. An overview of the science objectives for OMI can be found in Levelt et al. (2006). The OMI NO_2 Level 2 data product (version 2.0) used here is made available by the Aura Validation Data Center (http://avdc.gsfc.nasa.gov/), although other retrieval schemes are available (e.g. from KNMI, see http://www.knmi.nl/omi).

The algorithm for retrieving vertical NO_2 column densities used here is described in Bucsela et al. (2006). Briefly, a Differential Optical Absorption Spectroscopy (DOAS) fitting algorithm is applied to the measured backscattered radiation spectrum to produce a slant NO_2 column density. Air-mass factors, which transform the slant column density to a vertical column density, are calculated based on viewing geometry, surface albedo, cloud and aerosol conditions, and shape of the NO_2 vertical profile. These are determined separately for unpolluted regions where the contribution is mainly stratospheric, and polluted regions. In the Level 2 version 2.0 product used here, a priori vertical column densities are computed monthly across a geographical grid using the NASA Global Modeling Initiative (GMI) model. Tropospheric column densities are estimated for each pixel using the total vertical column density and the unpolluted column. The surface reflectivity climatology used is the OMI/AURA Surface Reflectance Climatology Level 3 Global 0.5° lat/lon Grid product available at http://www.knmi.nl/omi/research/product/.

For this analysis, we first removed overpass matchups that are affected by row anomalies (documented at http://disc.sci.gsfc.nasa.gov/Aura/data-holdings/OMI). The data for each day were then sorted by distance of cross-track position to station location, and the overpass with the closest position was selected for matchup with the ground-based data. The resulting daily overpasses are all within 0.2° latitude and 0.6° longitude of the station location, and are within 0.1° latitude and 0.3° longitude (or 27 km) 95% of the time.

We use station overpass data from North Toronto (43.77° latitude, −79.41° longitude), Newmarket (44.04°, −79.48°), and Egbert (44.23°, −79.78°). These three sites represent an urban site, a suburban site, and a rural site respectively, located in an approximately northerly transect from the city of Toronto, Canada's most populated city. The sites are sufficiently separated that they are always represented by different OMI pixels. These sites were also chosen since they are the locations of routine air quality measurements operated by the provincial and federal governments whose data could be compared with the satellite NO_2 retrievals.

We also use the OMI cloud fraction, included with the OMI NO_2 Level 2 station overpass product. The determination of cloud fractions by OMI is discussed in Stammes et al. (2008). Briefly, the cloud fraction determined by OMI is the fraction of the pixel which is required to be covered by cloud in order to match the reflectance measured by the satellite (where clouds are assumed to be Lambertian reflectors with a fixed albedo of 0.8). The algorithm used for the product provided with the station overpass data is based on the O_2–O_2 absorption, described in Acarreta et al. (2004).

2.2. Ground station observations

Hourly NO and NO_2 data at the North Toronto and Newmarket stations are publicly available from 2000 onward and updated in real time at http://www.airqualityontario.com. Here we use May–September data from 2005 to 2010, which has been subject to quality control by the Ontario Ministry of the Environment. Since the satellite passes over these stations around 13:30 local time, averages of the ground station data from 12 pm to 2 pm were taken in order to represent the average conditions captured by the satellite at overpass time. NO and NO_2 measurements are made by chemiluminescent analyzers that satisfy the requirements of the US EPA as equivalent or reference methods, but are subject to positive artifacts due to the use of molybdenum converters for NO_2 (Dunlea et al., 2007; Steinbacher et al., 2007; Winer et al., 1974). NO, NO_2, and NO_y ($= NO + NO_2 + HNO_3 +$ peroxy acyl nitrates + alkyl nitrates + particulate nitrates + others) data from Egbert were collected using two chemiluminescent instruments using an NO_2-specific photolytic converter for the determination of NO_2, and molybdenum converter positioned at the front of the inlet for the determination of NO_y. More selective conversion of NO_2, achieved using photolysis, is more important at the rural station, where the NO_y budget consists of relatively more oxidized forms of nitrogen. Measurements from Egbert were available from 2005 to 2009. Hourly meteorological data were obtained from the National Climate Data and Information Archive operated by Environment Canada.

3. Results and discussion

3.1. Correlation of satellite-retrieved tropospheric NO_2 VCD with ground-level observations

We first discuss the correlation between the tropospheric NO_2 column densities over each station with their respective ground-level measurements. As mentioned in the Introduction, OMI-derived tropospheric NO_2 column densities (and subsequently derived ground-level NO_2 concentrations) have previously been compared to ground-level in situ measurements (Boersma et al., 2009; Kramer et al., 2008; Lamsal et al., 2008). The assumption in these studies is that pollutant concentrations at the surface, especially in polluted areas, represent the majority of the pollutant density through the whole vertical column. Furthermore, since the OMI pixel size is the smallest of the available satellite-based methods for NO_2 detection, it should best capture the spatial gradients across an urban–suburban–rural scale.

In this study, we focus on summer time conditions (May–September) for two reasons. First, this is the photochemically relevant season for ozone production. Second, it was observed that in the study region, wintertime retrievals from OMI are significantly compromised due to high calculated cloud fractions, and many months only have a single observation with a cloud fraction less than 0.2. During the summer, the number of days with a cloud fraction less than 0.2 compared to the days with a cloud fraction greater than 0.2 is about equal.

Fig. 1 (top panel), shows the change in the Pearson correlation coefficient between the OMI-retrieved tropospheric VCDs and the ground station data from North Toronto, Newmarket, and Egbert as a function of the cloud fraction (CF) bin within which the observations fall. The correlation of daily data at each individual site ranges in significance ($R = 0.53$ to $R = 0.87$ for CF bins < 0.2). Fig. 1 (bottom panel) plots the fraction of ground-level data that falls within each bin. The shapes of the curves from all three locations are similar; as the cloud fraction increases, the correlation weakens. This corresponds to increased retrieval uncertainty due to the influence of clouds. For cloud fractions less than around 0.6, the retrieved tropospheric column is primarily dependent on the measurement, the surface reflectivity and the shape of the a priori profile, and less dependent on the integrated a priori profile. As the cloud fraction goes to 1, tropospheric column estimates become more dependent on the a priori profile below the top of the clouds and less dependent on the measurement. In most analyses, a cloud fraction cutoff between 0.2 and 0.5 is usually selected; in the current analysis, this would result in discarding approximately 30–50% of the data. The choice of cloud fraction cutoff used requires a balance between minimizing retrieval uncertainty due to clouds, and statistical confidence in the analysis performed (which will be related to sample size). We are not aware of any previous studies that discuss the effects of this somewhat subjective criterion. The loss of this amount of data is important to consider, especially if the discarded data is not representative of the overall population.

Fig. 2 shows the correlation between monthly averaged OMI-retrieved tropospheric VCDs and ground station measurements during the summer. Only days when the cloud fraction was less than 0.2 were included for both the OMI and ground-based monthly averages. Including all three stations in the correlation results in $R = 0.86$. While the daily correlation at any one site is usually less robust (Fig. 1), averaging over time improves the correlation, as would be expected with decreasing the sources of variability not accounted for by both measurements. This strong correlation illustrates the capability of the OMI-retrieved tropospheric VCD NO_2 to represent ground-level pollutant conditions over longer temporal scales (months) and across a spatial concentration gradient under cloud-free conditions ($CF < 0.2$). Note that least orthogonal distance regression analysis (allowing for error in both the dependent and independent variables, where the error is estimated as the standard deviation of the monthly averages in both x and y) shows that slopes calculated at individual sites are not statistically different from the slope from the pooled data across all three sites.

3.2. Selection bias due to cloud screening

Selecting cloud-free days for analysis could be expected to impart biases in the ground station data for multiple reasons; they could result from photochemical/heterogeneous reactions that convert the species of interest, or from meteorological controls on ground-level mixing ratios (e.g. boundary layer height, atmospheric transport). Fig. 3 uses box plots to show the distribution of the summertime ground station data at North Toronto, Newmarket, and Egbert on cloud-free days ($CF < 0.2$) compared to the distribution on cloudy days ($CF > 0.2$). The number of days in each category is roughly equal at all of the sites. In all cases, the median from the cloud-free data set is lower than the median from the cloudy data set. The nonparametric Wilcoxon–Mann–Whitney two-sample rank test was performed to evaluate the two-tail null hypothesis that the medians of each data set are not

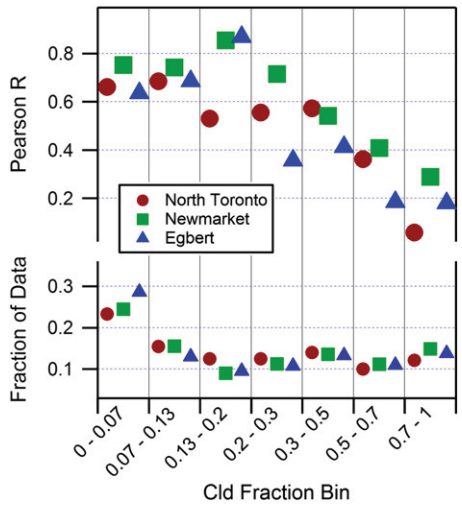

Fig. 1. Pearson coefficients for correlation of daily NO_2 tropospheric columns with 12:00–14:00 ground station measurements, as a function of cloud fraction bins (top panel). Fraction of data in each bin for the correlation analysis (illustrating that each bin has approximately the same number of observations) (bottom panel). Circles (red) = North Toronto, squares (green) = Newmarket, triangles (blue) = Egbert. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

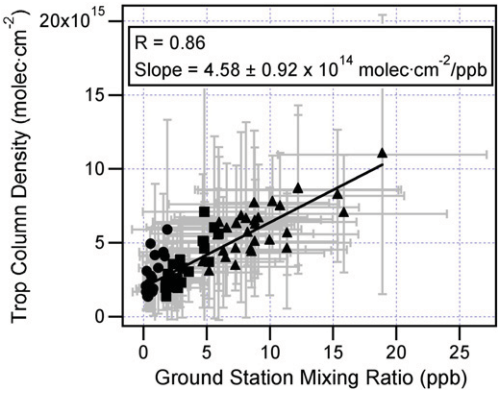

Fig. 2. Monthly averaged OMI tropospheric column densities vs. monthly averaged 12–2 pm ground station observations across North Toronto (triangles), Newmarket (squares), and Egbert (circles). Error bars represent one standard deviation of the mean in x and y.

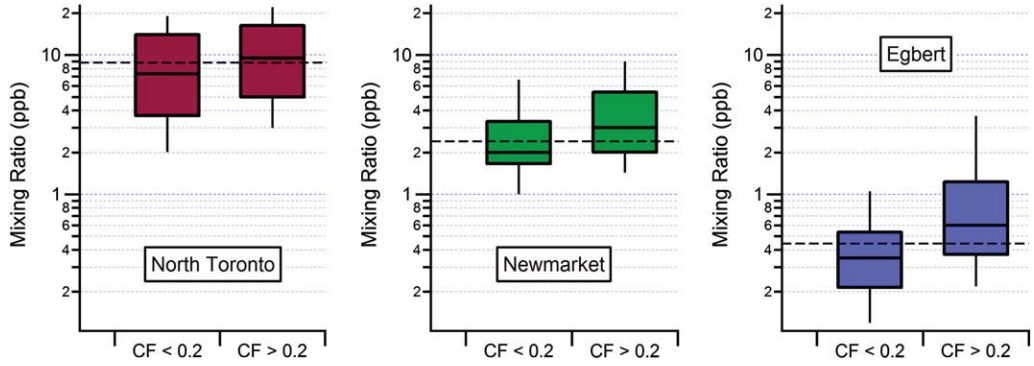

Fig. 3. Box plots of the distribution of daily ground station data for cloud-free days and cloudy days as calculated by the OMI cloud fraction (whiskers are drawn to the 10th and 90th percentile). The dotted line represents the median of the entire dataset.

different. In all three cases, the null hypothesis is rejected at $p = 0.01$. The same conclusion is obtained by performing *t*-tests on the log-transformed data (resulting in approximately normal distributions).

The results of this statistical analysis clearly show the difference between the distributions of the data that are excluded based on cloud fraction (CF > 0.2) and of the data that are retained (CF < 0.2). At each site, focusing on cloud-free conditions results in average pollutant mixing ratios that are biased low. Of the locations considered in this analysis, the relative effect of the bias increases with distance from the urban source region (with the calculated Wilcoxon–Mann–Whitney *p*-value becoming smaller). The relative bias is most dramatic for Egbert, where the median mixing ratio on cloudy days is almost twice the median for cloud-free days (0.6 ppb and 0.35 ppb respectively), and the 90th percentile value is four times higher on cloudy days. In contrast, the absolute bias is largest at the urban site, and smallest at the rural site.

The results presented here have significant implications for the use of satellite-derived measurements of ground-level pollution over the long term. For example, annual average pollution levels in an area could be significantly underestimated if a cloud selection bias exists. The 12:00–14:00 mean summer NO_2 was calculated for each year using only cloud-free days (CF < 0.2) and compared to the summer mean that includes all days (May–Sept.). At North Toronto, the summer means using only cloud-free days (7.2 ppb to 11.8 ppb) are biased low by 12% ± 6% (±1 standard deviation) on average compared to the mean from all days (8.8 ppb to 12.1 ppb). At Newmarket and Egbert, this underestimation is even more dramatic (in a relative sense), with the mean summer mixing ratio underestimated by 20% ± 7% (cloud free means of 2.4 to 3.9 vs. means from all data of 3.3 ppb to 4.4 ppb) and 40% ± 10% (cloud free means of 0.3 ppb to 0.8 ppb vs. means from all data of 0.8 to 1.2 ppb) respectively. Here we define the percent bias as:

$$\%\ Bias = \left[\frac{mean\ \ from\ \ all\ \ days - mean\, from\ \ \ cloud\ \ free\ \ days}{mean\ \ from\ \ all\ \ days}\right] \times 100\,\% \quad (1)$$

From the perspective of air quality, these relative changes in NO_x would have significant impacts on predicted rates of photochemical ozone production. However, our analysis also shows that for the locations considered in this study, there is no significant bias at any of the sites when 24-h averaging times are used at the ground station (instead of 12:00–14:00). Hence for epidemiological studies, which often focus on 24-h averages instead of a specific time of day, the impact of the bias may be less significant.

To investigate possible reasons for the selection bias, we look at the distribution of NO_x and NO_y at the Egbert location, where the bias is most pronounced (NO_y data is not collected routinely at the other stations included in this analysis). The proportion of NO_x found as NO_2 depends on the levels of ozone and the photolysis rate of NO_2 (J_{NO2}), both of which are likely to be higher at the surface under cloud-free conditions. For example, Flynn et al. (2010) found that during the 2006 Texas Air Quality Study, clouds and aerosols decreased the NO_2 photolysis rate by 17% in the Houston area compared to cloud-free days, resulting in a 35% decrease in the ozone production rate. While the global impact of clouds on photolysis rates depends strongly on their vertical distribution (lower altitude clouds have been shown to enhance photolysis rates on average throughout the troposphere due to back-scattering effects (Mao et al., 2003; Tie et al., 2003)), radiation in the boundary layer below the clouds is always reduced.

Furthermore, the proportion of NO_y that is NO_x depends on the presence of oxidants. NO_x can form more oxidized species of nitrogen by reactions with OH and organic peroxy radicals. These conversions also rely on photochemistry, which would be enhanced at the surface under sunny skies, causing the lifetime of NO_2 to be shorter under cloud-free conditions compared to cloudy conditions. Evidence of this may be seen in Fig. 4, where box plots show the distribution of the ratios of NO_2/NO_x, NO_x/NO_y and total NO_y mixing ratios at Egbert on days with CF < 0.2 compared to days with CF > 0.2. The proportion of NO_x that is made up of NO_2 is lower under cloud-free conditions compared to cloudy conditions, likely due to higher J_{NO2}. Moreover, the fraction of NO_y that is composed of NO_x is lower under cloud-free conditions compared to cloudy conditions, suggesting a faster rate of oxidation.

The third panel in Fig. 4 compares the NO_y mixing ratios at Egbert on cloud-free and cloudy days. Since photochemical conversions of NO_2 conserve total NO_y, changes in the average amount of NO_y must be due to other reasons, such as changing emissions, deposition or pollutant transport conditions. Egbert is a receptor site, where there are relatively few local sources of NO_x, and most of its pollution is brought by atmospheric transport from major source regions to the southwest. Larger amounts of NO_y during cloudy conditions indicate there might be a difference in pollutant transport that is generally associated with cloudiness.

Fig. 5 shows rose plots of the net wind vector at the Egbert meteorological station from 11:00 to 15:00 EST on days when the cloud fraction was less than 0.2 compared to days when the cloud fraction was greater than 0.2. Meteorological data from May to September 2005–2009 were used. The net wind vectors are calculated for the 4-h period by summing the magnitude of each x- and y-component of the wind at each hour, then calculating the resultant vector by trigonometry. This figure shows that the wind is more often coming from the northwest on the cloud-free days, while the cloudy days are more significantly influenced by southerly winds.

Air from the north at this location is associated with cleaner background conditions, while air from the south is associated with pollutant flow from Toronto and other populated regions in southern Ontario and the US. The closer proximity of upwind NO_x sources on cloudy days

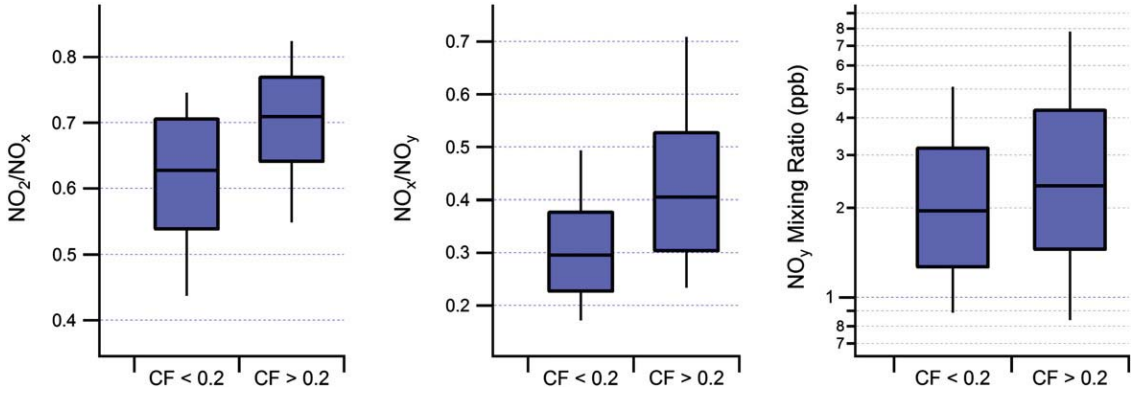

Fig. 4. Box plots of the distribution of the ratios of NO_2/NO_x (left panel), the ratios of NO_x/NO_y (middle panel), and overall NO_y mixing ratios (right panel) on cloud-free and cloudy days as calculated based on the OMI cloud fraction (whiskers as in Fig. 3).

could also explain the higher NO_x/NO_y ratio, indicating a shorter time for photochemistry since emission. The correlation of northerly transport with cloud-free days could be associated in general with movements of the continental polar air mass, carrying cool dry air and stable atmospheric conditions (which would lead to sunny skies), whereas air from the south may be associated in general with maritime tropical air masses from low latitudes which are typically unstable (leading to cloudy skies). The correlation of pollutant transport with cloudy conditions has been observed elsewhere. Crawford et al. (2003), report enhanced CO mixing ratios in the free troposphere over the North Pacific in and around clouds. This was attributed to frontal disturbances (associated with cloudiness) which cause uplift into the free troposphere from Asian sources at the surface. This observation, in combination with the present results, emphasizes the importance of sampling atmospheric composition during both cloudy and cloud-free conditions.

3.3. Relating a selection bias at the surface to a selection bias in tropospheric VCD

In the previous section we have identified that a selection bias exists at the ground-level when monitoring station data are sampled based on the OMI cloud fraction retrieval, which could impact ground-level climatologies inferred from satellite observations. We now discuss how this bias at the surface might relate back to a bias in the tropospheric column densities. If NO_2 mixing ratios at the surface are higher under cloudy skies due to a shallower boundary layer, or if reduced photolysis below the cloud is compensated by higher photolysis above the clouds then a selection bias may not manifest in the vertical column. However, scaling the tropospheric column to an in situ mixing ratio in these cases would need careful consideration.

If, on the other hand, days with higher ground-level concentrations also have higher vertical column densities, then it would be valuable to estimate whether this bias would be significant in comparison with the column retrieval error. At the most urban location (North Toronto), where the relative bias is the smallest, the difference between cloudy days and clear days (~2 ppb) would translate to a boundary layer column density of about 5×10^{15} molec cm^{-2} (assuming mean pressure of 0.93 atm, a boundary layer height of 1 km, and mean temperature of 290 K). This value is on the order of the standard deviation of monthly tropospheric column averages (see Fig. 2), and is larger than the monthly average error in NO_2 tropospheric column density (median $=3\times10^{15}$ molec cm^{-2}). At the rural site (Egbert), where the relative bias is the largest, the difference between cloudy days and clear days appears small (0.3 ppb) but still corresponds to a boundary layer column density of around 7×10^{14} molec cm^{-2}. This is smaller than both the standard deviation of monthly tropospheric column averages and the monthly average error in NO_2 tropospheric column densities (median $=2.5\times10^{15}$ molec cm^{-2}), in which case a bias in the tropospheric column density may not be discernible. However, it still represents a significant fraction of the overall monthly average tropospheric column density at the site.

Future work involving collocated observations of cloudiness (or NO_2 photolysis), boundary layer height, and nitrogen oxide mixing ratios may provide further insight into the relationship between a ground-level bias and a bias in the tropospheric column densities.

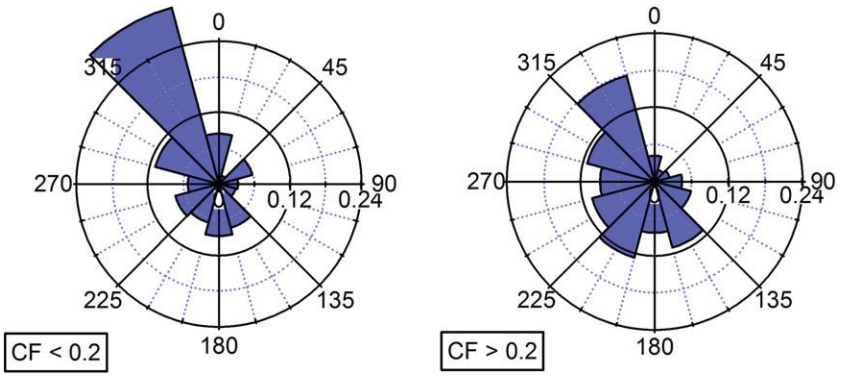

Fig. 5. Wind rose plots of net wind vectors at Egbert during the 11:00–15:00 period for cloud-free days (left panel) and cloudy days (right panel) as calculated by the OMI cloud fraction.

Regional chemistry model output could also be used in the future to investigate the relationship of ground-level selection biases to vertical column biases as a function of meteorology and transport.

4. Summary and conclusion

Using tropospheric NO_2 vertical column densities obtained by OMI and ground station NO_2 observations at three sites across the greater Toronto area, it has been shown that the satellite-retrieved data represents the variability in ground-level pollutant conditions relatively well under cloud-free conditions (CF<0.2). The correlation at any individual site ranges in significance, but averaging across urban spatial scales and monthly time scales results in stronger agreement. This type of averaging smoothes out much of the uncorrelated variability that exists in a daily comparison at an individual site. For example, the spatial heterogeneity in pollutant concentrations at a point within the satellite footprint may be smoothed out by using average monthly concentrations, due to the inclusion of variable wind and atmospheric transport conditions.

Removing days with high cloud fractions is necessary when relating satellite-retrieved trace gas columns to surface concentrations due to the interferences caused by clouds. It was shown here that the correlation of daily tropospheric VCDs and ground station NO_2 varied as a function of cloud fraction used to screen the data, likely due to the amount of data retained, and the accuracy of the retrieval. At CF<0.2, the amount of data discarded is around 50% at all three stations considered in this study. The removal of this data could significantly influence pollutant climatologies calculated from satellite observations. The fraction of data removed and its implications will vary across station location and meteorological conditions.

In the study region considered here, it was shown that the distributions of NO_2 on cloud-free days and cloudy days, as diagnosed by an OMI cloud fraction of 0.2, were significantly different at all stations. Summer midday NO_2 on cloud-free days was, on average, lower than on cloudy days, and the relative magnitude of this difference increases with distance from the urban center. The reasons for this change in distribution were investigated by using NO_x and NO_y observations at the rural site, which showed that changes in the rates of photolysis and oxidation reactions likely play a role. Meteorological conditions at the rural site are such that cloud-free days often correlate with atmospheric transport from the north, characterized by clean, pollutant-free air, whereas cloudier days are influenced more by pollutant flow from the south. This meteorological influence on mixing ratio distributions would be most important at receptor sites with few local NO_x emissions, and less important at urban centers dominated by local emissions of similar magnitude in all directions. The magnitude of this bias during winter, when NO_2 levels are highest, may be even more pronounced, but this was not explored due to the high number of days that are influenced by cloud fractions greater than 0.2 during the winter at these locations.

This investigation highlights the caution that must be applied when using satellite observations of ground-level pollution where a significant amount of data must be removed due to clouds. This effect should be kept in mind when using satellite observations in epidemiological studies, which rely on spatial gradients in pollution observed from space to assign differential exposure levels to populations at urban and regional scales. However, it will also strongly depend on metrics used to assign exposure, since it was observed that no significant bias exists in ground-level 24-h daily averages screened for cloud fraction retrieved by OMI at the overpass time. The influence of this screening will be different at each location and for each remotely-sensed pollutant and should be explored in future analyses. Without testing for the bias by screening ground-level data with the satellite cloud filter, as we have done, there is no obvious way to predict the magnitude and direction of a bias for an arbitrary site. Thus appropriate adjustment factors for climatologies cannot easily be derived, underscoring the importance of future satellite measurements that are less susceptible to interference from clouds.

Acknowledgments

Access to air quality monitoring data is provided by the Ontario Ministry of the Environment. We also gratefully acknowledge support from the Natural Sciences and Engineering Research Council of Canada.

References

Acarreta, J. R., De Haan, J. F., & Stammes, P. (2004). Cloud pressure retrieval using the O_2–O_2 absorption band at 477 nm. *Journal of Geophysical Research*, *109*, D05204.

Beirle, S., Boersma, K. F., Platt, U., Lawrence, M. G., & Wagner, T. (2011). Megacity emissions and lifetimes of nitrogen oxides probed from space. *Science*, *333*, 1737–1739.

Boersma, K. F., Jacob, D. J., Trainic, M., Rudich, Y., DeSmedt, I., Dirksen, R., et al. (2009). Validation of urban NO_2 concentrations and their diurnal and seasonal variations observed from the SCIAMACHY and OMI sensors using in situ surface measurements in Israeli cities. *Atmospheric Chemistry and Physics*, *9*, 3867–3879.

Bucsela, E. J., Celarier, E. A., Wenig, M. O., Gleason, J. F., Veefkind, J. P., Boersma, K. F., et al. (2006). Algorithm for NO_2 vertical column retrieval from the Ozone Monitoring Instrument. *IEEE Transactions on Geoscience and Remote Sensing*, *44*, 1245–1257.

Bucsela, E. J., Perring, A. E., Cohen, R. C., Boersma, K. F., Celarier, E. A., Gleason, J. F., et al. (2008). Comparison of tropospheric NO(2) from in situ aircraft measurements with near-real-time and standard product data from OMI. *Journal of Geophysical Research*, *113*, D16S31.

Crawford, J., Olson, J., Davis, D., Chen, G., Barrick, J., Shetter, R., et al. (2003). Clouds and trace gas distributions during TRACE-P. *Journal of Geophysical Research*, *108*(D21), 8818.

Dunlea, E. J., Herndon, S. C., Nelson, D. D., Volkamer, R. M., San Martini, F., Sheehy, P. M., et al. (2007). Evaluation of nitrogen dioxide chemiluminescence monitors in a polluted urban environment. *Atmospheric Chemistry and Physics*, *7*, 2691–2704.

Fishman, J., Bowman, K. W., Burrows, J. P., Richter, A., Chance, K. V., Edwards, D. P., et al. (2008). Remote sensing of tropospheric pollution from space. *Bulletin of the American Meteorological Society*, *89*, 805–821.

Flynn, J., Lefer, B., Rappenglück, B., Leuchner, M., Perna, R., Dibb, J., et al. (2010). Impact of clouds and aerosols on ozone production in southeast Texas. *Atmospheric Environment*, *44*, 4126–4133.

Gupta, P., & Christopher, S. A. (2008). An evaluation of Terra-MODIS sampling for monthly and annual particulate matter air quality assessment over the southeastern United States. *Atmospheric Environment*, *42*, 6465–6471.

Jaeglé, L., Quinn, P. K., Bates, T. S., Alexander, B., & Lin, J. -T. (2011). Global distribution of sea salt aerosols: New constraints from in situ and remote sensing observations. *Atmospheric Chemistry and Physics*, *11*, 3137–3157.

Kramer, L. J., Leigh, R. J., Remedios, J. J., & Monks, P. S. (2008). Comparison of OMI and ground-based in situ and MAX-DOAS measurements of tropospheric nitrogen dioxide in an urban area. *Journal of Geophysical Research*, *113*, D16S39.

Lamsal, L. N., Martin, R. V., van Donkelaar, A., Steinbacher, M., Celarier, E. A., Bucsela, E., et al. (2008). Ground-level nitrogen dioxide concentrations inferred from the satellite-borne Ozone Monitoring Instrument. *Journal of Geophysical Research*, *113*, D16308.

Lee, C., Martin, R. V., van Donkelaar, A., Lee, H., Dickerson, R. R., Hains, J. C., et al. (2011). SO_2 emissions and lifetimes: Estimates from inverse modeling using in situ and global, space-based (SCIAMACHY and OMI) observations. *Journal of Geophysical Research*, *116*, D06304.

Levelt, P. F., Hilsenrath, E., Leppelmeier, G. W., van den Oord, G. H. J., Bhartia, P. K., Tamminen, J., et al. (2006). Science objectives of the Ozone Monitoring Instrument. *IEEE Transactions on Geoscience and Remote Sensing*, *44*, 1199–1208.

Mao, H., Wang, W., Liang, X., & Talbot, R. (2003). Global and seasonal variations of O-3 and NO_2 photodissociation rate coefficients. *Journal of Geophysical Research*, *108*(D7), 4216.

Martin, R. V. (2008). Satellite remote sensing of surface air quality. *Atmospheric Environment*, *42*, 7823–7843.

Myriokefalitakis, S., Vrekoussis, M., Tsigaridis, K., Wittrock, F., Richter, A., Brühl, C., et al. (2008). The influence of natural and anthropogenic secondary sources on the glyoxal global distribution. *Atmospheric Chemistry and Physics*, *8*, 4965–4981.

Noguchi, K., Itoh, H., Shibasaki, T., Hayashida, S., Uno, I., Ohara, T., et al. (2009). Comparison of tropospheric NO_2 observations by GOME and the air-quality monitoring network around Tokyo, Japan. *Journal of the Remote Sensing Society of Japan*, *29*, 398–409.

Stammes, P., Sneep, M., de Haan, J. F., Veefkind, J. P., Wang, P., & Levelt, P. F. (2008). Effective cloud fractions from the Ozone Monitoring Instrument: Theoretical framework and validation. *Journal of Geophysical Research*, *113*, D16S38.

Stavrakou, T., Müller, J.-., De Smedt, I., Van Roozendael, M., Van Der Werf, G. R., Giglio, L., et al. (2009). Global emissions of non-methane hydrocarbons deduced from SCIAMACHY formaldehyde columns through 2003–2006. *Atmospheric Chemistry and Physics*, *9*, 3663–3679.

Steinbacher, M., Zellweger, C., Schwarzenbach, B., Bugmann, S., Buchmann, B., Ordóñez, C., et al. (2007). Nitrogen oxide measurements at rural sites in Switzerland: Bias of conventional measurement techniques. *Journal of Geophysical Research*, *112*, D11307.

Tie, X., Madronich, S., Walters, S., Zhang, R., Rasch, P., & Collins, W. (2003). Effect of clouds on photolysis and oxidants in the troposphere RID A-2942-2011. *Journal of Geophysical Research*, *108*(D20), 4642.

van Donkelaar, A., Martin, R. V., Brauer, M., Kahn, R., Levy, R., Verduzco, C., et al. (2010). Global estimates of ambient fine particulate matter concentrations from satellite-based aerosol optical depth: Development and application. *Environmental Health Perspectives*, *118*, 847–855.

Wenig, M. O., Cede, A. M., Bucsela, E. J., Celarier, E. A., Boersma, K. F., Veefkind, J. P., et al. (2008). Validation of OMI tropospheric NO(2) column densities using direct-Sun mode Brewer measurements at NASA Goddard Space Flight Center. *Journal of Geophysical Research*, *113*, D16S45.

Winer, A. M., Peters, J. W., Smith, J. P., & Pitts, J. N., Jr. (1974). Response of commercial chemiluminescent NO NO_2 analyzers to other nitrogen containing compounds. *Environmental Science and Technology*, *8*, 1118–1121.

Zhou, Y., Brunner, D., Boersma, K. F., Dirksen, R., & Wang, P. (2009). An improved tropospheric NO_2 retrieval for OMI observations in the vicinity of mountainous terrain. *Atmospheric Measurement Techniques*, *2*, 401–416.