Evaluation of version 3.0B of the BEHR OMI NO_2 product

Response to Anonymous Referee #1

Joshua L. Laughner, Qindan Zhu, and Ronald C. Cohen

November 7, 2018

We thank the reviewer for their positive response and careful reading. The suggestions offered and questions raised are well taken, and we have done our best to incorporate them into the paper.

Responses to specific comments follow. The reviewer's comments will be shown in red, our response in blue, and changes made to the paper are shown in black block quotes. Unless otherwise indicated, page and line numbers correspond to the original paper. Figures, tables, or equations referenced as "Rn" are numbered within this response; if these are used in the changes to the paper, they will be replaced with the proper number in the final paper. Figures, tables, and equations numbered normally refer to the numbers in the original discussion paper.

This manuscript seems to focus largely on how the updated model chemistry/emissions and the daily vs. monthly averaged a-priori profiles impact the performance of the BEHR NO2 retrieval. But there were other pertinent updates in the retrieval: vary- ing tropopause height, directional surface reflectance, and a new combined surface pressure dataset. Are the differences in retrieval performance against observations as a result of these updates very minor compared to the differences attributed to the updated model chemistry and profile temporal resolution? It seems worthwhile men- tioning/commenting on. I wondered whether there were any specific instances/cases where some of these other updates could be relatively more important.

Our decision to focus most strongly on the modeled NO_2 profiles was due primarily to two reasons:

1. In discussions with other members of the NO_2 remote sensing community, concerns were raised about the accuracy of such high spatial and temporal resolution profiles.

2. In Laughner et al. (2018), we showed that the effect of the a priori profiles on the NO_2 VCDs has one of the two greatest effects on the NO_2 VCDs.

Because accurately simulating day-by-day high resolution NO_2 profiles accurately is challenging, and because to our knowledge, this is the first time an NO_2 product using such profiles has been generated for such a large time period and domain, we felt it was most important to validate that component separately. The other scientifically interesting modifications either are or will likely be standard for state-of-the-art NO_2 retrievals (BRDF,

Vasilkov et al. 2017; surface pressure, Zhou et al. 2009; variable tropopause, Bucsela et al. 2013).

During testing, we had found that using the hypsometric surface pressure correction recommended by Zhou et al. (2009) (instead of the simple scale height relationship used in v3.0A and earlier) had a comparatively large effect on the agreement between BEHR and aircraft + Pandora VCDs during the DISCOVER-CO campaign, which took place in the Rocky Mountains. This is interesting, because the Zhou et al. (2009) method was originally intended to downscale very coarse resolution ($3^\circ \times 2^\circ$) modeled surface pressure to OMI pixel resolution; we did not expect it to provide a large advantage over using an already high resolution surface elevation database with a scale height calculation. We have added a short section describing this to the supplement:

> "While we did not carry out an explicit test of how each change to the BEHR algorithm between v2.1C and v3.0B affected the comparison vs. aircraft and Pandora data, we did investigate the effect of different methods of computing the surface pressure of the OMI pixels. The AMF calculation requires a priori knowledge of the average surface pressure of the each OMI pixel, as the location of the surface affects the shape of the scattering weights (e.g. a low reflectivity surface high up in the atmosphere will cause the scattering weights to decrease more rapidly with decreasing altitude than a surface lower down in the atmosphere).
>
> In BEHR v3.0A and earlier versions, this surface pressure was computed by averaging surface elevation data from the GLOBE database (Hastings and Dunbar, 1999) within the OMI pixel, which is then converted to from elevation to pressure using a 7.4 km scale height. In v3.0B, surface pressure taken from the same WRF-Chem model that supplied the NO_2 profiles is adjusted using the same average GLOBE surface elevation in the method described by Zhou et al. (2009). The Zhou et al. (2009) method was originally intended to downscale very coarse ($\sim 3^\circ \times 2^\circ$) modeled surface pressure to OMI pixels using a high resolution terrain database, therefore the effect of using it with already high resolution modeled surface pressure has not been tested.
>
> Laughner et al. (2018) showed that switching to the (Zhou et al., 2009) method increased BEHR NO_2 VCDs by ~ 5 to 10% over the Rocky Mountains during the summer months. This is a small but systematic change, and so was investigated as a way to correct the low bias in BEHR NO_2 VCDs vs. aircraft and Pandora measurements during the DISCOVER-AQ Colorado campaign.
>
> Figure R1 shows regressions of BEHR VCDs against aircraft + Pandora VCDs using both the scale height and hypsometric equation methods of computing surface pressure. The latter method improves the slope by $\sim$ 18%, and while there is a small increase in most BEHR VCDs, the reduction of the 4 largest aircraft VCDs has a larger effect on the slope. The aircraft VCDs change because, when computing a VCD from the aircraft profiles, we integrate from the OMI pixel surface pressure to its tropopause pressure, for consistency between the aircraft VCD calculation and BEHR AMF calculation.

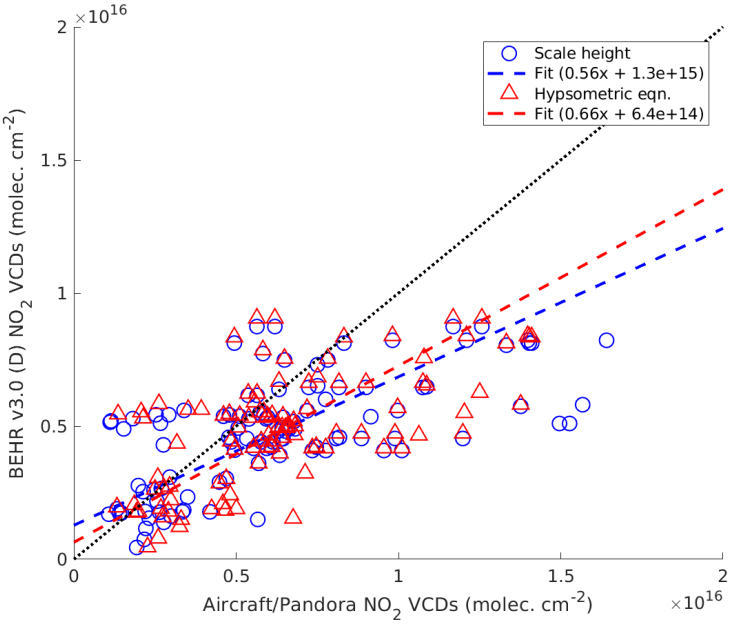

Figure R1: Regression of BEHR v3.0B (D) VCDs vs. aircraft + matched Pandora VCDs for the DISCOVER-CO campaign. The blue series and fit used surface pressured computed using a 7.4 km scale height to convert GLOBE elevations into pressures; the red series used the hypsometric equation as in Zhou et al. (2009).

> As the two methods of calculating surface pressure do not significantly alter the BEHR VCDs in this comparison, we cannot say explicitly that the surface pressure calculated with the hypsometric equation improves the BEHR retrieval. However, using that surface pressure does lead to greater consistency between BEHR and aircraft VCDs when also applied as the lower limit for integrating the aircraft profiles."

Abstract: I suggest including the detail that BEHR is focused on retrievals over North America in the abstract.

We have added:

> "Version 3.0B of the Berkeley High Resolution (BEHR) OMI NO_2 product is designed to accurately retrieve daily variation in the high spatial resolution mapping of tropospheric column NO_2 **over continental North America between** 25° **N and** 50° **N.**"

Introduction: The introduction includes some background on previous evaluation ef- forts for the NASA SP and KNMI DOMINO products. Given the focus of this paper on evaluating this latest version of the BEHR algorithm, some details on how previ- ous versions of this algorithm have been evaluated (and its performance) also seems relevant. This could add more motivation/context for the necessity of updating the al- gorithm (in addition to the

already cited work from Laughner et al. 2016 that focused on the importance of daily profiles).

We have added the following paragraph near the end of the introduction:

> "Russell et al. (2011) evaluated the original BEHR algorithm over California using data from the Arctic Research of the Composition of the Troposphere from Aircraft and Satellites (ARCTAS-CA) field campaign. As the ARCTAS-CA campaign did not include a large number of tropospheric profiles, Russell et al. (2011) computed aircraft-derived NO_2 VCDs from times when the aircraft was flying in the boundary layer. Assuming a well-mixed boundary layer, Russell et al. (2011) extrapolated the measurements within the boundary layer to the surface, and combined with measurements in the free troposphere from the remainder of the ARCTAS-CA campaign, were able to estimate tropospheric NO_2 VCDs from aircraft measurements for a larger number of coincident OMI pixels than would have been possible with traditional aircraft profiles, at the expense of increased uncertainty in the aircraft-derived VCDs. Russell et al. (2011) found that both the original BEHR product had similar agreement as the NASA SP v1 product with the aircraft data (both with slopes near 1), but BEHR had better correlation (R^2 0.83 vs. 0.72). Since then, the plethora of aircraft campaigns and expansion of the Pandora ground based spectrometer network across the United States has provided better datasets to evaluate the BEHR product in a variety of locations."

p. 5, l. 29-30: The authors mention using the GEOS-Chem global chemistry model to extend aircraft profiles to the surface. It could be relevant here to include at least the horizontal resolution of the model output used. Is the model output identical to the model experiment run cited in Nault et al. (2017)? It wasn't clear to me whether these authors were directly using output from that experiment.

This is data from the same GEOS-Chem simulation as Nault et al. 2017. We have clarified this:

> "We use modeled NO_2 profiles from the "updated + 33%" GEOS-Chem simulation described in Nault et al. (2017) (v9.02 of the GEOS-Chem global chemical transport model (Bey et al., 2001) at 2.5° × 2° resolution, with updated HNO_3, HO_2NO_2, and N_2O_5 chemistry and lightning emission rates)."

p. 6, l. 18: I wonder whether some summary statistics might more easily advance the authors' argument that model output from v3.0 "show better agreement" than model output from v2.1. At the moment, there is only a description of the more obvious qualitative details.

We have added Table R1 and the following paragraph after the second paragraph of section 3.1:

> "
>
> We evaluate the agreement quantitatively by calculating the mean absolute bias between the average WRF and aircraft profiles (Table R1). We divide the profiles

	BL ($p > 775$ hPa)	BL (no SEAC4RS)	FT ($p \leq 775$ hPa)	FT (no SEAC4RS)
V2	965	902	71	86
V3 Monthly	530	609	74	87
V3 Daily	482	618	108	66

Table R1: Mean absolute bias between each of the types of simulated NO_2 profiles and the aircraft profiles shown in Fig. 1. Values are given for the boundary layer (BL) and free troposphere (FT), with the divide at 775 hPa ($\sim$ 2 km). All values are in parts per trillion by volume (pptv).

> into boundary layer (BL) and free troposphere (FT), as different processes (e.g. anthropogenic vs. lightning emissions) govern them. As the SEAC4RS campaign has an obvious error in the free troposphere (which will be discussed below), we calculate these values with and without the SEAC4RS campaign. In the BL, the version 3 profiles have one-half to two-thirds the bias of the version 2 profiles (depending if SEAC4RS is excluded). In the free troposphere, there is little difference in the mean bias between profile types, unless SEAC4RS is included, in which case the daily profiles have a 33% greater bias."

p. 6, l. 31: "since the strongest lighting occurs. . ." Replace "lighting" with "lightning"
Corrected, thank you.

p. 9, l. 29: I suggest inserting "modeled" between "daily" and "profiles" so the point is very clear.
Added.

p. 12, l. 2: The word "adequately" is used rather subjectively here. I know it is discussed later in the manuscript, but I wonder whether there is any way to quantitatively evaluate just how well the WRF-Chem NO2 profiles are indeed capturing the day-to- day variability. If not, perhaps this caveat could be more clear, and you could mention here that you propose some suggestions later in the manuscript.

Unfortunately, developing a quantitative metric that accurately evaluates the important aspects of the agreement between the daily variability in the modeled NO_2 profiles and the real world is quite difficult. Structure functions have been used with success before (Follette-Cook et al., 2015), but are insufficient for our application: since structure functions measure average difference vs. separation, a simulated plume and a real plume with similar shapes but going in different directions will still have similar structure functions, even though they differ in a crucial way.

Developing the necessary methods to quantify how well the simulated daily profiles match reality would be an interesting and worthwhile endeavor, unfortunately it is outside the scope of this paper. Therefore we have added the following text to Sect. 3.2:

> "While we recognize that this conclusion is highly qualitative, the specific character of agreement that is important for these profiles (overall plume size and direction, rather than exact agreement between modeled and real concentrations

or column densities) is rather difficult to evaluate quantitatively. We recognize that developing such methods is necessary and offer several possible approaches in Sect. 5."

p. 12, l. 12-15: This explanation of negative VCDs in the DISCOVER-CO dataset is unclear to me.

We have expanded this explanation and reorganized the surrounding text slightly to accomodate it:

"For the DISCOVER-CO aircraft comparison, negative VCDs were removed. Negative VCDs occur when the estimated stratospheric NO_2 column is greater than the total NO_2 column, thus $V_{trop} = V_{total} - V_{strat} < 0$; they cannot be introduced by the AMF correction of the tropospheric SCD to VCD as the AMF is a multiplicative factor and always > 0. Since all versions of BEHR use the same stratospheric NO_2 column as their respective NASA SP products, an error in stratospheric subtraction will be present in all products, and it cannot be corrected in the BEHR retrieval. Aircraft VCDs, by their nature, cannot be negative, so for these comparisons we remove the negative VCDs so as to avoid increasing the regression slopes by trying to fit these erroneous points. (However, we do note that this is a special case where individual pixels or small groups of pixels are being compared against other VCDs. Most applications of BEHR data should retain the negative VCDs to avoid transforming the essentially Gaussian random stratospheric error into a systematic error by removing part of the bell curve.) Since the stratospheric VCDs are added back to the BEHR or NASA SP tropospheric VCDs for comparison with the Pandora VCDs, negative VCDs are not an issue with Pandora comparisons."

p. 12, l. 28-30: The authors introduce the notation of "BEHR v.30 (M)" and "(D)" for the first time here, I think. While it's fair to say it is obvious, the authors could explicitly clarify that "(M)" refers to the product using monthly average profiles, and "(D)" refers to the product using daily profiles

We have added the following sentence to the end of Sect. 4:

"Throughout, BEHR v3.0 (M) refers to BEHR using monthly NO_2 profiles; likewise, BEHR v3.0 (D) refers to the product using daily NO_2 profiles."

Table 2 and 3: I was wondering whether there would be value in reporting correlation coefficients in addition to slopes. Can the authors explain why they haven't included these in their evaluation of the product performance? For example, I wondered whether they can demonstrate that in addition to improving a bias, using daily profiles (D) explain more of the variability than monthly profiles (M) alone.

We do report R^2 values in the supplement (Table S3). We chose not to include them in the main paper as this section was already fairly dense (with a three-way comparison between

NASA, BEHR v2, and BEHR v3) and the slopes are the main indicator of the accuracy of each products' columns. Further, there is no clear pattern in the R^2 values. We have added the following sentence in Sect. 4 stating this:

> "We will focus on the regression slopes here; intercepts and R^2 values are given in Table S3 in the supplement; however we note that there is not a clear pattern of any one product having a consistently better R^2 value than the others."

Section 4.2: Can the authors clarify why they have chosen to separate their evaluation by looking at all the Pandora data, vs. just the Pandora data during coincident aircraft spirals? Is this meant to demonstrate how continued long-term monitoring is superior to short-term campaign coverage for evaluation purposes?

We should clarify that both cases include aircraft and Pandora data. Our main purpose was to provide both a regression that used all available data and one that weights the Pandora and aircraft data equally; when using all Pandora data, the sheer number of data points essentially overwhelms the aircraft data. Since both measurements have strengths and weaknesses for comparison with satellite columns, we wanted to show both a comparison over long time periods (all Pandora data) and one that gave the aircraft data fair weight. We have expanded Sect. 4 to describe this:

> "For the DISCOVER campaigns, we compare BEHR against aircraft-derived and Pandora VCDs together, calculating a single regression line for the combined dataset. These two measurements have unique strengths and weaknesses for comparison against satellite VCDs: Pandoras give a precise column measurement and can be deployed for long time periods, but have a very small footprint (leading to possible representativeness errors) and provide a total, not tropospheric, column. Aircraft profiles have a footprint more similar to an OMI pixel size, but introduce uncertainty due to missing parts of the profile (near the surface and in the upper troposphere in the DISCOVER campaigns) and cannot be deployed for long term, routine observations.
>
> In order to take advantage of each methods' strenghts, we use two comparisons, in one, only Pandora data that has a coincident aircraft profile is include ("matched"), in the other, all cloud-free Pandora data is used ("all"). We do so because, when including all Pandora data, the number of Pandora comparisons available will overwhelm the number of available aircraft profiles in the regression. Therefore the regressions using all Pandora data are representative of longer time periods, but weighted strongly towards the Pandora data, and the regressions using only the coincident data represent shorter time periods, but give more weight to the aircraft data."

We also remind the reader that these regressions are aircraft+Pandora in Sect. 4.2:

> "Using **aircraft data plus** just Pandora data coincident with aircraft spirals, v2.1 performs better.... However, using **aircraft data plus** all Pandora data, v3.0 (D) performs better...."

References

Bey, I., Jacob, D. J., Yantosca, R. M., Logan, J. A., Field, B. D., Fiore, A. M., Li, Q., Liu, H. Y., Mickley, L. J., and Schultz., M. G.: Global modeling of tropospheric chemistry with assimilated meteorology, J. Geophys. Res., 106, 23 073–23 096, 2001.

Bucsela, E., Krotkov, N., Celarier, E., Lamsal, L., Swartz, W., Bhartia, P., Boersma, K., Veefkind, J., Gleason, J., and Pickering, K.: "A new tropospheric and stratospheric NO_2 retrieval algorithm for nadir-viewing satellite instruments: applications to OMI, Atmos. Meas. Tech., 6, 2607–2626, doi:10.5194/amt-6-2607-2013, 2013.

Follette-Cook, M. B., Pickering, K. E., Crawford, J. H., Duncan, B. N., Loughner, C. P., Diskin, G. S., Fried, A., and Weinheimer, A. J.: Spatial and temporal variability of trace gas columns derived from WRF/Chem regional model output: Planning for geostationary observations of atmospheric composition, Atmos. Environ., 118, 28–44, doi:10.1016/j.atmosenv.2015.07.024, URL https://doi.org/10.1016/j.atmosenv.2015.07.024, 2015.

Hastings, D. and Dunbar, P.: Global Land One-kilometer Base Elevation (GLOBE) Digital Elevation Model, Documentation, Volume 1.0. National Oceanic and Atmospheric Administration, National Geophysical Data Center, 325 Broadway, Boulder, Colorado 80303, U.S.A., 1999.

Laughner, J. L., Zhu, Q., and Cohen, R. C.: The Berkeley High Resolution Tropospheric NO_2 Product, Earth System Science Data Discussions, 2018, 1–33, doi:10.5194/essd-2018-66, URL https://www.earth-syst-sci-data-discuss.net/essd-2018-66/, 2018.

Nault, B. A., Laughner, J. L., Wooldridge, P. J., Crounse, J. D., Dibb, J., Diskin, G., Peischl, J., Podolske, J. R., Pollack, I. B., Ryerson, T. B., Scheuer, E., Wennberg, P. O., and Cohen, R. C.: Lightning NO_x Emissions: Reconciling Measured and Modeled Estimates With Updated NO_x Chemistry, Geophys. Res. Lett., doi:10.1002/2017GL074436, 2017.

Russell, A. R., Perring, A. E., Valin, L. C., Bucsela, E. J., Browne, E. C., Wooldridge, P. J., and Cohen, R. C.: A high spatial resolution retrieval of NO_2 column densities from OMI: method and evaluation, Atmos. Chem. Phys., 11, 8543–8554, doi:10.5194/acp-11-8543-2011, URL https://doi.org/10.5194/acp-11-8543-2011, 2011.

Vasilkov, A., Qin, W., Krotkov, N., Lamsal, L., Spurr, R., Haffner, D., Joiner, J., Yang, E.-S., and Marchenko, S.: Accounting for the effects of surface BRDF on satellite cloud and trace-gas retrievals: a new approach based on geometry-dependent Lambertian equivalent reflectivity applied to OMI algorithms, Atmospheric Measurement Techniques, 10, 333–349, doi:10.5194/amt-10-333-2017, 2017.

Zhou, Y., Brunner, D., Boersma, K. F., Dirksen, R., and Wang, P.: An improved tropospheric NO_2 retrieval for OMI observations in the vicinity of mountainous terrain, Atmos. Meas. Tech., 2, 401–416, doi:10.5194/amt-2-401-2009, URL https://www.atmos-meas-tech.net/2/401/2009/, 2009.

Evaluation of version 3.0B of the BEHR OMI NO_2 product

Response to Anonymous Referee #2

Joshua L. Laughner, Qindan Zhu, and Ronald C. Cohen

November 8, 2018

We thank the reviewer for their positive comments. We acknowledge that there is a large volume of information presented here, and believe that the suggestions we have incorporated should help make the major points more apparent to the reader.

Responses to specific comments follow. The reviewer's comments will be shown in red, our response in blue, and changes made to the paper are shown in black block quotes. Unless otherwise indicated, page and line numbers correspond to the original paper. Figures, tables, or equations referenced as "Rn" are numbered within this response; if these are used in the changes to the paper, they will be replaced with the proper number in the final paper. Figures, tables, and equations numbered normally refer to the numbers in the original discussion paper.

Section 4. The comparison of VCD should include some scatterplots. Presenting the comparison as a table (Table 2) only is a bit difficult to follow. Especially, the results with separated Pandora and aircraft VCDs should be shown. Some of these plots could go in the supplementary material.

Scatterplots with separate aircraft and Pandora data were already included in the supplement. Figures S1-S6 show separate aircraft and, where available, Pandora comparisons for each of the 6 campaigns used for validation. We have added a figure showing scatter plots for the combined aircraft and Pandora data to the main paper (Fig. R1).

You use a quite precise spatial collocation criteria (pandora site within the OMI pixel, so no spatial smoothing in practice) and then you time average Pandora observations ±1 h from the OMI observation, which is quite large time frame. Can you open a little bit about this choice? How does the results change with a shorter time interval?

The ±1 h averaging is the same as done in Goldberg et al. (2017). Part of our reason for using that averaging window is to make our results comparable with that study. Given the average wind for cities where the DISCOVER-AQ campaigns took place (4.5 m/s, from WRF simulations), a shorter averaging time window, e.g. ±0.5 h, might make more sense as this means an airmass would travel ∼16 km, which is similar to the length of an OMI pixel at nadir in either dimension (13 or 24 km). However, no slope changes by > 0.05 (max $\sim 8\%$) using the shorter window, so we prefer to be consistent with Goldberg et al. (2017) so that readers can compare our results. We have added a sentence to sect. 4 explaining this:

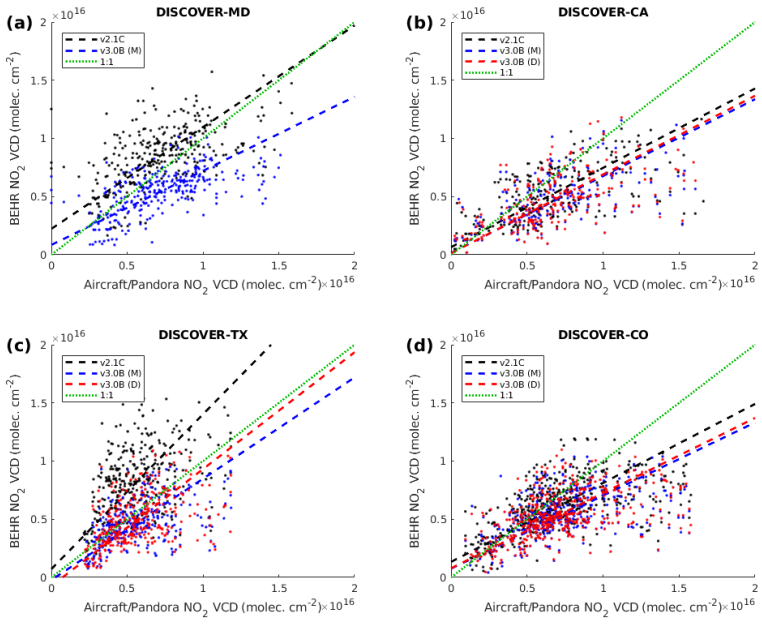

Figure R1: Scatter plots of the BEHR v2.1C, v3.0B (M), and (where available) v3.0B (D) VCDs against coincident aircraft and Pandora VCDs. All Pandora VCDs are used for these plots. Each panel is one campaign: the **(a)** Maryland, **(b)** California, **(c)**, Texas, and **(d)** Colorado DISCOVER-AQ campaigns. For slopes, see Table 3; for intercepts and R^2 values, see Table S3.

> "As stated in sect. 2.3, we average all data within 1 hour of OMI overpass (i.e. 13:30 local time ±1 h) to be consistent with Goldberg et al. (2017). A shorter averaging window (±0.5 h) was tested; the maximum effect on the slope was $\sim$ 8% with most of the "matched" data showing differences of $\leq$ 5% and the "all" data changing by $\leq$ 3.5% in all but one case."

You mention several times in the text that changes in emission information as input have a role in the discrepancies you observe between different versions of the algorithm. Could you actually show them? For example, plotting or mentioning the quantitative the emission changes over the areas of study.

While the emissions files used in the original BEHR product are no longer available, we have reproduced their likely value based on the description in Russell et al. (2011) and Russell et al. (2012) and added a figure to the supplement (Fig R2).

Fig. 1: the plot is quite small, it could be a bit bigger so that the the different lines can be better separated? Also, the legend can be shown only once.

We have done our best to enlarge this figure within the space constraints of the AMT template.

P12 L11 only Pandora data that has a coincident aircraft profile is include -¿ only Pandora

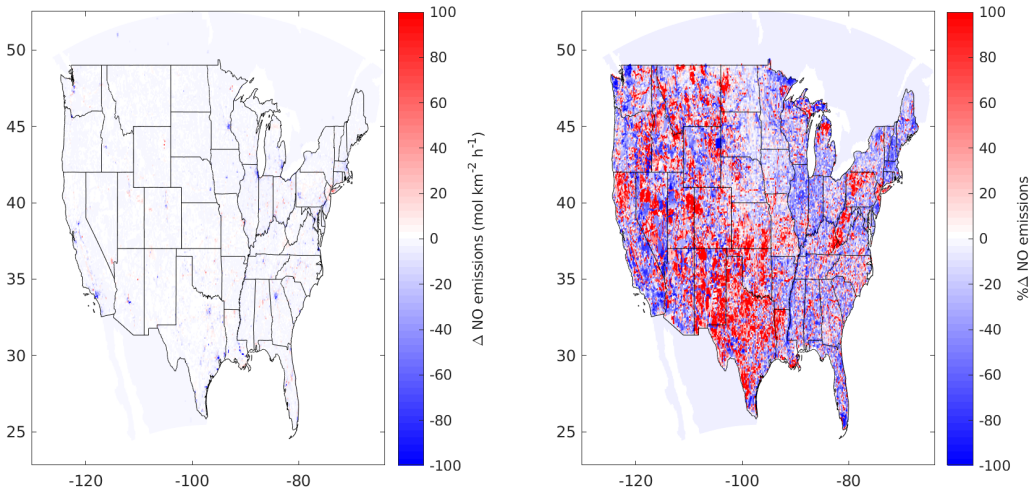

Figure R2: Absolute **(a)** and percent **(b)** change in WRF-Chem NO emissions between the EPA NEI 2005 inventory, unscaled, and the EPA NEI 2011 inventory, scaled to 2012 levels.

data that have a coincident aircraft profile are included

Corrected, thank you.

Conclusions P17 L12 I think you refer to Sentinel 4, as Sentinel 5 is not planned on a geostationary orbit.

Yes, thank you for the correction.

Figure 1 and 3: the different panels would benefit from a title mentioning the different sites considered in the plot

Added.

References

Goldberg, D. L., Lamsal, L. N., Loughner, C. P., Swartz, W. H., Lu, Z., and Streets, D. G.: A high-resolution and observationally constrained OMI NO_2 satellite retrieval, Atmos. Chem. Phys., 17, 11 403–11 421, doi:10.5194/acp-17-11403-2017, URL `https://doi.org/10.5194/acp-17-11403-2017`, 2017.

Russell, A. R., Perring, A. E., Valin, L. C., Bucsela, E. J., Browne, E. C., Wooldridge, P. J., and Cohen, R. C.: A high spatial resolution retrieval of NO_2 column densities from OMI: method and evaluation, Atmos. Chem. Phys., 11, 8543–8554, doi:10.5194/acp-11-8543-2011, URL `https://doi.org/10.5194/acp-11-8543-2011`, 2011.

Russell, A. R., Valin, L. C., and Cohen, R. C.: Trends in OMI NO_2 observations over the United States: effects of emission control technology and the economic recession, Atmos. Chem. Phys., 12, 12 197–12 209, doi:10.5194/acp-12-12197-2012, 2012.

Evaluation of version 3.0B of the BEHR OMI NO_2 product

Joshua L. Laughner[1], Qindan Zhu[2], and Ronald C. Cohen[1,2]

[1]Department of Chemistry, University of California, Berkeley, Berkeley, CA 94720
[2]Department of Earth and Planetary Sciences, University of California, Berkeley, Berkeley, CA 94720

Correspondence: Ronald C. Cohen (rccohen@berkeley.edu)

Abstract. Version 3.0B of the Berkeley High Resolution (BEHR) OMI NO_2 product is designed to accurately retrieve daily variation in the high spatial resolution mapping of tropospheric column NO_2 over continental North America between 25° N and 50° N. To assess the product, we compare against in situ aircraft profiles and Pandora vertical column densities (VCDs). We also compare the WRF-Chem simulation used to generate the a priori NO_2 profiles against observations. We find that using daily NO_2 profiles improves the VCDs retrieved in urban areas relative to low resolution or monthly a priori by amounts that are large compared to current uncertainties in NO_x emissions and chemistry (of order 10% to 30%). Based on this analysis, we offer suggestions to consider when designing retrieval algorithms and validation procedures for upcoming geostationary satellites.

1 Introduction

NO_x ($\equiv NO + NO_2$) is an atmospheric trace gas emitted by anthropogenic activity (predominantly combustion, e.g. motor vehicles and power plants), lightning, biomass burning, and soil microbes. It plays an important role in air quality, as a major controlling factor in ozone and aerosol production, as well as being toxic itself.

Satellite observations of NO_2 have proven to be extremely useful in constraining anthropogenic (e.g. Richter et al., 2005; Kim et al., 2006, 2009; van der A et al., 2008; Konovalov et al., 2010; Russell et al., 2010; Beirle et al., 2011; Castellanos and Boersma, 2012; Russell et al., 2012; Zhou et al., 2012; McLinden et al., 2014; Lu et al., 2015; Liu et al., 2016, 2017; Miyazaki et al., 2012, 2017), lightning (e.g. Beirle et al., 2004; Martin et al., 2007; Beirle et al., 2010; Bucsela et al., 2010; Miyazaki et al., 2014; Pickering et al., 2016; Nault et al., 2017), soil (e.g. Bertram et al., 2005; van der A et al., 2008; Hudman et al., 2010, 2012; Zörner et al., 2016), and biomass burning (e.g. Mebust et al., 2011; Huijnen et al., 2012; Mebust and Cohen, 2013, 2014; Bousserez, 2014; Schreier et al., 2014; Castellanos et al., 2015; van Marle et al., 2017) emissions.

Satellite observations of NO_2 relate absorption of light in the $\sim$ 400–460 nm range of reflected Earthshine radiances to a total column measurement of NO_2 using differential optical absorption spectroscopy (DOAS, Boersma et al., 2001; Richter and Wagner, 2011) or a similar technique (e.g. van Geffen et al., 2015). Most applications of satellite NO_2 observations to constrain emissions or otherwise study air quality are focused on the tropospheric contribution to the total column; therefore the stratospheric column must be removed. Several methods have been implemented to do so (e.g. Boersma et al., 2007; Bucsela et al., 2013). The tropospheric slant column density (SCD) is then converted to a vertical column density (VCD) through the

use of an air mass factor (AMF, McKenzie et al., 1991; Slusser et al., 1996; Burrows et al., 1999; Palmer et al., 2001) that accounts for the effect of path length, surface reflectivity and elevation, NO_2 vertical distribution, clouds, and aerosols.

There have been numerous studies evaluating OMI NO_2 products against in situ aircraft profiles and ground based column measurements. This is not meant to be an exhaustive list, but to provide a summary of the results of evaluations of existing standard OMI NO_2 products.

The first-generation NASA Standard Product (SP) and KNMI DOMINO products were evaluated by Bucsela et al. (2008) and Hains et al. (2010) using aircraft profiles from multiple campaigns and Russell et al. (2011) using an extrapolation method with ARCTAS-CA aircraft data. These studies all identified a high bias in the DOMINO VCDs; by comparing the DOMINO a priori profiles to aircraft and lidar profiles Hains et al. (2010) found evidence that this was caused by insufficient vertical mixing in the DOMINO a priori profiles, which was corrected in DOMINO v2.

Lamsal et al. (2014) undertook a detailed evaluation of the NASA SP v2, primarily focusing on data from the Deriving Information on Surface Conditions from COlumn and VERtically Resolved Observations Relevant to Air Quality (DISCOVER-AQ) campaign in Baltimore, MD, USA. This work combined evaluation of the a priori profile against aircraft measurements along with validation of OMI VCDs with aircraft and ground-based VCDs. They found that the NASA SP v2 VCDs were generally biased low in urban areas and high in rural or suburban areas. This is consistent with the effect of coarse a priori profiles (Russell et al., 2011); in a large urban area like the Baltimore/Washington D.C. urban corridor, a coarse profile can capture the average urban characteristic profile, but on the edge, a coarse profile cannot capture the transition from urban to rural.

Krotkov et al. (2017) and Goldberg et al. (2017) both evaluated the NASA SP v3, primarily using ground based VCD observations. They found it to be biased low by $\sim 50\%$ in the Baltimore area (Goldberg et al., 2017) and 50% or more Hong Kong (Krotkov et al., 2017), but better than SP v2 in remote areas, due to the improved total column fitting implemented in version 3. Ialongo et al. (2016) also compared versions 2 and 3 of the NASA SP and version 2 of DOMINO against ground based column measurements in Helsinki, one of only a few studies at high latitudes ($> 60°$). They found that SP v3 was biased 30% low, while the version 2 products were not. They attributed this to cancellation of errors in the version 2 products, namely the high bias in the total OMI columns corrected by van Geffen et al. (2015), and the representativeness mismatch between OMI pixels and Pandora measurements.

Russell et al. (2011) evaluated the original BEHR algorithm over California using data from the Arctic Research of the Composition of the Troposphere from Aircraft and Satellites (ARCTAS-CA) field campaign. As the ARCTAS-CA campaign did not include a large number of tropospheric profiles, Russell et al. (2011) computed aircraft-derived NO_2 VCDs from times when the aircraft was flying in the boundary layer. Assuming a well-mixed boundary layer, Russell et al. (2011) extrapolated the measurements within the boundary layer to the surface, and combined with measurements in the free troposphere from the remainder of the ARCTAS-CA campaign, were able to estimate tropospheric NO_2 VCDs from aircraft measurements for a larger number of coincident OMI pixels than would have been possible with traditional aircraft profiles, at the expense of increased uncertainty in the aircraft-derived VCDs. Russell et al. (2011) found that both the original BEHR product had similar agreement as the NASA SP v1 product with the aircraft data (both with slopes near 1), but BEHR had better correlation (R^2

0.83 vs. 0.72). Since then, the plethora of aircraft campaigns and expansion of the Pandora ground based spectrometer network across the United States has provided better datasets to evaluate the BEHR product in a variety of locations.

Here we present an evaluation of version 3.0B of the BErkeley High Resolution (BEHR) OMI NO_2 retrieval. Version 3.0B implements several changes over v2.1C:

- Daily profiles for selected years
- Updated 12 km WRF-Chem NO_2 profiles with a more complete chemical mechanism (Zare et al., 2018), updated anthropogenic emissions, lightning NO_x emissions added
- Use of v3.0 NASA Standard Product (SP) tropospheric SCDs
- Directional surface reflectance
- Variable tropopause height
- Surface pressure combining a high resolution terrain database with WRF-simulated surface pressure

The motivation for this upgrade stems from ideas developed in Laughner et al. (2016), where we showed that daily, high resolution a priori profiles are necessary for a retrieval to simultaneously retrieve NO_x VCDs and lifetime to accuracies better than 30%. As our goal is to study the relationship between changes in NO_x VCDs/emissions and NO_x lifetime across the US, and resolving open questions requires higher relative precision and high accuracy than prior retrievals, we have developed a new product with daily 12 km a priori profiles. Therefore, in this work, we first evaluate the simulated WRF-Chem profiles against aircraft measurements and OMI SCDs to demonstrate that the daily profiles accurately represent the real atmosphere. We then directly evaluate the retrieved VCDs using both aircraft and Pandora observations and show that v3.0 is generally superior to v2.1C, and that using daily profiles improves the overall quality of the retrieval.

2 Methods: models and observations

2.1 BEHR

The BEHR OMI NO_2 retrieval is described in detail in Laughner et al. (2018f). Briefly, the BEHR retrieval calculates a tropospheric air mass factor (AMF) using high resolution a priori input data for surface reflectance, surface elevation, and NO_2 vertical profiles; the NO_2 profiles are simulated with WRF-Chem (Sect. 2.2). To capture the day-to-day variation in NO_2 profiles, daily profiles are used. Currently, 2005, 2007–2009, and 2012–2014 are available. Other years will be posted as processing is completed. A second subproduct uses monthly average profiles (simulated for 2012) to retrieval all years of the OMI data record.

The BEHR AMF is used to convert the tropospheric slant column densities (SCDs) available in the NASA OMI NO_2 standard product to tropospheric vertical column densities (VCDs). For full details of the AMF calculation, see Laughner et al. (2018f). The BEHR product is available for download as HDF version 5 files at behr.cchem.berkeley.edu.

2.2 WRF-Chem

The WRF-Chem model version used to simulate the a priori NO_2 profiles for BEHR v3.0B is v3.5.1 (Grell et al., 2005). The model domain is 405 (east-west) by 254 (north-south) 12 km grid cells centered on 39° N 97° W with 29 vertical levels. Meterological initial, boundary, and nudging conditions are taken from the North American Regional Reanalysis (NARR) product; boundary conditions and four-dimension data analysis (FDDA) nudging (Liu et al., 2006) is applied every 3 hours. Temperature, water vapor, and U/V winds are nudged with nudging coefficients of 0.0003 s^{-1}.

The chemical mechanism used is described in Zare et al. (2018), which has a very detailed description of alkyl nitrate and nighttime chemistry. Methyl peroxynitrate (MPN) chemistry was added (Browne et al., 2011) to improve upper tropospheric chemistry. Anthropogenic emissions are from the National Emissions Inventory, 2011, scaled by EPA annual total emissions (EPA, 2016) to the model year. Biogenic emissions are from the Model for Emissions of Gases and Aerosols from Nature (Guenther et al., 2006). Lightning emissions are parameterized following Laughner and Cohen (2017) for a simulation with FDDA active (500 mol NO $flash^{-1}$, 2x base flashrate).

Chemical initial and boundary conditions are interpolated to the WRF grid using the MOZBC utility (https://www2.acom.ucar.edu/wrf-chem/wrf-chem-tools-community). For 2007 and later model years, chemical data is obtained from the MOZART model runs available at https://www.acom.ucar.edu/wrf-chem/mozart.shtml. For 2005 and 2006, chemical data is obtained from a GEOS-Chem model run, described in Laughner et al. (2018f).

2.3 Pandora ground-based columns

Evaluation of satellite NO_2 VCDs usually uses one of two methods. First, total satellite columns can be directly compared to a ground-based column measurement, such as a Pandora spectrometer (Herman et al., 2009) or multi-axis DOAS (MAX-DOAS) instrument (Hönninger et al., 2004). In the case of a direct-sun measurement, such as a Pandora spectrometer, the AMF required is only a geometric AMF to account for the path length difference between the slant and vertical column, since the multiple scattering that necessitates the use of a more complex AMF in the satellite retrieval is a much smaller signal than the direct-sun signal (Herman et al., 2009).

We compare against Pandora ground based column measurements taken during the four DISCOVER-AQ campaigns. For each OMI overpass, pixels are matched with Pandora sites that lie within the pixel boundaries defined by the FoV75 corners in the OMPIXCOR product (Kurosu and Celarier, 2010). Only pixels meeting the criteria in Table 1 are used. If multiple valid pixels from the same overpass encompass the Pandora site, their VCDs are averaged. As in Goldberg et al. (2017), the stratospheric VCD from the NASA Standard Product is added to the tropospheric VCD to obtain a total column, since the Pandora columns do not separate stratospheric and tropospheric contributions.

Pandora observations are matched in time to the OMI observations using the exact time of observation for each pixel given in the OMI data files. As in Goldberg et al. (2017), Pandora observations ± 1 h from the OMI observation are averaged.

Data field	Condition
XTrackQualityFlags	Must be 0
VcdQualityFlags	Must be an even number
CloudFraction	Must be ≤ 0.2
BEHRAMFTrop	Must be a non-fill value $> 10^{-6}$

Table 1. Criteria that OMI pixels must meet to be used in any comparison.

2.4 In situ aircraft profiles

The other common method of evaluating satellite VCDs is to use in situ measurements of NO_2 by instrumented aircraft that flies a vertical profile to calculate a VCD by integrating the NO_2 concentrations vertically. Ideally, the aircraft should fly a spiral path that provides a complete vertical sampling of the troposphere over a ground footprint similar in scale to the satellite pixel; the DISCOVER-AQ campaigns held in Maryland, California, Texas, and Colorado between 2011 and 2014 were designed to provide this sampling over the lower troposphere. In other cases, the VCD calculated from integrating the aircraft profiles is often matched to satellite pixels in which the boundary layer is sampled (e.g. Bucsela et al., 2008; Hains et al., 2010), on the assumption that upper troposphere (UT) sampling from adjacent pixels is sufficient.

We calculate tropospheric VCDs from in situ NO_2 profiles measured from aircraft. We use six campaigns: the four DISCOVER-AQ campaigns (https://www-air.larc.nasa.gov/missions/discover-aq/discover-aq.html) in Maryland (2011), California (2013), Texas (2013), and Colorado (2014), the Southeast Nexus campaign (2013, southeast US, SENEX Science Team, 2013), and the Studies of Emissions and Atmospheric Composition, Clouds, and Climate Coupling by Regional Surveys (SEAC4RS, 2013, Toon et al., 2016). For the DISCOVER-AQ and SEAC4RS campaigns, we use 1 second NO_2 data from the TD-LIF instrument is used (Nault et al., 2015; Wooldridge et al., 2010; Day et al., 2002; Thornton et al., 2000). For the SENEX campaign, we use 1 second data from the chemiluminescence instrument (Ryerson et al., 1999).

We draw on methodology from several papers (Bucsela et al., 2008; Hains et al., 2010; Lamsal et al., 2014) for our approach. Similar to (Hains et al., 2010), only profiles with a minimum radar altitude < 500 m and at least 20 measurements below 3 km above ground level (AGL) are used. In the DISCOVER-AQ campaigns, individual profiles are demarcated in the data by a profile number. In the SENEX and SEAC4RS data, profiles were identified manually as periods when the aircraft was consistently ascending or descending. The profile measurements are binned to the same pressure levels used in the BEHR algorithm and the final profile uses the median of each bin.

Profiles are spatially matched to OMI pixels if any of the 1 second measurements in the bottom 3 km AGL lies within the FoV75 pixel boundaries. As with Pandora data, OMI pixels must meet the criteria in Table 1 to be included; all VCDs from valid pixels intersecting the profile are averaged to yield a single VCD to compare against the profile. Only profiles with an mean observation time of all points in the bottom 3 km AGL within 1.5 h of the mean OMI observation time for the orbit are used.

To calculate a VCD from the in situ measurements, the aircraft profiles are integrated from the average surface pressure to the average tropopause pressure of the matched pixels. The surface and tropopause pressure are used from the product being evaluated, i.e. aircraft profiles are integrated between BEHR surface and tropopause pressure for comparison with BEHR VCDs and NASA surface and tropopause pressures for comparison with NASA VCDs. For BEHR v2.1C comparisons, 200 hPa is used as the fixed tropopause pressure. Aircraft profiles that do not span the necessary vertical extent are extended similarly to Lamsal et al. (2014). The aircraft profile is extended to the surface by using the ratio of modeled concentrations at each of the missing levels to the lowest level with aircraft data to scale the bottom bin with aircraft data. Missing profile levels above the top of the aircraft profile are replaced with model data. We use modeled NO_2 profiles from the "updated + 33%" GEOS-Chem simulation described in Nault et al. (2017) (v9.02 of the GEOS-Chem global chemical transport model (Bey et al., 2001) at $2.5° \times 2°$ resolution, with updated HNO_3, ~~with the updated~~ HO_2NO_2, and N_2O_5 chemistry and lightning emission rates~~described in Nault et al. (2017)~~). The NO_2 profiles are monthly averages of model output from 2012 sampled between 12:00 and 14:00 local standard time. We avoid using the a priori WRF-Chem profiles for this so that the aircraft VCDs are independent of the retrieved VCDs.

We also used the extrapolation method from Hains et al. (2010), where the median of the top 10 and bottom 10 points are extrapolated to the tropopause and surface pressures, respectively. The median of the top 10 points must be < 100 pptv. As in Hains et al. (2010), a detection limit of 3 pptv is assumed, and if the median to be extrapolated is less than 3 pptv, it is set to one-half of the detection limit, 1.5 pptv.

In addition, we directly compare the a priori profiles to the in situ aircraft profiles. This is done as in Laughner and Cohen (2017); for each 1 second data point in the aircraft data, the nearest WRF-Chem output time is selected, and the model grid cell containing the aircraft location is sampled. This effectively samples the model output as if the aircraft were flying through the model world.

We use a similar set of aircraft campaigns here as for the VCD evaluation (Sect. 2.4); the only difference being that we use the Deep Convective Clouds and Chemistry (Barth et al., 2015) instead of SENEX. The DC3 campaign focused on outflow from convective systems (i.e. thunderstorms) and so is used to evaluate the lightning NO_x parameterization. The DC3 campaign had better UT sampling but far fewer profiles than SENEX. The DISCOVER-AQ campaigns focused on satellite validation, flying repeated spirals over 6–8 sites during each campaign; however, for the average comparison, we use all data, not just that taken during the spirals.

3 WRF-Chem profile evaluation

3.1 Comparison with in situ aircraft profiles

Figure 1 shows campaign averaged profiles matched with WRF profiles from the four DISCOVER-AQ campaigns, the DC3 campaign, and the SEAC4RS campaign. We compare the monthly average NO_2 profiles from BEHR v2.1C and v3.0B for all campaigns, as well as the daily v3.0B profiles. The plots shown only use data between 12:00 and 15:00 local standard time,

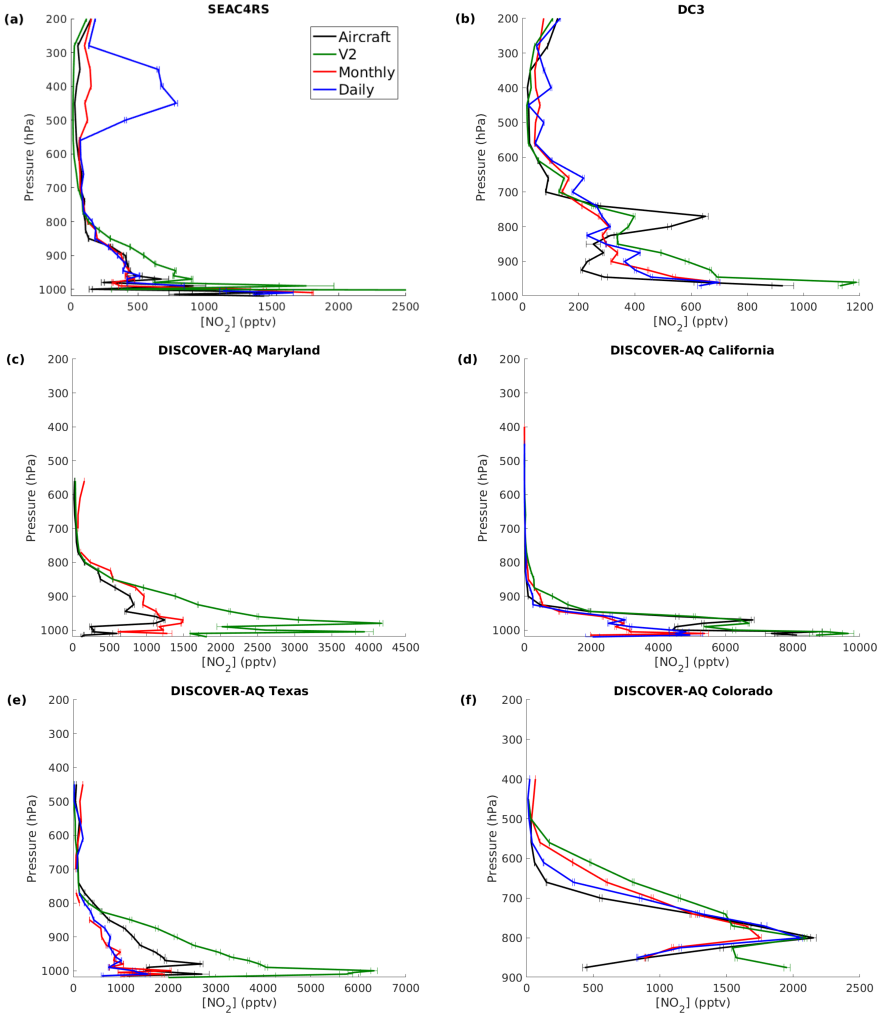

Figure 1. Comparison of average WRF-Chem and aircraft NO_2 profiles from the **(a)** SEAC4RS, **(b)** DC3, and DISCOVER-AQ campaigns, the latter in **(c)** Maryland, **(d)** California, **(e)** Texas, and **(f)** Colorado. Aircraft profiles are shown in black, BEHR v2.1 profiles in green, BEHR v3.0 monthly profiles in red, and (where available) BEHR v3.0 daily profiles in blue. The WRF and aircraft data are matched as described in Sect. 2.4 and binned by pressure. Uncertainties are 1 standard deviation of all profiles averaged. Note that for SEAC4RS the v2 profile reaches a maximum of $\sim$ 8000 pptv, off the plot axes.

	BL ($p > 775$ hPa)	BL (no SEAC4RS)	FT ($p \leq 775$ hPa)	FT (no SEAC4RS)
V2	965	902	71	86
V3 Monthly	530	609	74	87
V3 Daily	482	618	108	66

Table 2. Mean absolute bias between each of the types of simulated NO_2 profiles and the aircraft profiles shown in Fig. 1. Values are given for the boundary layer (BL) and free troposphere (FT), with the divide at 775 hPa ($\sim$ 2 km). All values are in parts per trillion by volume (pptv).

since the v3.0 monthly average profiles are calculated as a weighted average that only includes contributions from ± 1 h from OMI overpass; this way all profiles get a fair comparison to the observations.

In general, the v3.0 profiles show better agreement with observed profiles than the v2.1 profiles, except during the California DISCOVER-AQ campaign. The most dramatic example is the Maryland DISCOVER-AQ campaign, where the factor of ~ 2 reduction in NO_2 concentration (likely due to updating emissions from 2005 to 2012, Fig. S10) brings the modeled profiles into substantially better agreement with the observed profiles. In the California DISCOVER-AQ campaign, the v2.1 profiles managed to capture an elevated layer of NO_2 that the v3.0 profiles did not; though we note that transport in California's central valley is notorious difficult to model (Hu et al., 2010, and references therein). In Texas, the v3.0 profiles and v2.1 profiles lie on opposite sides of the observed profiles, possibly suggesting that emissions in Houston did not decrease as much in fact as in the NEI inventory driving the v3.0 WRF simulations. In Colorado, both the v3.0 and v2.1 profile match observations reasonably well. The daily profiles do a better job capturing the decrease in NO_2 between 750 and 600 hPa than the v3.0 monthly or v2.1 profiles; this may be due to day-to-day variability in recirculation from the upslope/downslope winds (e.g. Sullivan et al., 2016).

We evaluate the agreement quantitatively by calculating the mean absolute bias between the average WRF and aircraft profiles (Table 2). We divide the profiles into boundary layer (BL) and free troposphere (FT), as different processes (e.g. anthropogenic vs. lightning emissions) govern them. As the SEAC4RS campaign has an obvious error in the free troposphere (which will be discussed below), we calculate these values with and without the SEAC4RS campaign. In the BL, the version 3 profiles have one-half to two-thirds the bias of the version 2 profiles (depending if SEAC4RS is excluded). In the free troposphere, there is little difference in the mean bias between profile types, unless SEAC4RS is included, in which case the daily profiles have a 33% greater bias.

We include the SEAC4RS and DC3 campaigns to check the simulation of lightning NO_x in the profiles. The daily profiles show similar agreement to the DC3 observations as in Laughner and Cohen (2017). Restricting the DC3 data to 12:00–15:00 local standard time as we have done here reduces the strength of the lightning signal, since the strongest ~~lighting~~ lightning occurs after OMI overpass (Lay et al., 2007; Williams et al., 2000). Compared to Laughner and Cohen (2017), the discrepancy between modeled and observed profiles decreased around 500 hPa, increased around 400 hPa, and is similarly small around 200 hPa. Surprisingly, the difference between the v2.1 and v3.0 profile around 200 hPa is not as significant as the difference

between the lightning and no-lighting cases in Laughner and Cohen (2017). This is unexpected as the v2.1 profiles did not include lightning NO_x emission. It is possible that convection of greater surface NO_x concentrations is driving the v2.1 UT concentration.

The SEAC4RS campaign covers the southeast US, which has very active lightning (Hudman et al., 2007; Travis et al., 2016). The daily profiles demonstrate a substantial overestimate in UT NO_2 (between 600 and 200 hPa). This is centered in the SE US; model-measurement discrepancies between 600 and 200 hPa in the rest of the country are < 500 pptv (not shown). As discussed in Laughner et al. (2018f), the southeast US exhibits greater NO_2 VCDs (and therefore smaller AMFs) when using daily profiles; that is opposite with the profiles seen here, as greater NO_2 at higher altitudes results in larger AMFs. Laughner et al. (2018f) showed that the 3 month average daily shape factor over the SE US had less contribution from UT NO_2 than the monthly profiles; this indicates that on average pixels in the SE US are not influenced by lightning, but that the SEAC4RS sampling tended to select for convective outflow. However, this does indicate that the simulation of the UT in the southeast US is biased high.

To investigate the cause of this bias, we compare the WRF lightning flash density to that measured by the Earth Networks Total Lightning Network (ENTLN). ENTLN is a ground-based lightning observation network with more than 900 sensors deployed in the contiguous US. The sensors record lightning-produced stokes as well as accurate time and location. Strokes are then clustered into a flash if they are within 700 milliseconds and 10 kilometers. The detection coefficient is larger than 70% across southern contiguous US (Rudlosky, 2015).

For the comparison, the WRF-Chem simulation is that described in Sect. 2.2. ENTLN and WRF-Chem are sampled from May 13 to Jun 23 2012 over the middle and east US domain, where active lightning events are detected. Both observed and simulated lightning flashes are converted to flash density by dividing flash counts by corresponding grid areas and time range.

Figure 2a and b show the spatial distribution of flash density in number per km^2 per day observed by ENTLN and simulated by WRF-Chem. The largest biases are located over the southeast US (outlined by red on the map). In this region, WRF-Chem substantially overestimates flash density in general and a detection coefficient of 70% for ENTLN cannot account for the discrepancy. The simulated flash density is the highest primarily along the coast, which is not detected by ENTLN.

The scatter plot of daily flash density over the southeast US from two datasets in Fig. 2c demonstrates that the WRF-Chem consistently overestimates flashes in the southeast US over the study period. However, outside of the southeast US, the agreement improves. The simulation captures the spatial pattern over the regional scale (Fig. 2a–b) and the simulated flash densities are consistent with the observed flash densities and the correlation improves as well (Fig. 2d).

Currently, the cause of the discrepancies between the flash density from WRF-Chem simulation and ENTLN observation, is unknown. However, it is clear that it is the flash density, rather than the per-flash production rate of NO, is the cause of the disagreement in the UT between the daily profiles and SEAC4RS data. Further research is required to optimize the lightning parameterizations and improve flash density simulations in the southeast US for our model simulation.

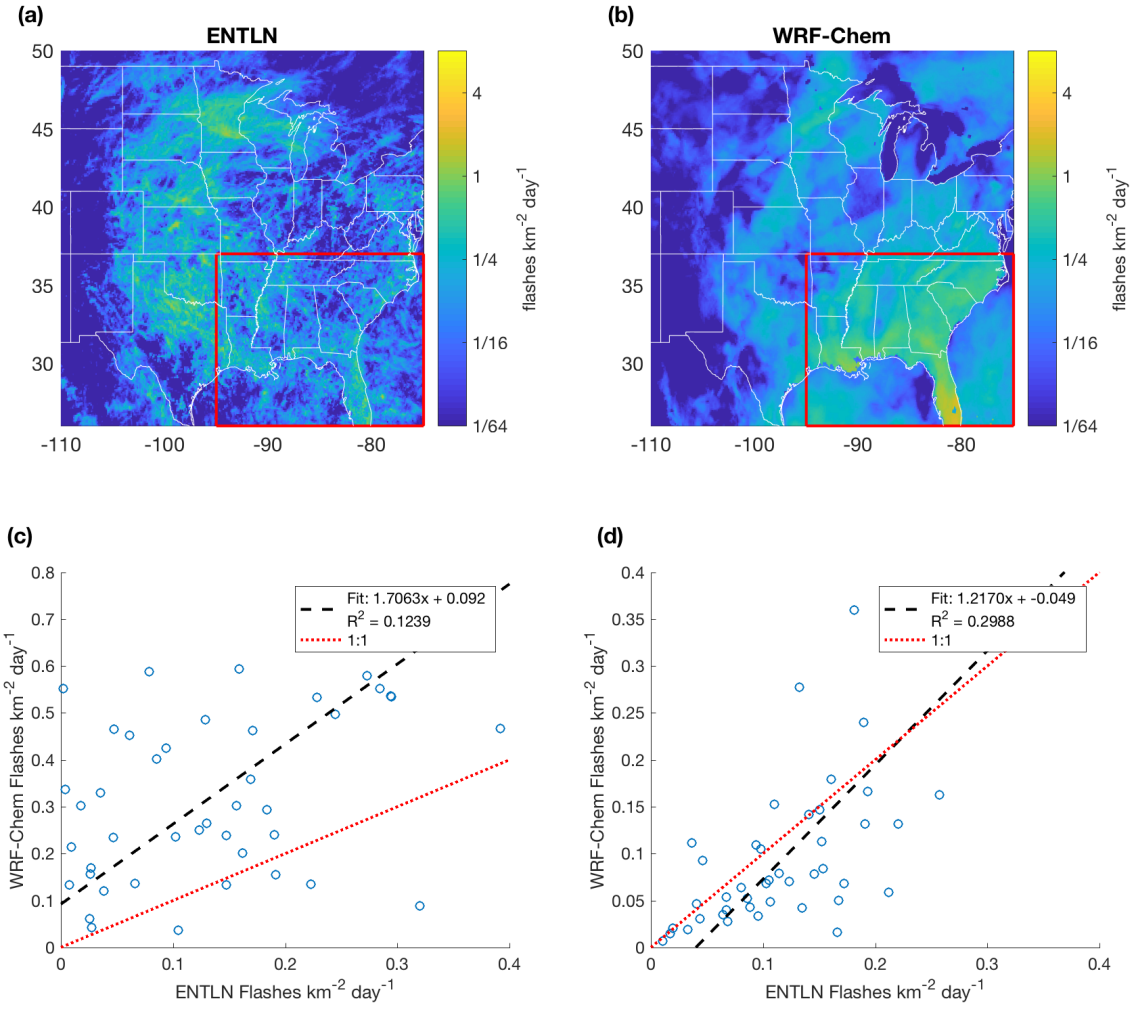

Figure 2. Comparison between observed and simulated flash density from May 13 to June 23 2012. **(a)** and **(b)** show the mean flash density averaged over the study period from ENTLN and WRF-Chem, respectively. Both are gridded at 12 km grid spacing. **(c)** and **(d)** show the correlation between total flash density per day between WRF and ENTLN in **(c)** the southeast US (denoted by the red box in **a** and **b**) and **(d)** elsewhere in CONUS.

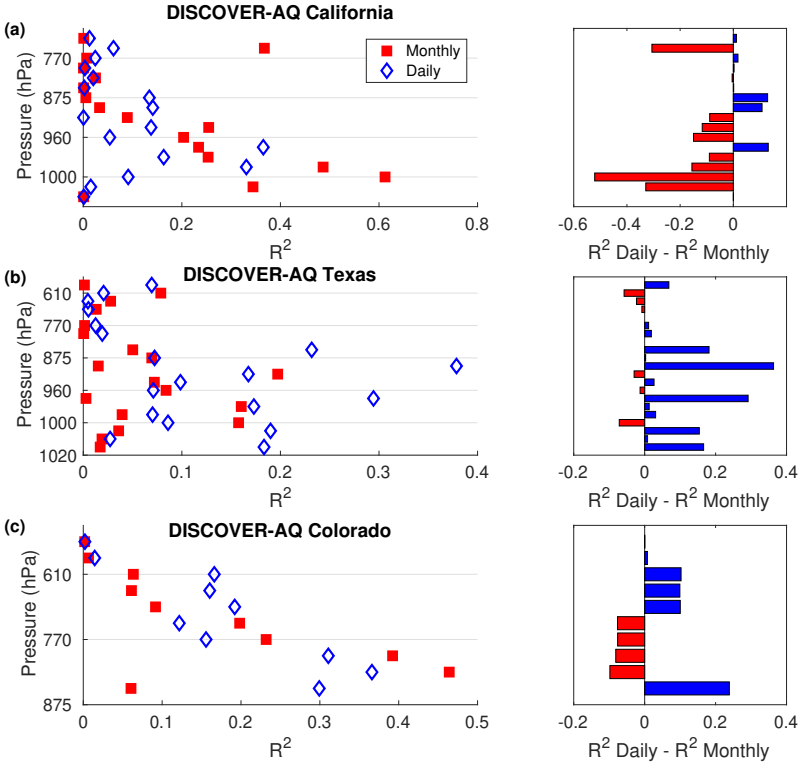

Figure 3. R^2 values for correlation between aircraft data and spatiotemporally matched WRF-Chem data for the **(a)** DISCOVER-CA, **(b)** DISCOVER-TX, and **(c)** DISCOVER-CO campaigns, binned by pressure. Left column: absolute R^2 values for each bin. Right column: the difference in R^2 values using monthly average and daily profiles for each bin.

3.2 Evaluation of variability in daily profiles

As demonstrated in Laughner et al. (2016), simulating the day-to-day variability in the a priori NO_2 profiles can have a significant impact on the retrieved NO_2 VCDs, due primarily to the day-to-day variation in wind speed and direction driving outflow from emissions sources, e.g. cities and power plants. To examine how well WRF-Chem captures the day-to-day variability in NO_2 profiles, we compare aircraft data from three DISCOVER-AQ campaigns and the matched WRF-Chem data (Sect. 2.4). For each profile in the DISCOVER data, we binned the NO_2 concentrations by pressure and calculated the correlation between WRF-Chem and aircraft NO_2 concentrations (one data point per profile per pressure bin). The results are shown in Fig. 3.

In California (Fig. 3a), the monthly average profiles correlate better with the aircraft data. However, as mentioned before, the Californian Central Valley is known to be difficult to model accurately (Hu et al., 2010). In Colorado (Fig. 3c), the daily profiles do a slightly better job overall, getting the variability at the surface and in an elevated layer more accurately than the monthly average profiles. The difference in Texas is quite dramatic (Fig. 3b), with the daily modeled profiles performing substantially

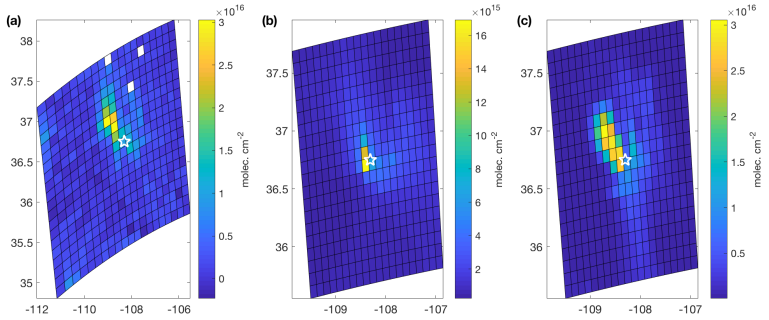

Figure 4. A comparison of OMI SCDs **(a)** and WRF monthly average **(b)** and daily **(c)** VCDs. The star marks the location of the Four Corners power plant. Data is from 4 Mar 2007.

better. This suggests that daily profiles are able to capture variability caused by small, concentrated urban plumes much more effectively than monthly average profiles.

As a second check, we also compare WRF-Chem tropospheric VCDs to OMI SCDs to evaluate the general accuracy of wind direction and speed in the daily model profiles. The OMI SCDs do not depend on modeled vertical profiles, and so constitute an independent check on the plume direction. In order to have strong isolated NO_x sources, we use Atlanta, Chicago, Las Vegas, Los Angeles, New York, and the Four Corners power plant for this study. For each of these sites, 5 days from 2007 are randomly chosen. If insufficient OMI SCDs are available for any day ($> 10\%$ of OMI pixels are cloud covered or in the row anomaly), another day is randomly chosen.

For each day, the agreement between the relative spatial distribution of WRF-Chem VCDs and OMI SCDs is manually evaluated, focusing on whether the model plume is advected in the same direction as the OMI SCDs indicate. Each day's agreement is evaluated qualitatively as good or bad. This, whether the WRF-Chem daily VCDs are significantly different from the monthly average WRF-Chem VCDs, and the confidence in the comparison are recorded for each comparison. Because of the number of factors that affect the absolute magnitude of SCDs, we look for qualitative, rather than statistically quantitative, agreement between the modeled VCDs and OMI SCDs. This is relevant since Laughner et al. (2016) noted that it is primarily the plume shape that drives the day-to-day variability in AMFs, therefore a direct, qualitative evaluation of the plume shape is desirable.

Figures 4 and 5 show two example comparisons, one good (Fig. 4) and one poor (Fig. 5). By studying randomly chosen days for 6 large NO_x sources, we find that about 67–73% of days with sufficient data to be evaluated show good agreement between the OMI SCDs and WRF-Chem daily VCDs. (The range is due to different levels of confidence filtering.) This indicates that the WRF-Chem simulated NO_2 profiles are adequately capturing the day-to-day variability due to wind speed and direction. While we recognize that this conclusion is highly qualitative, the specific character of agreement that is important for these profiles (overall plume size and direction, rather than exact agreement between modeled and real concentrations or column densities)

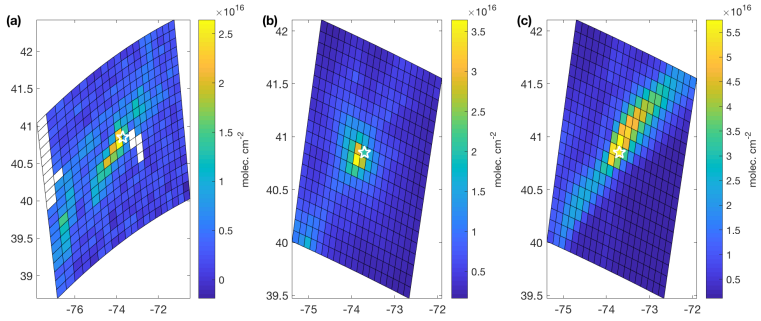

Figure 5. A comparison of OMI SCDs **(a)** and WRF monthly average **(b)** and daily **(c)** VCDs. The star marks the location of New York, NY, USA. Data is from 29 Sept 2007.

is rather difficult to evaluate quantitatively. We recognize that developing such methods is necessary and offer several possible approaches in Sect. 5.

Both comparisons (vs. OMI SCDs and aircraft measurements) show that daily WRF-Chem profiles do, on average, a better job than monthly average profiles capturing the day-to-day variation in profile shape. Therefore, the core improvement in BEHR v3.0, the transition to daily, high-resolution a priori profiles, is fundamentally sound. Daily profiles are especially important for applications that focus on upwind/downwind differences in NO_2 columns around a NO_x source (Laughner et al., 2016) and, as we will see in Sect. 4, generally improve the retrieval in dense urban areas.

4 Column density evaluation

For the DISCOVER campaigns, we compare BEHR against aircraft-derived and Pandora VCDs together, calculating a single regression line for the combined dataset. ~~We~~ These two measurements have unique strengths and weaknesses for comparison against satellite VCDs: Pandoras give a precise column measurement and can be deployed for long time periods, but have a very small footprint (leading to possible representativeness errors) and provide a total, not tropospheric, column. Aircraft profiles have a footprint more similar to an OMI pixel size, but introduce uncertainty due to missing parts of the profile (near the surface and in the upper troposphere in the DISCOVER campaigns) and cannot be deployed for long term, routine observations.

In order to take advantage of each methods' strenghts, we use two comparisons, in one, only Pandora data that has a coincident aircraft profile ~~is include~~ are included ("matched"), in the other, all cloud-free Pandora data is used ("all"). We do so because, when including all Pandora data, the number of Pandora comparisons available will overwhelm the number of available aircraft profiles in the regression. Therefore the regressions using all Pandora data are representative of longer time periods, but weighted strongly towards the Pandora data, and the regressions using only the coincident data represent shorter time periods, but give more weight to the aircraft data. As stated in sect. 2.3, we average all data within 1 hour of OMI overpass (i.e. 13:30 local time ± 1 h) to be consistent with Goldberg et al. (2017). A shorter averaging window (± 0.5 h) was tested;

the maximum effect on the slope was $\sim 8\%$ with most of the "matched" data showing differences of $\leq 5\%$ and the "all" data changing by $\leq 3.5\%$ in all but one case.

Slopes and their 1σ uncertainties for combined aircraft and Pandora VCDs are shown in Table 3. To evaluate the southeast US, we use the SENEX and SEAC4RS campaigns, which only have aircraft data. These results are shown in Table 4. More details (slope, intercepts, R^2 values) can be found in Tables S1, S2, and S3. Figure 6 shows scatter plots of the BEHR vs. aircraft and all Pandora data for the four DISCOVER-AQ campaigns. Scatter plots showing aircraft and Pandora data separately are available in Sect. S1 of the supplement.

For the DISCOVER-CO aircraft comparison, negative VCDs were removed. ~~Such VCDs result from an overestimated stratosphere; since~~ Negative VCDs occur when the estimated stratospheric NO_2 column is greater than the total NO_2 column, thus $V_{trop} = V_{total} - V_{strat} < 0$; they cannot be introduced by the AMF correction of the tropospheric SCD to VCD as the AMF is a multiplicative factor and always > 0. Since all versions of BEHR use the same ~~stratosphere~~ stratospheric NO_2 column as their respective NASA SP products, an error in stratospheric subtraction will be present in all products, and ~~since they~~ it cannot be corrected in the BEHR retrieval~~, do not contribute useful information to the evaluation. To evaluate the southeast US, we use the SENEX and SEAC4RS campaigns, which only have aircraft data~~. Aircraft VCDs, by their nature, cannot be negative, so for these comparisons we remove the negative VCDs so as to avoid increasing the regression slopes by trying to fit these erroneous points. (However, we do note that this is a special case where individual pixels or small groups of pixels are being compared against other VCDs. Most applications of BEHR data should retain the negative VCDs to avoid transforming the essentially Gaussian random stratospheric error into a systematic error by removing part of the bell curve.) Since the stratospheric VCDs are added back to the BEHR or NASA SP tropospheric VCDs for comparison with the Pandora VCDs, negative VCDs are not an issue with Pandora comparisons. ~~These results are shown in Table 4. More details (slope, intercepts, R^2 values) can be found in Tables S1, S2, and S3.~~

In the following sections, we will evaluate the new BEHR v3.0 VCDs from three perspectives: performance compared to the current NASA SP, performance compared to the previous version of BEHR, and performance using daily a priori profiles compared to using monthly a priori profiles. Throughout, BEHR v3.0 (M) refers to BEHR using monthly NO_2 profiles; likewise, BEHR v3.0 (D) refers to the product using daily NO_2 profiles. We will focus on the regression slopes here; intercepts and R^2 values are given in Table S3 in the supplement; however we note that there is not a clear pattern of any one product having a consistently better R^2 value than the others.

4.1 Comparison vs. SP v3.0

For all the DISCOVER campaigns, BEHR v3.0 shows better agreement with both aircraft and Pandora measurements than the NASA SP v3.0 (slopes closer to 1). This is expected, since these campaigns generally centered on one or more cities, and a key feature of the BEHR retrieval are the ~ 12 km a priori profiles ($\sim 10x$ high resolution than the NASA SP v3.0 profiles) which better capture the urban profile shape.

In the SENEX and SEAC4RS campaigns, BEHR's performance is more mixed. These campaigns include the southeast US, where we found that the WRF-Chem simulation that generated the a priori profiles overestimated the lightning flash density

Campaign	Product	Slope (Matched)	Slope (All)
DISCOVER-MD	BEHR v3.0B (D)	N/A	N/A
	BEHR v3.0B (M)	0.80 ± 0.08	0.64 ± 0.03
	BEHR v2.1C	1.3 ± 0.1	0.87 ± 0.05
	SP v3.0	0.79 ± 0.08	0.50 ± 0.03
DISCOVER-CA	BEHR v3.0B (D)	0.49 ± 0.04	0.68 ± 0.04
	BEHR v3.0B (M)	0.51 ± 0.04	0.66 ± 0.04
	BEHR v2.1C	0.57 ± 0.05	0.68 ± 0.04
	SP v3.0	0.41 ± 0.04	0.54 ± 0.03
DISCOVER-TX	BEHR v3.0B (D)	0.69 ± 0.07	1.00 ± 0.06
	BEHR v3.0B (M)	0.60 ± 0.05	0.87 ± 0.05
	BEHR v2.1C	1.1 ± 0.1	1.33 ± 0.08
	SP v3.0	0.53 ± 0.05	0.74 ± 0.05
DISCOVER-CO	BEHR v3.0B (D)	0.66 ± 0.06	0.66 ± 0.03
	BEHR v3.0B (M)	0.70 ± 0.06	0.63 ± 0.03
	BEHR v2.1C	0.74 ± 0.06	0.68 ± 0.03
	SP v3.0	0.53 ± 0.05	0.50 ± 0.02

Table 3. Slopes and 1σ uncertainties of BEHR vs. combined aircraft (extended with GEOS-Chem profiles) and Pandora VCDs. Matched slopes use only Pandora data approximately coincident with aircraft profiles to get similar sampling; all uses all valid Pandora data. Outliers and negative VCDs are removed before computing slopes.

Campaign	Product	Slope (GEOS-Chem)	Slope (Extrap.)
SENEX	BEHR v3.0B (D)	2.3 ± 0.5	1.7 ± 0.5
	BEHR v3.0B (M)	1.0 ± 0.2	0.9 ± 0.3
	BEHR v2.1C	1.4 ± 0.4	1.5 ± 0.5
	SP v3.0	1.1 ± 0.2	0.8 ± 0.3
SEAC4RS	BEHR v3.0B (D)	0.9 ± 0.4	0.7 ± 0.3
	BEHR v3.0B (M)	1.2 ± 0.4	1.0 ± 0.3
	BEHR v2.1C	2.6 ± 0.5	2.5 ± 0.7
	SP v3.0	1.0 ± 0.3	0.8 ± 0.3

Table 4. Slopes and 1σ uncertainties for RMA regression of satellite VCDs against in situ calculated VCDs. Both methods of extending the profiles (using GEOS-Chem modeled profiles or extrapolating the top/bottom ten points) are included. Outliers are removed before calculating these parameters.

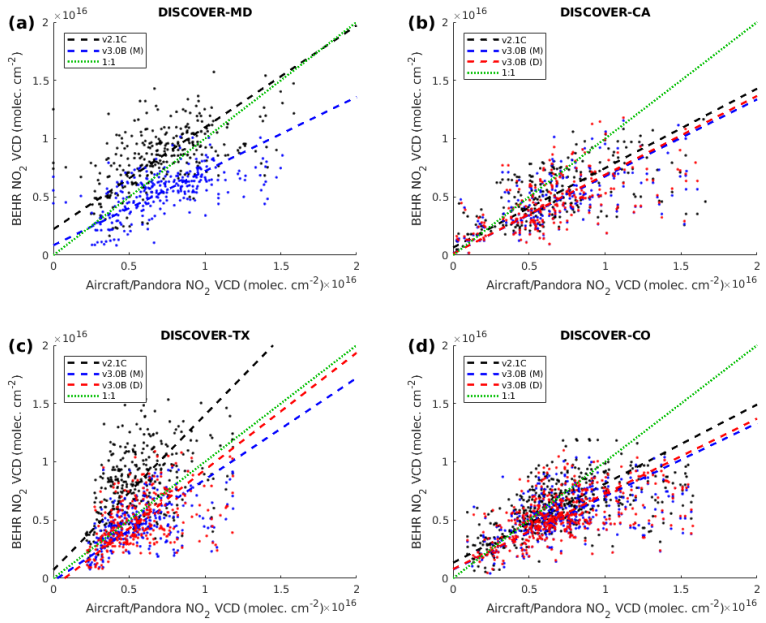

Figure 6. Scatter plots of the BEHR v2.1C, v3.0B (M), and (where available) v3.0B (D) VCDs against coincident aircraft and Pandora VCDs. All Pandora VCDs are used for these plots. Each panel is one campaign: the **(a)** Maryland, **(b)** California, **(c)**, Texas, and **(d)** Colorado DISCOVER-AQ campaigns. For slopes, see Table 3; for intercepts and R^2 values, see Table S3.

(Sect. 3.1). In SEAC4RS, whether BEHR v3.0 (M) performs better or worse than the NASA SP v3.0 depends on the method used to extend the profile (Sect. 2.4). This indicates that uncertainty in the measurement is greater than the difference between these two products. BEHR v3.0 (D) performs poorly in the SENEX campaign; this will be explored in Sect. 4.3. Overall, BEHR v3.0 (M) is not significantly affected by the overestimated lightning flash density in the southeast US, as the monthly average profiles smooth out the overlarge UT lightning NO_2 signal.

4.2 Comparison vs. BEHR v2.1

Using aircraft data plus just Pandora data coincident with aircraft spirals, v2.1 performs better in all DISCOVER campaigns except MD. However, using aircraft data plus all Pandora data, v3.0 (D) performs better than or similar to v2.1 in all DISCOVER where daily profiles are available. The Pandoras provide more observations than the aircraft profiles, and, due to their small footprint, are more sensitive to narrow, highly concentrated NO_2 plumes. The v2.1 profiles used 2005 emissions; as seen in Fig. 1, this led to too much NO_2 being placed at the surface, which will increase the retrieved VCD. This suggests that the better performance of v2.1 in some cases is due to cancellation of errors; overestimated surface NO_2 is canceling out the lack of temporal variation in the profiles. That is, the higher average surface concentration in the v2.1 profiles may be similar to the in-plume concentrations resolved by the daily v3.0 profiles.

In v3.0, when daily profiles are available, the agreement is similar to or better than v2.1 if aircraft and all Pandora data is used. Therefore, daily profiles are able to capture at least some enhancements in surface NO_2 where and when they occur, without overestimating the average profile. This is not evident using aircraft and just the coincident Pandora data because of the smaller number of comparisons. As the comparison expands (using all Pandora data), the improvement becomes evident. The better performance of daily profiles suggests that even though Laughner et al. (2016) did not see large effects in a multi-month average using daily instead of monthly profiles, that daily profiles will provide a more accurate representation of urban VCDs over longer averaging periods.

BEHR v3.0 performs better in the SENEX and SEAC4RS comparisons than v2.1 (excluding 3.0 (D) in SENEX). The v2.1 profiles did not include lightning emissions, as it was a limitation of WRF-Chem at that time (Laughner et al., 2018f). This indicates that, even though the contribution of lightning to the southeast US profiles is too large, the inclusion of lightning NO_2 in the profiles did improve the representation of the southeast US. Laughner et al. (2018f) also showed that implementing a variable tropopause pressure decreased VCDs in the southeast US during summer; this also would help reduce the high bias compared to SENEX and SEAC4RS seen in BEHR v2.1.

4.3 Comparison of BEHR v3.0 (M) vs BEHR v3.0 (D) in the SE US

In the SENEX campaign, v3.0 (D) performs significantly worse than v3.0 (M). From Fig. 1 we know that the daily a priori profiles overestimate the UT NO_2, and from Fig. 2 we know that this is due to a significant overestimate of the flash density in our WRF simulation. The comparison in Table 4 would seem to indicate that this overestimate has a severe impact on the retrieved VCDs, but we must also consider the uncertainty in the SENEX-derived VCDs.

Figure 7a shows the ensemble of profiles from SENEX used to calculate VCDs. The circles mark levels that had to be calculated using model data for $> 50\%$ of the profiles. In SENEX, that is all levels about ~ 700 hPa, which means that the SENEX aircraft data provides very little constraint on the UT. The lightning contribution to the SENEX columns must come from the GEOS-Chem monthly averages or extrapolation from a lower altitude, which means the spatial and temporal variation is lost.

Figure 7c shows the effect of using the WRF-Chem a priori profiles instead of the GEOS-Chem profiles to extend the SENEX profiles. The WRF-Chem profiles do include spatial and temporal variation of the UT, but using them reinforces the AMF errors, moving all points away from the 1:1 line. Without either in situ measurements of the UT in the southeast US or Pandora total column observations we cannot separate the errors in AMF caused by the overestimated UT NO_2 in the a priori profiles from the error caused by the lack of spatiotemporal variation in the extended aircraft profiles. For example, the error in the cluster of points below the 1:1 line in Fig. 7c could be corrected if either the UT NO_2 in the a priori profile was reduced, decreasing the AMFs and so increasing the BEHR VCDs, or if the aircraft profile had less NO_2, thus moving the points left onto the 1:1 line. (In the case, there would still be a discrepancy between the BEHR VCD and the VCD derived from combining aircraft and WRF-Chem profiles, suggesting that the WRF-Chem UT NO_2 is still too great.)

Other campaigns do have better sampling of the UT, e.g. SEAC4RS (Fig. 7b,d,f), but do not have as many profiles in the southeast US (Fig. 7e,f). Therefore, we must currently assign an uncertainty of $\pm 100\%$ to VCDs retrieved with daily profiles in

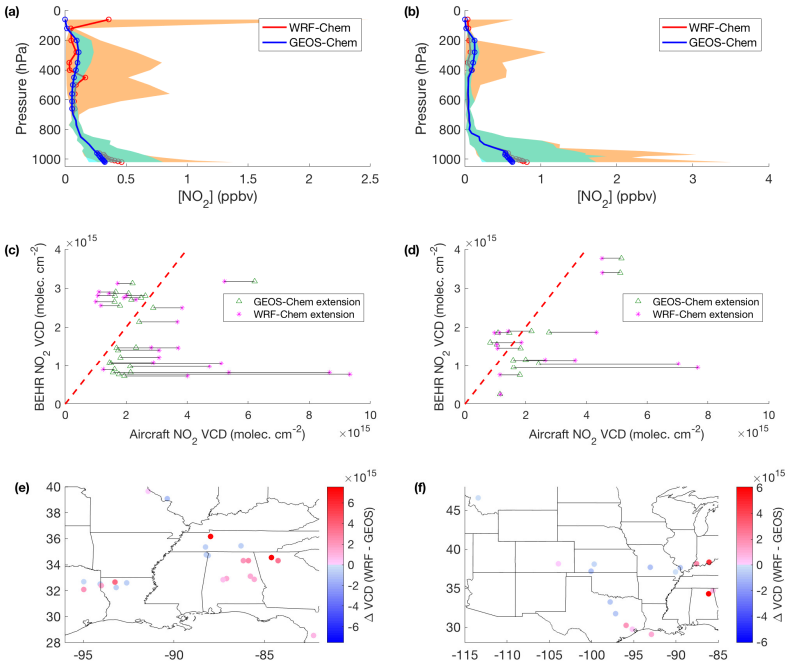

Figure 7. (a,b) The profiles used to calculate the aircraft VCDs extended using WRF-Chem or GEOS-Chem profiles; the solid line is the median of all profiles, the shading represents the 10th and 90th percentiles for each binned level. Circles indicate levels that were derived from the models in at least 50% of the profiles. **(c,d)** Comparison of BEHR v3.0 (D) VCDs vs. aircraft-derived VCDs using GEOS-Chem and WRF-Chem profiles to extend the profile to the surface and tropopause. The black lines connect corresponding comparisons between the two methods and the red dashed line represents the 1:1 agreement. **(e,f)** Difference between aircraft VCDs extended with WRF-Chem and GEOS-Chem profiles. **(a,c,e)** are for the SENEX campaign, **(b,d,f)** are for SEAC4RS.

the southeast US (east of 95° W and south of 37.5° N). This is almost certainly overly conservative, as Laughner et al. (2018f) showed that the frequency distribution of UT NO_2 in the southeast apriori profiles was skewed to lower values in the daily profiles, and a three month average using daily a priori profiles resulted in greater VCDs than using monthly a priori profiles, which would not be the case if the daily profiles always overestimated the UT. This suggests that days with little or no lightning in both the real world and WRF-Chem simulations are more numerous than days with significant lightning contribution, and so a multi-month average using daily profiles would in fact accurately capture this. However, without long term independent column measurements in the southeast, we cannot confirm this hypothesis. Future work will focus on improving the simulation of lightning in the southeast US. If successful, improved WRF-Chem profiles for the southeast can be implemented.

4.4 Comparison of BEHR v3.0 (M) vs BEHR v3.0 (D) in urban areas

In the DISCOVER campaigns, BEHR v3.0 (D) using daily profiles has regression slopes similar to or closer to 1 than BEHR v3.0 (M) using monthly profiles except in the DISCOVER-CA aircraft comparisons. There is a clear improvement in DISCOVER-TX using daily profiles. This suggests that the daily profiles are capturing small, concentrated plumes in the urban area (Fig. 3b), which is improving the retrieval overall in an urban area with many highly concentration industrial NO_x sources. Therefore, we argue that daily profiles improve the retrieval in many ways, not only for applications that select for upwind/downwind pixels as shown in Laughner et al. (2016), but also for multi-month averages in dense urban areas.

5 Discussion: future efforts to validate daily profiles

Using space-based SCDs to evaluate the spatial distribution of NO_2 in a CTM is powerful (Sect. 3.2), because both provide a spatially continuous field of NO_2 columns. As we have shown here, this makes a qualitative evaluation straightforward and illustrative. However, a quantitative metric is more challenging to devise, as the direct correlation of model and satellite columns is less important than the more abstract agreement between the overall plume direction and extent. As we have shown here, daily, high-resolution profiles provide important benefits to an NO_2 retrieval; therefore, development of more quantitative methods to evaluate model performance in this manner should be a priority.

There are several possibilities. First, an algorithm that identifies the plume and computes the direction and length of its major axis could be used. This would allow a comparison of the direction and extent of the plumes more directly. Such an algorithm would not be trivial to develop; comparisons such as the one shown in Fig. 5a,c would likely be difficult for the algorithm to distinguish the plume direction accurately.

Second, this problem could be treated as an image recognition problem. A neural network could be trained on modeled VCDs and SCDs. A training set of good and bad days could be constructed from the WRF-Chem simulations used in BEHR v3.0D. Development of this approach is beyond the scope of this paper.

Third, dense sensor networks (e.g. Shusterman et al., 2016; Kim et al., 2018) may also be useful to evalulate daily profiles by permitting a simpler correlation test between modeled and observed surface concentrations than is possible between modeled VCDs and observed SCDs. Development of these networks is a topic of active research. This method may be necessary for future retrievals, especially over the US and European domains, where decreasing NO_x emissions mean that the contrast between plumes and background in SCDs is much weaker now than in 2007.

6 Conclusions

We have evaluated version 3.0B of the BEHR OMI NO_2 product against multiple datasets. We find that the WRF simulation used to generate the a priori NO_2 profiles generally agrees well with the available aircraft data; however, the number of lightning flashes is significantly overestimated in the southeast US leading to an overestimate of the UT NO_2 in that region, although broadly consistent with ENTLN observations elsewhere. When compared against aircraft-derived and Pandora VCDs,

BEHR v3.0B performs better than SP v3.0, with regionally varying low biases of 0–51% compared to in situ and Pandora measurements. Using daily profiles yields better results than monthly profiles, except in the southeast US.

The lessons learned here are applicable to geostationary satellites scheduled to launch in the near future. Because the BEHR retrieval focuses on the continental United States, it serves as a useful prototype for future NO_2 retrievals from geostationary satellites such as GEMS (Bak et al., 2013; Choi and Ho, 2015), ~~Sentinel-5~~ Sentinel-4 (Ingmann et al., 2012), and TEMPO (Chance et al., 2013), which also will be inherently restricted to regional areas. This offers the opportunity to use higher-resolution a priori data than global retrievals.

Here, the results from the SENEX and SEAC4RS campaign here demonstrate that verifying the chemical transport model's reproduction of the day-to-day variability in lightning flashes is vital to obtain reliable results in such regions. With the sub-daily temporal resolution available to geostationary satellites, this will only become more important. Therefore, geostationary retrievals should evaluate the diurnal variation in lightning flashes in their a priori models using ground- and space- based lightning detectors (e.g. NLDN, ENTLN, or the GOES-R lightning mapper), and plans should be made to validate retrieved VCDs in multiple regions that have strong, but different, lightning influence. Such validations must include measurement of the UT NO_2 profile and/or total column observations in order to reliably separate errors in the a priori profiles from errors in the observations used for evaluation.

Evaluating the day-to-day performance of the a priori profiles in future geostationary retrievals is crucial. Daily profiles have been shown to significantly affect retrieved NO_2, especially in applications that systematically focus on NO_2 VCDs downwind of a source (Laughner et al., 2016), and we have shown here that daily profiles also improve performance in urban areas. With the WRF-Chem model configuration used here, urban NO_2 plumes are simulated with the correct spatial pattern $\sim 70\%$ of the time. Planned campaigns to evaluate geostationary satellite retrievals should be designed with an eye towards also evaluating the day-to-day accuracy of the a priori profiles.

Code and data availability. The analysis code for this paper is available at https://github.com/behr-github/BEHR-v3-evaluation/ (Laughner, 2018). Supporting datasets generated or used by this code is hosted by UC Dash (Laughner et al., 2018e). The BEHR v3.0B product is hosted as four subproducts by UC Dash (Laughner et al., 2018a, b, c, d) as well as on behr.cchem.berkeley.edu. The BEHR algorithm is available at https://github.com/CohenBerkeleyLab/BEHR-core/tree/master (Laughner and Zhu, 2018).

Competing interests. The authors declare no competing interests.

Acknowledgements. The authors gratefully acknowledge support from the NASA ESS Fellowship NNX14AK89H, NASA grant NNX15AE37G, and the TEMPO project grant SV3-83019.

We would like to acknowledge high-performance computing support from Cheyenne (doi:10.5065/D6RX99HX) provided by NCAR's Computational and Information Systems Laboratory, sponsored by the National Science Foundation. This research also used the Savio

computational cluster resource provided by the Berkeley Research Computing program at the University of California, Berkeley (supported by the UC Berkeley Chancellor, Vice Chancellor for Research, and Chief Information Officer).

We acknowledge use of the WRF-Chem preprocessor tools MOZBC, fire_emiss, etc. provided by the Atmospheric Chemistry Observations and Modeling (ACOM) laboratory of NCAR.

Finally, we acknowledge the use of in situ NO_2 measurements from the UC Berkeley TD-LIF instrument and the NOAA chemiluminescence instrument, as well as ground based NO_2 columns from multiple Pandora spectrometers operated by various groups throughout the DISCOVER-AQ campaigns.

References

Bak, J., Kim, J. H., Liu, X., Chance, K., and Kim, J.: Evaluation of ozone profile and tropospheric ozone retrievals from GEMS and OMI spectra, Atmos. Meas. Tech., 6, 239–249, https://doi.org/10.5194/amt-6-239-2013, 2013.

Barth, M. C., Cantrell, C. A., Brune, W. H., Rutledge, S. A., Crawford, J. H., Huntrieser, H., Carey, L. D., MacGorman, D., Weisman, M., Pickering, K. E., Bruning, E., Anderson, B., Apel, E., Biggerstaff, M., Campos, T., Campuzano-Jost, P., Cohen, R., Crounse, J., Day, D. A., Diskin, G., Flocke, F., Fried, A., Garland, C., Heikes, B., Honomichl, S., Hornbrook, R., Huey, L. G., Jimenez, J. L., Lang, T., Lichtenstern, M., Mikoviny, T., Nault, B., O'Sullivan, D., Pan, L. L., Peischl, J., Pollack, I., Richter, D., Riemer, D., Ryerson, T., Schlager, H., Clair, J. S., Walega, J., Weibring, P., Weinheimer, A., Wennberg, P., Wisthaler, A., Wooldridge, P. J., and Ziegler, C.: The Deep Convective Clouds and Chemistry (DC3) Field Campaign, Bull. Am. Met. Soc., 96, 1281–1309, https://doi.org/10.1175/bams-d-13-00290.1, 2015.

Beirle, S., Platt, U., Wenig, M., and Wagner, T.: NOx production by lightning estimated with GOME, Adv. Space Res., 34, 793 – 797, https://doi.org/https://doi.org/10.1016/j.asr.2003.07.069, http://www.sciencedirect.com/science/article/pii/S0273117704003576, trace Constituents in the Troposphere and Lower Stratosphere, 2004.

Beirle, S., Huntrieser, H., and Wagner, T.: Direct satellite observations of lightning-produced NO_x, Atmos. Chem. Phys., 10, 10 965–10 986, https://doi.org/10.5194/acp-10-10965-2010, 2010.

Beirle, S., Boersma, K., Platt, U., Lawrence, M., and Wagner, T.: "Megacity Emissions and Lifetimes of Nitrogen Oxides Probed from Space", Science, 333, 1737–1739, 2011.

Bertram, T. H., Heckel, A., Richter, A., Burrows, J. P., and Cohen, R. C.: Satellite measurements of daily variations in soil NO_x emissions, Geophys. Res. Lett., 32, https://doi.org/10.1029/2005gl024640, https://doi.org/10.1029/2005gl024640, 2005.

Bey, I., Jacob, D. J., Yantosca, R. M., Logan, J. A., Field, B. D., Fiore, A. M., Li, Q., Liu, H. Y., Mickley, L. J., and Schultz., M. G.: Global modeling of tropospheric chemistry with assimilated meteorology, J. Geophys. Res., 106, 23 073–23 096, 2001.

Boersma, F., Bucsela, E., Brinksma, E., and Gleason, J. F.: NO_2 *in: OMI Algorithm Theoretical Basis Document, Volume IV, OMI Trace Gas Algorithms* (K. Chance, ed.), Tech. rep., Smithsonian Astrophysical Observatory, 2001.

Boersma, K., Eskes, H., Veefkind, J., Brinksma, E., van der A, R., Sneep, M., van den Oord, G., Levelt, P., Stammes, P., Gleason, J., and Bucsela, E.: Near-real time retrieval of tropospheric NO_2 from OMI, Atmos. Chem. Phys., 7, 2103–2118, 2007.

Bousserez, N.: Space-based retrieval of NO_2 over biomass burning regions: quantifying and reducing uncertainties, Atmos. Meas. Tech., 7, 3431–3444, https://doi.org/10.5194/amt-7-3431-2014, https://doi.org/10.5194/amt-7-3431-2014, 2014.

Browne, E. C., Perring, A. E., Wooldridge, P. J., Apel, E., Hall, S. R., Huey, L. G., Mao, J., Spencer, K. M., Clair, J. M. S., Weinheimer, A. J., Wisthaler, A., and Cohen, R. C.: Global and regional effects of the photochemistry of $CH_3O_2NO_2$: evidence from ARCTAS, Atmos. Chem. Phys., 11, 4209–4219, https://doi.org/10.5194/acp-11-4209-2011, 2011.

Bucsela, E., Krotkov, N., Celarier, E., Lamsal, L., Swartz, W., Bhartia, P., Boersma, K., Veefkind, J., Gleason, J., and Pickering, K.: "A new tropospheric and stratospheric NO_2 retrieval algorithm for nadir-viewing satellite instruments: applications to OMI, Atmos. Meas. Tech., 6, 2607–2626, https://doi.org/10.5194/amt-6-2607-2013, 2013.

Bucsela, E. J., Perring, A. E., Cohen, R. C., Boersma, K. F., Celarier, E. A., Gleason, J. F., Wenig, M. O., Bertram, T. H., Wooldridge, P. J., Dirksen, R., and Veefkind, J. P.: Comparison of tropospheric NO2 from in situ aircraft measurements with near-real-time and standard product data from OMI, Journal of Geophysical Research: Atmospheres, 113, D16S31, https://doi.org/10.1029/2007JD008838, http://dx.doi.org/10.1029/2007JD008838, 2008.

Bucsela, E. J., Pickering, K. E., Huntemann, T. L., Cohen, R. C., Perring, A., Gleason, J. F., Blakeslee, R. J., Albrecht, R. I., Holzworth, R., Cipriani, J. P., Vargas-Navarro, D., Mora-Segura, I., Pacheco-Hernández, A., and Laporte-Molina, S.: Lightning-generated NOx seen by the Ozone Monitoring Instrument during NASA's Tropical Composition, Cloud and Climate Coupling Experiment (TC4), Journal of Geophysical Research: Atmospheres, 115, n/a–n/a, https://doi.org/10.1029/2009JD013118, http://dx.doi.org/10.1029/2009JD013118, d00J10, 2010.

Burrows, J. P., Weber, M., Buchwitz, M., Rozanov, V., Ladstätter-Weißenmayer, A., Richter, A., DeBeek, R., Hoogen, R., Bramstedt, K., Eichmann, K.-U., and Eisinger, M.: The Global Ozone Monitoring Experiment (GOME): Mission Concept and First Scientific Results, J. Atmos. Sci., 56, 151–175, https://doi.org/10.1175/1520-0469(1999)056<0151:TGOMEG>2.0.CO;2, 1999.

Castellanos, P. and Boersma, K. F.: Reductions in nitrogen oxides over Europe driven by environmental policy and economic recession, Sci. Rep., 2, https://doi.org/10.1038/srep00265, https://doi.org/10.1038/srep00265, 2012.

Castellanos, P., Boersma, K. F., Torres, O., and de Haan, J. F.: OMI tropospheric NO_2 air mass factors over South America: effects of biomass burning aerosols, Atmospheric Measurement Techniques, 8, 3831–3849, https://doi.org/10.5194/amt-8-3831-2015, https://www.atmos-meas-tech.net/8/3831/2015/, 2015.

Chance, K., Liu, X., Suleiman, R. M., Flittner, D. E., Al-Saadi, J., and Janz, S. J.: Tropospheric emissions: monitoring of pollution (TEMPO), https://doi.org/10.1117/12.2024479, 2013.

Choi, Y.-S. and Ho, C.-H.: Earth and environmental remote sensing community in South Korea: A review, Remote Sensing Applications: Society and Environment, 2, 66–76, https://doi.org/10.1016/j.rsase.2015.11.003, 2015.

Computational and Information Systems Laboratory: Cheyenne: HPE/SGI ICE XA System (NCAR Community Computing). Boulder, CO: National Center for Atmospheric Research, https://doi.org/10.5065/D6RX99HX, 2017.

Day, D. A., Wooldridge, P. J., Dillon, M. B., Thornton, J. A., and Cohen, R. C.: A thermal dissociation laser-induced fluorescence instrument for in situ detection of NO2, peroxy nitrates, alkyl nitrates, and HNO3, J. Geophys. Res. Atmos., 107, ACH 4–1–ACH 4–14, https://doi.org/10.1029/2001JD000779, http://dx.doi.org/10.1029/2001JD000779, 2002.

EPA: Air Pollutant Emissions Trends Data, https://www.epa.gov/air-emissions-inventories/air-pollutant-emissions-trends-data, 2016.

Goldberg, D. L., Lamsal, L. N., Loughner, C. P., Swartz, W. H., Lu, Z., and Streets, D. G.: A high-resolution and observationally constrained OMI NO_2 satellite retrieval, Atmos. Chem. Phys., 17, 11 403–11 421, https://doi.org/10.5194/acp-17-11403-2017, https://doi.org/10.5194/acp-17-11403-2017, 2017.

Grell, G. A., Peckham, S. E., Schmitz, R., McKeen, S. A., Frost, G., Skamarock, W. C., and Eder, B.: Fully coupled "online" chemistry within the {WRF} model, Atmos. Environ., 39, 6957 – 6975, https://doi.org/10.1016/j.atmosenv.2005.04.027, 2005.

Guenther, A., Karl, T., Harley, P., Wiedinmyer, C., Palmer, P. I., and Geron, C.: Estimates of global terrestrial isoprene emissions using MEGAN (Model of Emissions of Gases and Aerosols from Nature), Atmos. Chem. Phys., 6, 3181–3210, https://doi.org/10.5194/acp-6-3181-2006, http://www.atmos-chem-phys.net/6/3181/2006/, 2006.

Hains, J. C., Boersma, K. F., Kroon, M., Dirksen, R. J., Cohen, R. C., Perring, A. E., Bucsela, E., Volten, H., Swart, D. P. J., Richter, A., Wittrock, F., Schoenhardt, A., Wagner, T., Ibrahim, O. W., van Roozendael, M., Pinardi, G., Gleason, J. F., Veefkind, J. P., and Levelt, P.: Testing and improving OMI DOMINO tropospheric NO2using observations from the DANDELIONS and INTEX-B validation campaigns, J. Geophys. Res., 115, https://doi.org/10.1029/2009jd012399, https://doi.org/10.1029/2009jd012399, 2010.

Herman, J., Cede, A., Spinei, E., Mount, G., Tzortziou, M., and Abuhassan, N.: NO_2 column amounts from ground-based Pandora and MFDOAS spectrometers using the direct-sun DOAS technique: Intercomparisons and application to OMI validation, J. Geophys. Res. Atmos., 114, https://doi.org/10.1029/2009jd011848, https://doi.org/10.1029/2009jd011848, 2009.

Hönninger, G., von Friedeburg, C., and Platt, U.: Multi axis differential optical absorption spectroscopy (MAX-DOAS), Atmos. Chem. Phys., 4, 231–254, https://doi.org/10.5194/acp-4-231-2004, https://www.atmos-chem-phys.net/4/231/2004/, 2004.

Hu, J., Ying, Q., Chen, J., Mahmud, A., Zhao, Z., Chen, S.-H., and Kleeman, M. J.: Particulate air quality model predictions using prognostic vs. diagnostic meteorology in central California, Atmos. Environ., 44, 215 – 226, https://doi.org/https://doi.org/10.1016/j.atmosenv.2009.10.011, http://www.sciencedirect.com/science/article/pii/S1352231009008590, 2010.

Hudman, R. C., Jacob, D. J., Turquety, S., Leibensperger, E. M., Murray, L. T., Wu, S., Gilliland, A. B., Avery, M., Bertram, T. H., Brune, W., Cohen, R. C., Dibb, J. E., Flocke, F. M., Fried, A., Holloway, J., Neuman, J. A., Orville, R., Perring, A., Ren, X., Sachse, G. W., Singh, H. B., Swanson, A., and Wooldridge, P. J.: Surface and lightning sources of nitrogen oxides over the United States: Magnitudes, chemical evolution, and outflow, Journal of Geophysical Research, 112, https://doi.org/10.1029/2006jd007912, https://doi.org/10.1029/2006jd007912, 2007.

Hudman, R. C., Russell, A. R., Valin, L. C., and Cohen, R. C.: Interannual variability in soil nitric oxide emissions over the United States as viewed from space, Atmos. Chem. Phys., 10, 9943–9952, https://doi.org/10.5194/acp-10-9943-2010, https://www.atmos-chem-phys.net/10/9943/2010/, 2010.

Hudman, R. C., Moore, N. E., Mebust, A. K., Martin, R. V., Russell, A. R., Valin, L. C., and Cohen, R. C.: Steps towards a mechanistic model of global soil nitric oxide emissions: implementation and space based-constraints, Atmos. Chem. Phys., 12, 7779–7795, https://doi.org/10.5194/acp-12-7779-2012, https://www.atmos-chem-phys.net/12/7779/2012/, 2012.

Huijnen, V., Flemming, J., Kaiser, J. W., Inness, A., Leitão, J., Heil, A., Eskes, H. J., Schultz, M. G., Benedetti, A., Hadji-Lazaro, J., Dufour, G., and Eremenko, M.: Hindcast experiments of tropospheric composition during the summer 2010 fires over western Russia, Atmos. Chem. Phys., 12, 4341–4364, https://doi.org/10.5194/acp-12-4341-2012, https://doi.org/10.5194/acp-12-4341-2012, 2012.

Ialongo, I., Herman, J., Krotkov, N., Lamsal, L., Boersma, K. F., Hovila, J., and Tamminen, J.: Comparison of OMI NO_2 observations and their seasonal and weekly cycles with ground-based measurements in Helsinki, Atmos. Meas. Tech., 9, 5203–5212, https://doi.org/10.5194/amt-9-5203-2016, https://doi.org/10.5194/amt-9-5203-2016, 2016.

Ingmann, P., Veihelmann, B., Langen, J., Lamarre, D., Stark, H., and Courrèges-Lacoste, G. B.: Requirements for the {GMES} Atmosphere Service and ESA's implementation concept: Sentinels-4/-5 and -5p, Remote Sens. of Environ., 120, 58–69, https://doi.org/10.1016/j.rse.2012.01.023, 2012.

Kim, J., Shusterman, A. A., Lieschke, K. J., Newman, C., and Cohen, R. C.: The BErkeley Atmospheric CO_2 Observation Network: field calibration and evaluation of low-cost air quality sensors, Atmos. Meas. Tech., 11, 1937–1946, https://doi.org/10.5194/amt-11-1937-2018, https://www.atmos-meas-tech.net/11/1937/2018/, 2018.

Kim, S.-W., Heckel, A., McKeen, S. A., Frost, G. J., Hsie, E.-Y., Trainer, M. K., Richter, A., Burrows, J. P., Peckham, S. E., and Grell, G. A.: Satellite-observed U.S. power plant NO_x emission reductions and their impact on air quality, Geophys. Res. Lett., 33, https://doi.org/10.1029/2006gl027749, https://doi.org/10.1029/2006gl027749, 2006.

Kim, S.-W., Heckel, A., Frost, G. J., Richter, A., Gleason, J., Burrows, J. P., McKeen, S., Hsie, E.-Y., Granier, C., and Trainer, M.: NO_2 columns in the western United States observed from space and simulated by a regional chemistry model and their implications for NO_x emissions, J. Geophys. Res. Atmos., 114, https://doi.org/10.1029/2008jd011343, https://doi.org/10.1029/2008jd011343, 2009.

Konovalov, I. B., Beekmann, M., Richter, A., Burrows, J. P., and Hilboll, A.: Multi-annual changes of NO_x emissions in megacity regions: nonlinear trend analysis of satellite measurement based estimates, Atmos. Chem. Phys., 10, 8481–8498, https://doi.org/10.5194/acp-10-8481-2010, https://www.atmos-chem-phys.net/10/8481/2010/, 2010.

Krotkov, N. A., Lamsal, L. N., Celarier, E. A., Swartz, W. H., Marchenko, S. V., Bucsela, E. J., Chan, K. L., Wenig, M., and Zara, M.: The version 3 OMI NO_2 standard product, Atmos. Meas. Tech., 10, 3133–3149, https://doi.org/10.5194/amt-10-3133-2017, https://www.atmos-meas-tech.net/10/3133/2017/, 2017.

Kurosu, T. P. and Celarier, E. A.: OMI/Aura Global Ground Pixel Corners 1-Orbit L2 Swath 13x24km V003, Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC), https://doi.org/10.5067/Aura/OMI/DATA2020, 2010.

Lamsal, L. N., Krotkov, N. A., Celarier, E. A., Swartz, W. H., Pickering, K. E., Bucsela, E. J., Gleason, J. F., Martin, R. V., Philip, S., Irie, H., Cede, A., Herman, J., Weinheimer, A., Szykman, J. J., and Knepp, T. N.: Evaluation of OMI operational standard NO_2 column retrievals using in situ and surface-based NO_2 observations, Atmos. Chem. Phys., 14, 11 587–11 609, https://doi.org/10.5194/acp-14-11587-2014, https://doi.org/10.5194/acp-14-11587-2014, 2014.

Laughner, J., Zhu, Q., and Cohen, R.: Berkeley High Resolution (BEHR) OMI NO2 - Gridded pixels, daily profiles, v3, UC Berkeley Dash, Dataset, https://doi.org/10.6078/D12D5X, 2018a.

Laughner, J., Zhu, Q., and Cohen, R.: Berkeley High Resolution (BEHR) OMI NO2 - Native pixels, daily profiles, UC Berkeley Dash, Dataset, https://doi.org/10.6078/D1WH41, 2018b.

Laughner, J., Zhu, Q., and Cohen, R.: Berkeley High Resolution (BEHR) OMI NO2 - Gridded pixels, monthly profiles, UC Berkeley Dash, Dataset, https://doi.org/10.6078/D1RQ3G, 2018c.

Laughner, J., Zhu, Q., and Cohen, R.: Berkeley High Resolution (BEHR) OMI NO2 - Native pixels, monthly profiles, UC Berkeley Dash, Dataset, https://doi.org/10.6078/D1N086, 2018d.

Laughner, J. L.: Analysis code for Laughner et al., AMT (revised), 2018, https://doi.org/10.5281/zenodo.1479985, 2018.

Laughner, J. L. and Cohen, R. C.: Quantification of the effect of modeled lightning NO_2 on UV–visible air mass factors, Atmos. Meas. Tech., 10, 4403–4419, https://doi.org/10.5194/amt-10-4403-2017, https://www.atmos-meas-tech.net/10/4403/2017/, 2017.

Laughner, J. L. and Zhu, Q.: CohenBerkeleyLab/BEHR-Core: BEHR Core code, https://doi.org/10.5281/zenodo.998275, 2018.

Laughner, J. L., Zare, A., and Cohen, R. C.: Effects of daily meteorology on the interpretation of space-based remote sensing of NO_2, Atmospheric Chemistry and Physics, 16, 15 247–15 264, https://doi.org/10.5194/acp-16-15247-2016, http://www.atmos-chem-phys.net/16/15247/2016/, 2016.

Laughner, J. L., Zhu, Q., and Cohen, R. C.: Supporting data for "Evaluation of version 3.0B of the BEHR OMI NO2 product", https://doi.org/10.6078/D1JT28, 2018e.

Laughner, J. L., Zhu, Q., and Cohen, R. C.: The Berkeley High Resolution Tropospheric NO_2 Product, Earth System Science Data Discussions, 2018, 1–33, https://doi.org/10.5194/essd-2018-66, https://www.earth-syst-sci-data-discuss.net/essd-2018-66/, 2018f.

Lay, E. H., Jacobson, A. R., Holzworth, R. H., Rodger, C. J., and Dowden, R. L.: Local time variation in land/ocean lightning flash density as measured by the World Wide Lightning Location Network, J. Geophys. Res. Atmos., 112, https://doi.org/10.1029/2006JD007944, 2007.

Liu, F., Beirle, S., Zhang, Q., Dörner, S., He, K., and Wagner, T.: NOx lifetimes and emissions of cities and power plants in polluted background estimated by satellite observations, Atmos. Chem. Phys., 16, 5283–5298, https://doi.org/10.5194/acp-16-5283-2016, 2016.

Liu, F., Beirle, S., Zhang, Q., van der A, R. J., Zheng, B., Tong, D., and He, K.: NO_x emission trends over Chinese cities estimated from OMI observations during 2005 to 2015, Atmos. Chem. Phys. Discuss., pp. 1–21, https://doi.org/10.5194/acp-2017-369, 2017.

Liu, Y., Bourgeois, A., Warner, T., Swerdlin, S., and Hacker, J.: Implementation of the observation-nudging based on FDDA into WRF for supporting AFEC test operations. 6th WRF Conference, NCAR, Boulder, CO, USA, 2006.

Lu, Z., Streets, D., de Foy, B., Lamsal, L., Duncan, B., and Xing, J.: "Emissions of nitrogen oxides from US urban areas: estimation from Ozone Monitoring Instrument retrievals for 2005–2014", Atmos. Chem. Phys., 15, 10 367–10 383, https://doi.org/10.5194/acp-15-10367-2015, 2015.

Martin, R., Sauvage, B., Folkins, I., Sioris, C., Boone, C., Bernath, P., and Ziemke, J.: Space-based constraints on the production of nitric oxide by lightning, J. Geophys. Res. Atmos., 112, https://doi.org/10.1029/2006JD007831, 2007.

McKenzie, R., Johnstone, P., McElroy, C., Kerr, J., and Solomon, S.: Altitude distributions of stratospheric constituents from ground-based measurements at twilight, J. Geophys. Res. Atmos., 96, 15 499–15 511, https://doi.org/10.1029/91JD01361, 1991.

McLinden, C. A., Fioletov, V., Boersma, K. F., Kharol, S. K., Krotkov, N., Lamsal, L., Makar, P. A., Martin, R. V., Veefkind, J. P., and Yang, K.: Improved satellite retrievals of NO_2 and SO_2 over the Canadian oil sands and comparisons with surface measurements, Atmos. Chem. Phys., 14, 3637–3656, https://doi.org/10.5194/acp-14-3637-2014, 2014.

Mebust, A. and Cohen, R.: Observations of a seasonal cycle in NO_x emissions from fires in African woody savannas, Geophys. Res. Lett., 40, 1451–1455, https://doi.org/10.1002/grl.50343, 2013.

Mebust, A. and Cohen, R.: Space-based observations of fire NO_x emissions coefficients: a global biome-scale comparison, Atmos. Chem. Phys., 14, 2509–2524, https://doi.org/10.5194/acp-14-2509-2014, 2014.

Mebust, A. K., Russell, A. R., Hudman, R. C., Valin, L. C., and Cohen, R. C.: Characterization of wildfire NO_x emissions using MODIS fire radiative power and OMI tropospheric NO_2 columns, Atmos. Chem. Phys., 11, 5839–5851, https://doi.org/10.5194/acp-11-5839-2011, 2011.

Miyazaki, K., Eskes, H., and Sudo, K.: Global NO_x emissions estimates derived from an assimilation of OMI tropospheric NO_2 columns, Atmos. Chem. Phys., 12, 2263–2288, https://doi.org/10.5194/acp-12-2263-2012, 2012.

Miyazaki, K., Eskes, H., Sudo, K., and Zhang, C.: Global lightning NO_x production estimated by an assimilation of multiple satellite data sets, Atmos. Chem. Phys., 14, 3277–3305, https://doi.org/10.5194/acp-14-3277-2014, 2014.

Miyazaki, K., Eskes, H., Sudo, K., Boersma, K. F., Bowman, K., and Kanaya, Y.: Decadal changes in global surface NO_x emissions from multi-constituent satellite data assimilation, Atmos. Chem. Phys., 17, 807–837, https://doi.org/10.5194/acp-17-807-2017, https://www.atmos-chem-phys.net/17/807/2017/, 2017.

Nault, B. A., Garland, C., Pusede, S. E., Wooldridge, P. J., Ullmann, K., Hall, S. R., and Cohen, R. C.: Measurements of $CH_3O_2NO_2$ in the upper troposphere, Atmos. Meas. Tech., 8, 987–997, https://doi.org/10.5194/amt-8-987-2015, 2015.

Nault, B. A., Laughner, J. L., Wooldridge, P. J., Crounse, J. D., Dibb, J., Diskin, G., Peischl, J., Podolske, J. R., Pollack, I. B., Ryerson, T. B., Scheuer, E., Wennberg, P. O., and Cohen, R. C.: Lightning NO_x Emissions: Reconciling Measured and Modeled Estimates With Updated NO_x Chemistry, Geophys. Res. Lett., https://doi.org/10.1002/2017GL074436, 2017.

Palmer, P., Jacob, D., Chance, K., Martin, R., Spurr, R., Kurosu, T., Bey, I., Yantosca, R., Fiore, A., and Li, Q.: "Air mass factor formulation for spectroscopic measurements from satellites: Applications to formaldehyde retrievals from the Global Ozone Monitoring Experiment", J. Geophys. Res. Atmos., 106, 14 539–14 550, 2001.

Pickering, K. E., Bucsela, E., Allen, D., Ring, A., Holzworth, R., and Krotkov, N.: Estimates of lightning NOx production based on OMI NO2 observations over the Gulf of Mexico, J. Geophys. Res. Atmos., 121, 8668–8691, https://doi.org/10.1002/2015JD024179, http://dx.doi.org/10.1002/2015JD024179, 2015JD024179, 2016.

Richter, A. and Wagner, T.: The Use of UV, Visible and Near IR Solar Back Scattered Radiation to Determine Trace Gases, in: The Remote Sensing of Tropospheric Composition from Space, edited by Burrows, J., Platt, U., and Borrell, P., Springer, New York, 2011.

Richter, A., Burrows, J. P., Nüß, H., Granier, C., and Niemeier, U.: Increase in tropospheric nitrogen dioxide over China observed from space, Nature, 437, 129–132, https://doi.org/10.1038/nature04092, https://doi.org/10.1038/nature04092, 2005.

Rudlosky, S. D.: Evaluating ENTLN Performance Relative to TRMM/LIS., Journal of Operational Meteorology, 3, 2015.

Russell, A. R., Valin, L. C., Bucsela, E. J., Wenig, M. O., and Cohen, R. C.: Space-based Constraints on Spatial and Temporal Patterns of NOxEmissions in California, 2005-2008, Environ. Sci. Technol., 44, 3608–3615, https://doi.org/10.1021/es903451j, https://doi.org/10.1021/es903451j, 2010.

Russell, A. R., Perring, A. E., Valin, L. C., Bucsela, E. J., Browne, E. C., Wooldridge, P. J., and Cohen, R. C.: A high spatial resolution retrieval of NO_2 column densities from OMI: method and evaluation, Atmos. Chem. Phys., 11, 8543–8554, https://doi.org/10.5194/acp-11-8543-2011, https://doi.org/10.5194/acp-11-8543-2011, 2011.

Russell, A. R., Valin, L. C., and Cohen, R. C.: Trends in OMI NO_2 observations over the United States: effects of emission control technology and the economic recession, Atmos. Chem. Phys., 12, 12 197–12 209, https://doi.org/10.5194/acp-12-12197-2012, 2012.

Ryerson, T. B., Huey, L. G., Knapp, K., Neuman, J. A., Parrish, D. D., Sueper, D. T., and Fehsenfeld, F. C.: Design and initial characterization of an inlet for gas-phase NOymeasurements from aircraft, J. Geophys. Res. Atmos., 104, 5483–5492, https://doi.org/10.1029/1998jd100087, https://doi.org/10.1029/1998jd100087, 1999.

Schreier, S. F., Richter, A., Kaiser, J. W., and Burrows, J. P.: The empirical relationship between satellite-derived tropospheric NO_2 and fire radiative power and possible implications for fire emission rates of NO_x, Atmos. Chem. Phys., 14, 2447–2466, https://doi.org/10.5194/acp-14-2447-2014, https://www.atmos-chem-phys.net/14/2447/2014/, 2014.

SENEX Science Team: SENEX 2013 WP-3D Data: NOAA/Earth System Research Laboratory, Chemical Sciences Division, https://esrl.noaa.gov/csd/groups/csd7/measurements/2013senex/P3/DataDownload/, 2013.

Shusterman, A. A., Teige, V. E., Turner, A. J., Newman, C., Kim, J., and Cohen, R. C.: The BErkeley Atmospheric CO_2 Observation Network: initial evaluation, Atmos. Chem. Phys., 16, 13 449–13 463, https://doi.org/10.5194/acp-16-13449-2016, https://www.atmos-chem-phys.net/16/13449/2016/, 2016.

Slusser, J., Hammond, K., Kylling, A., Stamnes, K., Perliski, L., Dahlback, A., Anderson, D., and DeMajistre, R.: Comparison of air mass computations, J. Geophys. Res. Atmos., 101, 9315–9321, https://doi.org/10.1029/96JD00054, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/96JD00054, 1996.

Sullivan, J. T., McGee, T. J., Langford, A. O., Alvarez, R. J., Senff, C. J., Reddy, P. J., Thompson, A. M., Twigg, L. W., Sumnicht, G. K., Lee, P., Weinheimer, A., Knote, C., Long, R. W., and Hoff, R. M.: Quantifying the contribution of thermally driven recirculation to a highozone event along the Colorado Front Range using lidar, J. Geophys. Res. Atmos., 121, 10,377–10,390, https://doi.org/10.1002/2016JD025229, 2016.

Thornton, J. A., Wooldridge, P. J., and Cohen, R. C.: Atmospheric NO_2: In Situ Laser-Induced Fluorescence Detection at Parts per Trillion Mixing Ratios, Anal. Chem., 72, 528–539, https://doi.org/10.1021/ac9908905, https://doi.org/10.1021/ac9908905, pMID: 10695138, 2000.

Toon, O. B., Maring, H., Dibb, J., Ferrare, R., Jacob, D. J., Jensen, E. J., Luo, Z. J., Mace, G. G., Pan, L. L., Pfister, L., Rosenlof, K. H., Redemann, J., Reid, J. S., Singh, H. B., Thompson, A. M., Yokelson, R., Minnis, P., Chen, G., Jucks, K. W., and Pszenny, A.: Planning, implementation, and scientific goals of the Studies of Emissions and Atmospheric Composition, Clouds and Climate Coupling by Regional Surveys (SEAC4RS) field mission, J. Geophys. Res. Atmos., 121, 4967–5009, https://doi.org/10.1002/2015jd024297, https://doi.org/10.1002/2015jd024297, 2016.

Travis, K. R., Jacob, D. J., Fisher, J. A., Kim, P. S., Marais, E. A., Zhu, L., Yu, K., Miller, C. C., Yantosca, R. M., Sulprizio, M. P., Thompson, A. M., Wennberg, P. O., Crounse, J. D., Clair, J. M. S., Cohen, R. C., Laughner, J. L., Dibb, J. E., Hall, S. R., Ullmann, K., Wolfe, G. M., Pollack, I. B., Peischl, J., Neuman, J. A., and Zhou, X.: Why do models overestimate surface ozone in the Southeast United States?, Atmos. Chem. Phys., 16, 13 561–13 577, https://doi.org/10.5194/acp-16-13561-2016, https://doi.org/10.5194/acp-16-13561-2016, 2016.

van der A, R. J., Eskes, H. J., Boersma, K. F., van Noije, T. P. C., Van Roozendael, M., De Smedt, I., Peters, D. H. M. U., and Meijer, E. W.: Trends, seasonal variability and dominant NOx source derived from a ten year record of NO2 measured from space, J. Geophys. Res. Atmos., 113, D04 302, https://doi.org/10.1029/2007JD009021, http://dx.doi.org/10.1029/2007JD009021, 2008.

van Geffen, J. H. G. M., Boersma, K. F., Van Roozendael, M., Hendrick, F., Mahieu, E., De Smedt, I., Sneep, M., and Veefkind, J. P.: Improved spectral fitting of nitrogen dioxide from OMI in the 405â€“465 nm window, Atmos. Meas. Tech., 8, 1685–1699, https://doi.org/10.5194/amt-8-1685-2015, https://www.atmos-meas-tech.net/8/1685/2015/, 2015.

van Marle, M. J. E., Kloster, S., Magi, B. I., Marlon, J. R., Daniau, A.-L., Field, R. D., Arneth, A., Forrest, M., Hantson, S., Kehrwald, N. M., Knorr, W., Lasslop, G., Li, F., Mangeon, S., Yue, C., Kaiser, J. W., and van der Werf, G. R.: Historic global biomass burning emissions for CMIP6 (BB4CMIP) based on merging satellite observations with proxies and fire models (1750–2015), Geosci. Model Dev., 10, 3329–3357, https://doi.org/10.5194/gmd-10-3329-2017, https://www.geosci-model-dev.net/10/3329/2017/, 2017.

Williams, E., Rothkin, K., Stevenson, D., and Boccippio, D.: Global Lightning Variations Caused by Changes in Thunderstorm Flash Rate and by Changes in the Number of Thunderstorms, J. Appl. Meteorol., 39, 2223–2230, https://doi.org/10.1175/1520-0450(2001)040<2223:glvcbc>2.0.co;2, https://doi.org/10.1175/1520-0450(2001)040<2223:glvcbc>2.0.co;2, 2000.

Wooldridge, P. J., Perring, A. E., Bertram, T. H., Flocke, F. M., Roberts, J. M., Singh, H. B., Huey, L. G., Thornton, J. A., Wolfe, G. M., Murphy, J. G., Fry, J. L., Rollins, A. W., LaFranchi, B. W., and Cohen, R. C.: Total Peroxy Nitrates (ΣPNs) in the atmosphere: the Thermal Dissociation-Laser Induced Fluorescence (TD-LIF) technique and comparisons to speciated PAN measurements, Atmos. Meas. Techn., 3, 593–607, https://doi.org/10.5194/amt-3-593-2010, https://www.atmos-meas-tech.net/3/593/2010/, 2010.

Zare, A., Romer, P. S., Nguyen, T., Keutsch, F. N., Skog, K., and Cohen, R. C.: A comprehensive organic nitrate chemistry: insights into the lifetime of atmospheric organic nitrates, Atmos. Chem. Phys. Discuss., 2018, 1–33, https://doi.org/10.5194/acp-2018-530, https://www.atmos-chem-phys-discuss.net/acp-2018-530/, 2018.

Zhou, Y., Brunner, D., Hueglin, C., Henne, S., and Staehelin, J.: Changes in OMI tropospheric NO_2 columns over Europe from 2004 to 2009 and the influence of meteorological variability, Atmos. Environ., 46, 482–495, https://doi.org/10.1016/j.atmosenv.2011.09.024, https://doi.org/10.1016/j.atmosenv.2011.09.024, 2012.

Zörner, J., Penning de Vries, M., Beirle, S., Sihler, H., Veres, P. R., Williams, J., and Wagner, T.: Multi-satellite sensor study on precipitation-induced emission pulses of NO_x from soils in semi-arid ecosystems, Atmos. Chem. Phys., 16, 9457–9487, https://doi.org/10.5194/acp-16-9457-2016, 2016.

Supplement to "Evaluation of version 3.0B of the BEHR OMI NO_2 product"

Joshua L. Laughner, Qindan Zhu, and Ronald C. Cohen

S1 VCD comparison detail

BEHR v3.0 intercepts and R^2 values are generally similar to or better than NASA SP v3.0, though the discrepancy in intercepts is greater when comparing against Pandora data alone. In theory, comparing against aircraft data, the intercepts would indicate a bias in the stratospheric separation or total column (for Pandora comparisons, it can only be in the total column). In practice, it is not fully orthogonal to errors in the AMF. However, the stratospheric separation and total column will still be a significant component to the intercept, so it is reasonable that the BEHR and NASA intercepts are similar, as both use the same stratospheric separation and total columns.

Campaign	Product	Extended with GEOS-Chem Slope	Intercept	R^2	Extended by extrapolation Slope	Intercept	R^2
DISCOVER-MD	BEHR v3.0B (D)	N/A	N/A	N/A	N/A	N/A	N/A
	BEHR v3.0B (M)	0.888	-8.96×10^{14}	0.149	0.998	1.28×10^{14}	0.0813
	BEHR v2.1C	1.92	-2.69×10^{15}	0.0747	2.27	-7.24×10^{14}	0.0921
	SP v3.0	0.774	-5.35×10^{14}	0.0806	0.884	3.67×10^{14}	0.0469
DISCOVER-CA	BEHR v3.0B (D)	0.627	1.12×10^{15}	0.567	0.716	9.6×10^{14}	0.587
	BEHR v3.0B (M)	0.665	9.91×10^{14}	0.521	0.775	7.98×10^{14}	0.554
	BEHR v2.1C	0.944	8.29×10^{14}	0.554	1.14	5.25×10^{14}	0.625
	SP v3.0	0.512	9.66×10^{14}	0.291	0.513	9.93×10^{14}	0.311
DISCOVER-TX	BEHR v3.0B (D)	0.483	2.66×10^{14}	0.295	0.512	4.54×10^{14}	0.275
	BEHR v3.0B (M)	0.407	3.9×10^{14}	0.321	0.425	5.54×10^{14}	0.238
	BEHR v2.1C	0.867	1.03×10^{15}	0.386	0.836	1.58×10^{15}	0.359
	SP v3.0	0.299	7.15×10^{14}	0.352	0.317	7.9×10^{14}	0.355
DISCOVER-CO ($V > 0$)	BEHR v3.0B (D)	1.03	-5.82×10^{14}	0.0279	1.01	1.26×10^{14}	0.0234
	BEHR v3.0B (M)	0.915	-5.27×10^{14}	0.0244	0.957	8.24×10^{13}	0.062
	BEHR v2.1C	1.84	-1.32×10^{15}	0.018	1.84	-1.65×10^{14}	0.0186
	SP v3.0	0.705	-4.04×10^{14}	0.0553	0.65	1.69×10^{14}	0.0423
SENEX	BEHR v3.0B (D)	2.25	-2.59×10^{15}	0.154	1.72	-1.2×10^{15}	0.0549
	BEHR v3.0B (M)	0.953	-4.6×10^{14}	0.265	0.879	-8.89×10^{13}	0.153
	BEHR v2.1C	1.43	-2.25×10^{14}	0.164	1.48	-2.44×10^{14}	0.111
	SP v3.0	1.07	-5.84×10^{14}	0.274	0.835	2.35×10^{13}	0.107
SEAC4RS	BEHR v3.0B (D)	0.888	-1.51×10^{14}	0.0122	0.695	5.89×10^{14}	0.0473
	BEHR v3.0B (M)	1.22	-4.56×10^{14}	0.228	0.971	5.81×10^{14}	0.376
	BEHR v2.1C	2.63	-1.55×10^{15}	0.597	2.53	-7.82×10^{13}	0.422
	SP v3.0	1.01	-3.97×10^{14}	0.236	0.823	4.31×10^{14}	0.154

Table S1: Slopes, intercepts, and R^2 values for RMA regression of satellite VCDs against in situ calculated VCDs. Outliers are removed before calculating these parameters; negative VCDs are retained unless noted.

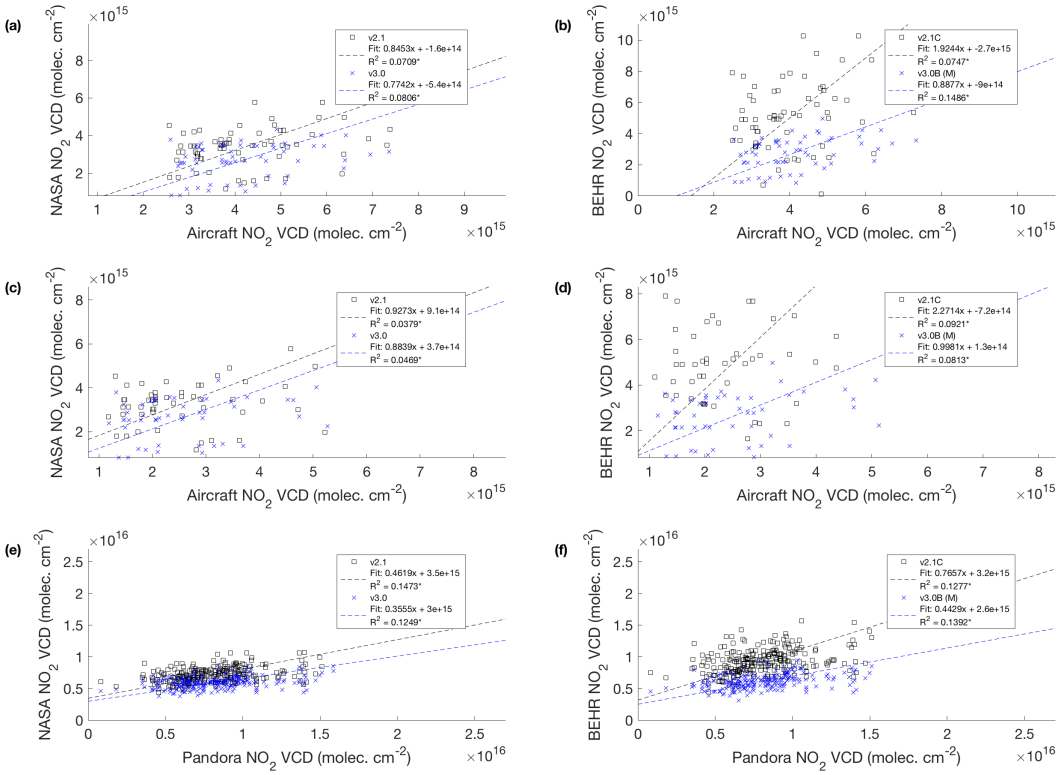

Figure S1: Scatter plots comparing **(a,c,e)** NASA Standard Product and **(b,d,f)** the BEHR product VCDs to **(a,b)** aircraft profiles extended with GEOS-Chem, **(c,d)** aircraft profiles extended by extrapolation, **(e,f)** Pandora columns measured during the DISCOVER-AQ Maryland campaign. An asterisk (*) after the R^2 value in the legend indicates the slope is statistically different from 0 at the 95% confidence level.

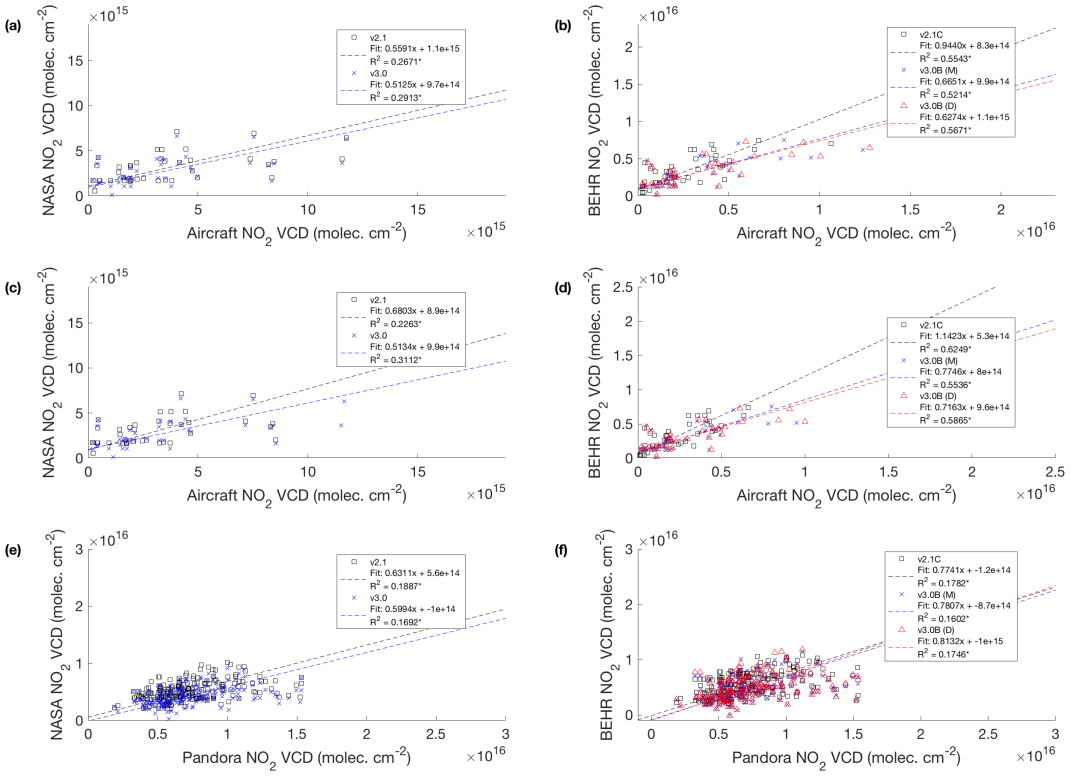

Figure S2: Scatter plots comparing **(a,c,e)** NASA Standard Product and **(b,d,f)** the BEHR product VCDs to **(a,b)** aircraft profiles extended with GEOS-Chem, **(c,d)** aircraft profiles extended by extrapolation, **(e,f)** Pandora columns measured during the DISCOVER-AQ California campaign. An asterisk (*) after the R^2 value in the legend indicates the slope is statistically different from 0 at the 95% confidence level.

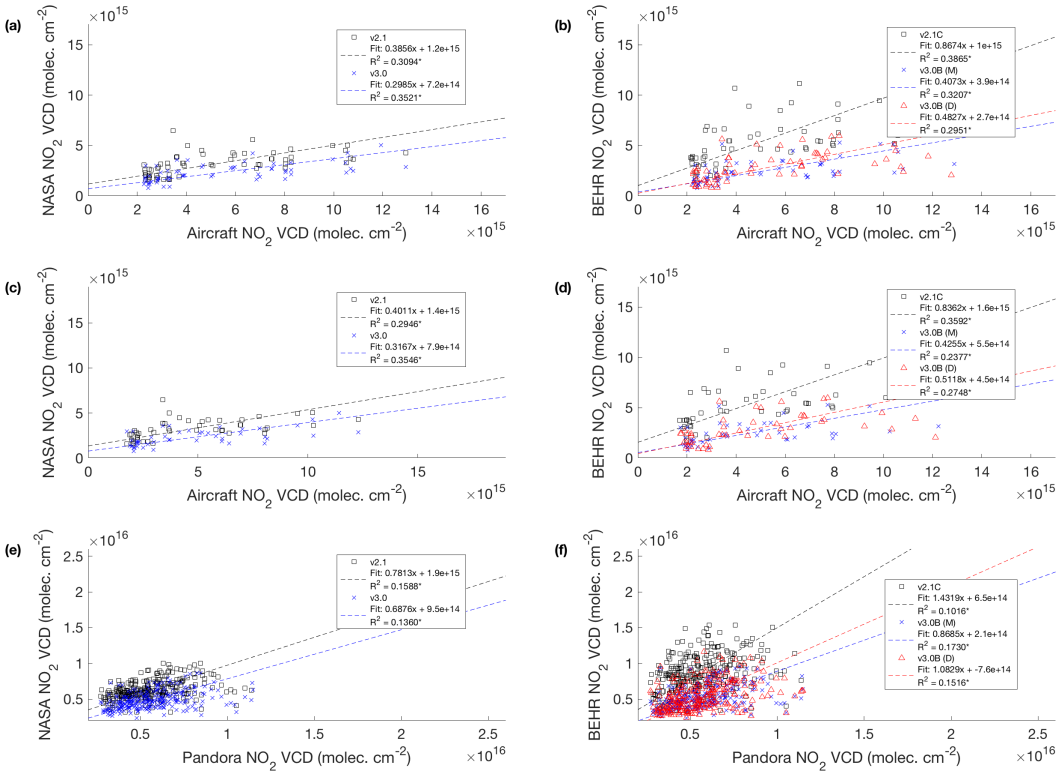

Figure S3: Scatter plots comparing **(a,c,e)** NASA Standard Product and **(b,d,f)** the BEHR product VCDs to **(a,b)** aircraft profiles extended with GEOS-Chem, **(c,d)** aircraft profiles extended by extrapolation, **(e,f)** Pandora columns measured during the DISCOVER-AQ Texas campaign. An asterisk (*) after the R^2 value in the legend indicates the slope is statistically different from 0 at the 95% confidence level.

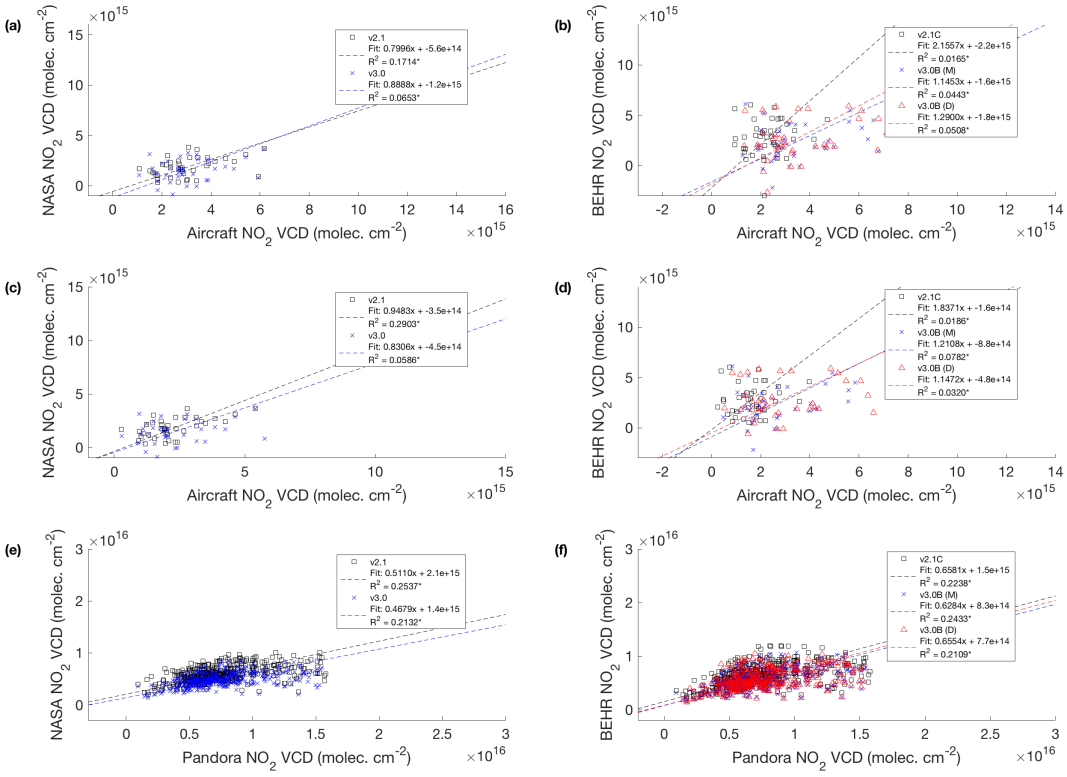

Figure S4: Scatter plots comparing **(a,c,e)** NASA Standard Product and **(b,d,f)** the BEHR product VCDs to **(a,b)** aircraft profiles extended with GEOS-Chem, **(c,d)** aircraft profiles extended by extrapolation, **(e,f)** Pandora columns measured during the DISCOVER-AQ Colorado campaign. Negative VCDs are not removed, in contrast to Table S3. An asterisk (*) after the R^2 value in the legend indicates the slope is statistically different from 0 at the 95% confidence level.

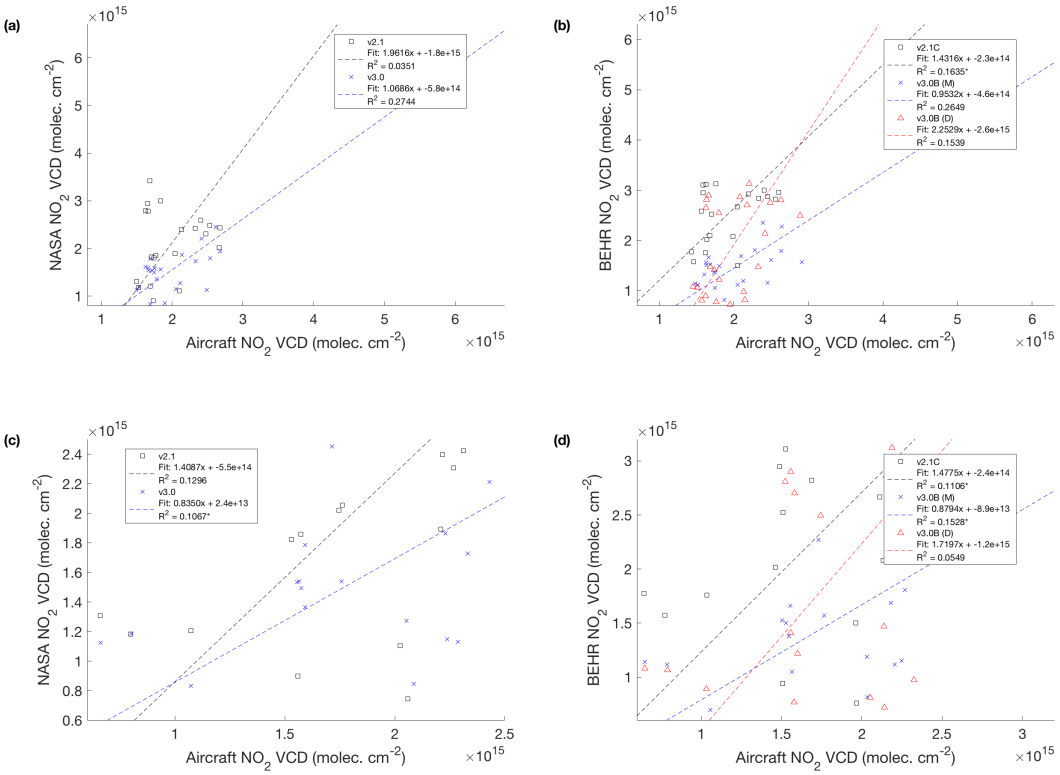

Figure S5: Scatter plots comparing **(a,c)** NASA Standard Product and **(b,d)** the BEHR product VCDs to **(a,b)** aircraft profiles extended with GEOS-Chem and **(c,d)** aircraft profiles extended by extrapolation measured during the SENEX campaign. An asterisk (*) after the R^2 value in the legend indicates the slope is statistically different from 0 at the 95% confidence level.

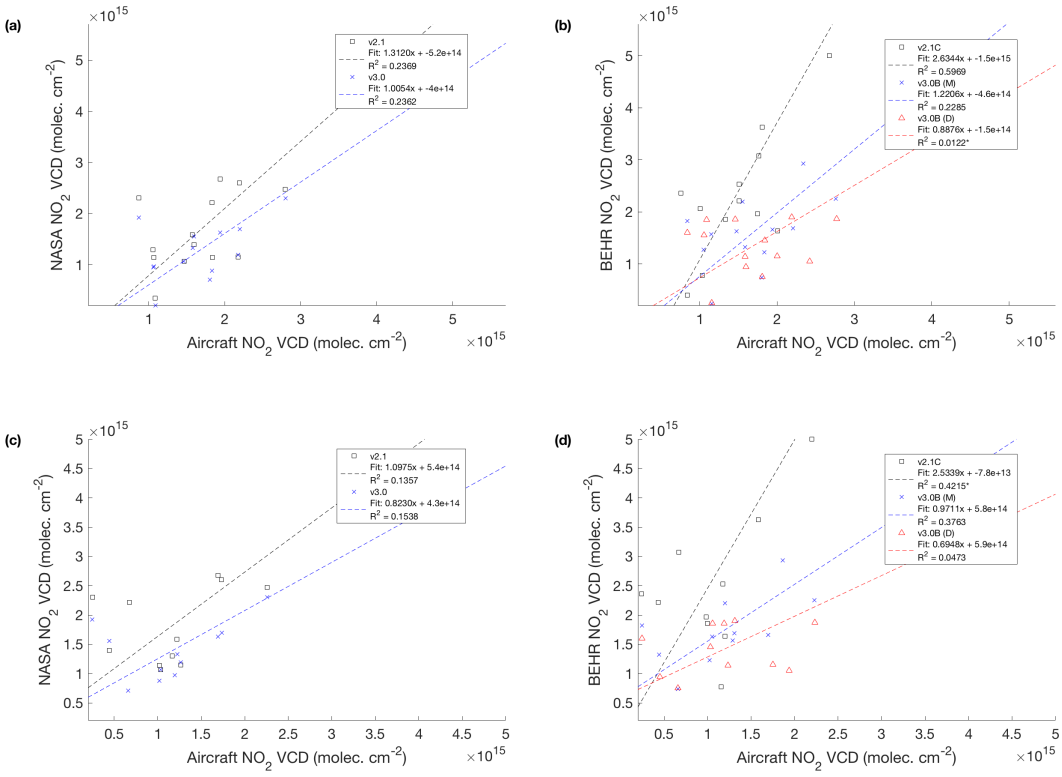

Figure S6: Scatter plots comparing **(a,c)** NASA Standard Product and **(b,d)** the BEHR product VCDs to **(a,b)** aircraft profiles extended with GEOS-Chem and **(c,d)** aircraft profiles extended by extrapolation measured during the SEAC4RS campaign. An asterisk (*) after the R^2 value in the legend indicates the slope is statistically different from 0 at the 95% confidence level.

Campaign	Product	Slope	Intercept	R^2
DISCOVER-MD	BEHR v3.0B (D)	N/A	N/A	N/A
	BEHR v3.0B (M)	0.443	2.55×10^{15}	0.139
	BEHR v2.1C	0.766	3.2×10^{15}	0.128
	SP v3.0	0.355	3.02×10^{15}	0.125
DISCOVER-CA	BEHR v3.0B (D)	0.813	-9.99×10^{14}	0.175
	BEHR v3.0B (M)	0.781	-8.66×10^{14}	0.16
	BEHR v2.1C	0.774	-1.22×10^{14}	0.178
	SP v3.0	0.599	-1.01×10^{14}	0.169
DISCOVER-TX	BEHR v3.0B (D)	1.08	-7.57×10^{14}	0.152
	BEHR v3.0B (M)	0.868	2.05×10^{14}	0.173
	BEHR v2.1C	1.43	6.54×10^{14}	0.102
	SP v3.0	0.688	9.54×10^{14}	0.136
DISCOVER-CO (V > 0)	BEHR v3.0B (D)	0.655	7.71×10^{14}	0.211
	BEHR v3.0B (M)	0.628	8.26×10^{14}	0.243
	BEHR v2.1C	0.658	1.5×10^{15}	0.224
	SP v3.0	0.468	1.38×10^{15}	0.213

Table S2: Slopes, intercepts, and R^2 values for RMA regression of satellite VCDs against Pandora VCDs. Outliers are removed before calculating these parameters.

Campaign	Product	Matched Data Slope	Matched Data Intercept	Matched Data R2	All Data Slope	All Data Intercept	All Data R2
DISCOVER-MD	BEHR v3.0B (D)	N/A	N/A	N/A	N/A	N/A	N/A
	BEHR v3.0B (M)	0.804	-4.16×10^{14}	0.407	0.637	8.46×10^{14}	0.361
	BEHR v2.1C	1.27	-2.92×10^{14}	0.292	0.874	2.23×10^{15}	0.304
	SP v3.0	0.788	-4.65×10^{14}	0.383	0.496	1.72×10^{15}	0.286
DISCOVER-CA	BEHR v3.0B (D)	0.493	1.3×10^{15}	0.525	0.677	1.27×10^{14}	0.321
	BEHR v3.0B (M)	0.509	1.18×10^{15}	0.514	0.661	1.5×10^{14}	0.317
	BEHR v2.1C	0.568	1.46×10^{15}	0.469	0.682	6.54×10^{14}	0.321
	SP v3.0	0.407	1.31×10^{15}	0.408	0.539	4.11×10^{14}	0.282
DISCOVER-TX	BEHR v3.0B (D)	0.692	1.35×10^{14}	0.157	~~1~~ 1.00	-7.39×10^{14}	0.208
	BEHR v3.0B (M)	0.598	3.22×10^{14}	0.217	0.871	-2.26×10^{14}	0.221
	BEHR v2.1C	1.06	1.07×10^{15}	0.19	1.33	7.16×10^{14}	0.173
	SP v3.0	0.527	5.25×10^{14}	0.12	0.736	2.45×10^{14}	0.136
DISCOVER-CO	BEHR v3.0B (D)	0.663	6.38×10^{14}	0.394	0.664	6.92×10^{14}	0.267
	BEHR v3.0B (M)	0.696	2.68×10^{14}	0.347	0.628	7.56×10^{14}	0.277
	BEHR v2.1C	0.741	1.09×10^{15}	0.422	0.679	1.33×10^{15}	0.326
	SP v3.0	0.531	4.32×10^{14}	0.312	0.498	1.05×10^{15}	0.224

Table S3: Slopes, intercepts, and ~~1σ uncertainties~~ R^2 values of BEHR vs. combined aircraft (extended with GEOS-Chem profiles) and Pandora VCDs. Matched slopes use only Pandora data approximately coincident with aircraft profiles to get similar sampling; all uses all valid Pandora data. Outliers and negative VCDs are removed before computing slopes.

S2 Impact of hypsometric surface pressure correction

While we did not carry out an explicit test of how each change to the BEHR algorithm between v2.1C and v3.0B affected the comparison vs. aircraft and Pandora data, we did investigate the effect of different methods of computing the surface pressure of the OMI pixels. The AMF calculation requires a priori knowledge of the average surface pressure of the each OMI pixel, as the location of the surface affects the shape of the scattering weights (e.g. a low reflectivity surface high up in the atmosphere will cause the scattering weights to decrease more rapidly with decreasing altitude than a surface lower down in the atmosphere).

In BEHR v3.0A and earlier versions, this surface pressure was computed by averaging surface elevation data from the GLOBE database (Hastings and Dunbar, 1999) within the OMI pixel, which is then converted to from elevation to pressure using a 7.4 km scale height. In v3.0B, surface pressure taken from the same WRF-Chem model that supplied the NO_2 profiles is adjusted using the same average GLOBE surface elevation in the method described by Zhou et al. (2009). The Zhou et al. (2009) method was originally intended to downscale very coarse ($\sim 3^\circ \times 2^\circ$) modeled surface pressure to OMI pixels using a high resolution terrain database, therefore the effect of using it with already high resolution modeled surface pressure has not been tested.

Laughner et al. (2018) showed that switching to the (Zhou et al., 2009) method increased BEHR NO_2 VCDs by ~ 5 to 10% over the Rocky Mountains during the summer months. This is a small but systematic change, and so was investigated as a way to correct the low bias in BEHR NO_2 VCDs vs. aircraft and Pandora measurements during the DISCOVER-AQ Colorado campaign.

Figure S7 shows regressions of BEHR VCDs against aircraft + Pandora VCDs using both the scale height and hypsometric equation methods of computing surface pressure. The latter method improves the slope by $\sim 18\%$, and while there is a small increase in most BEHR VCDs, the reduction of the 4 largest aircraft VCDs has a larger effect on the slope. The aircraft VCDs change because, when computing a VCD from the aircraft profiles, we integrate from the OMI pixel surface pressure to its tropopause pressure, for consistency between the aircraft VCD calculation and BEHR AMF calculation.

As the two methods of calculating surface pressure do not significantly alter the BEHR VCDs in this comparison, we cannot say explicitly that the surface pressure calculated with the hypsometric equation improves the BEHR retrieval. However, using that surface pressure does lead to greater consistency between BEHR and aircraft VCDs when also applied as the lower limit for integrating the aircraft profiles.

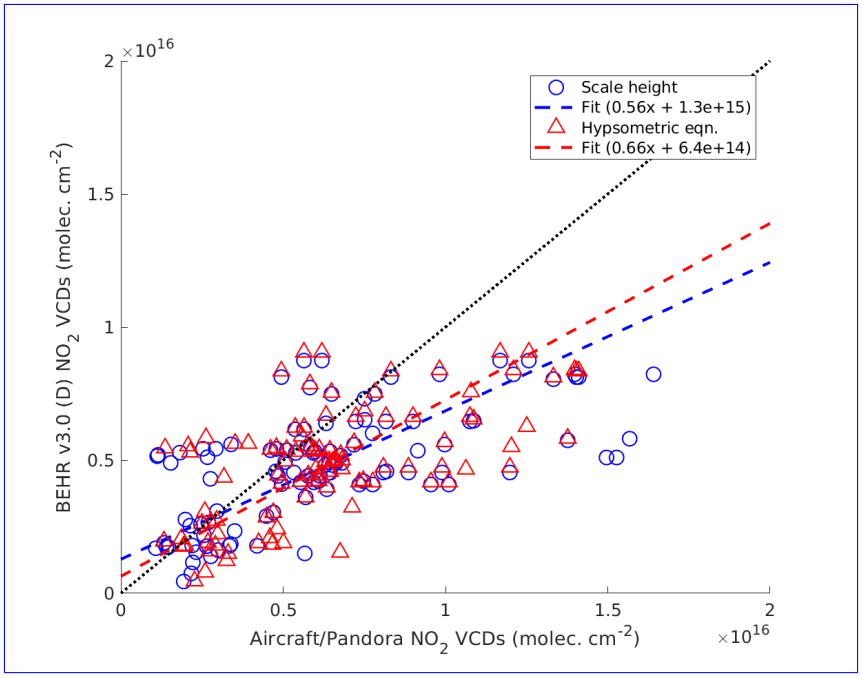

Figure S7: Regression of BEHR v3.0B (D) VCDs vs. aircraft + matched Pandora VCDs for the DISCOVER-CO campaign. The blue series and fit used surface pressured computed using a 7.4 km scale height to convert GLOBE elevations into pressures; the red series used the hypsometric equation as in Zhou et al. (2009).

S3 WRF Lightning - Individual Events

The analysis of a individual convective event taking place near the boundary of Alabama and Georgia on June 14 2012 in shown in Fig. S8. The spatial extent of flashes simulated by WRF-Chem is much broader than that measured by ENTLN, and outside of a few grid cells, the ENTLN flash counts are substantially less than the WRF-Chem simulation.

Outside of the southeast US, although the overall agreement in flash density improves, on smaller scales, we still see that flash density observed by ENTLN is concentrated in the convective core while the simulated flash density spreads over the convective area and fails to reproduce the gradient across the convective core (Fig. S9). The simulated flash density in the convection core is lower than observation despite the total flash counts are still comparable.

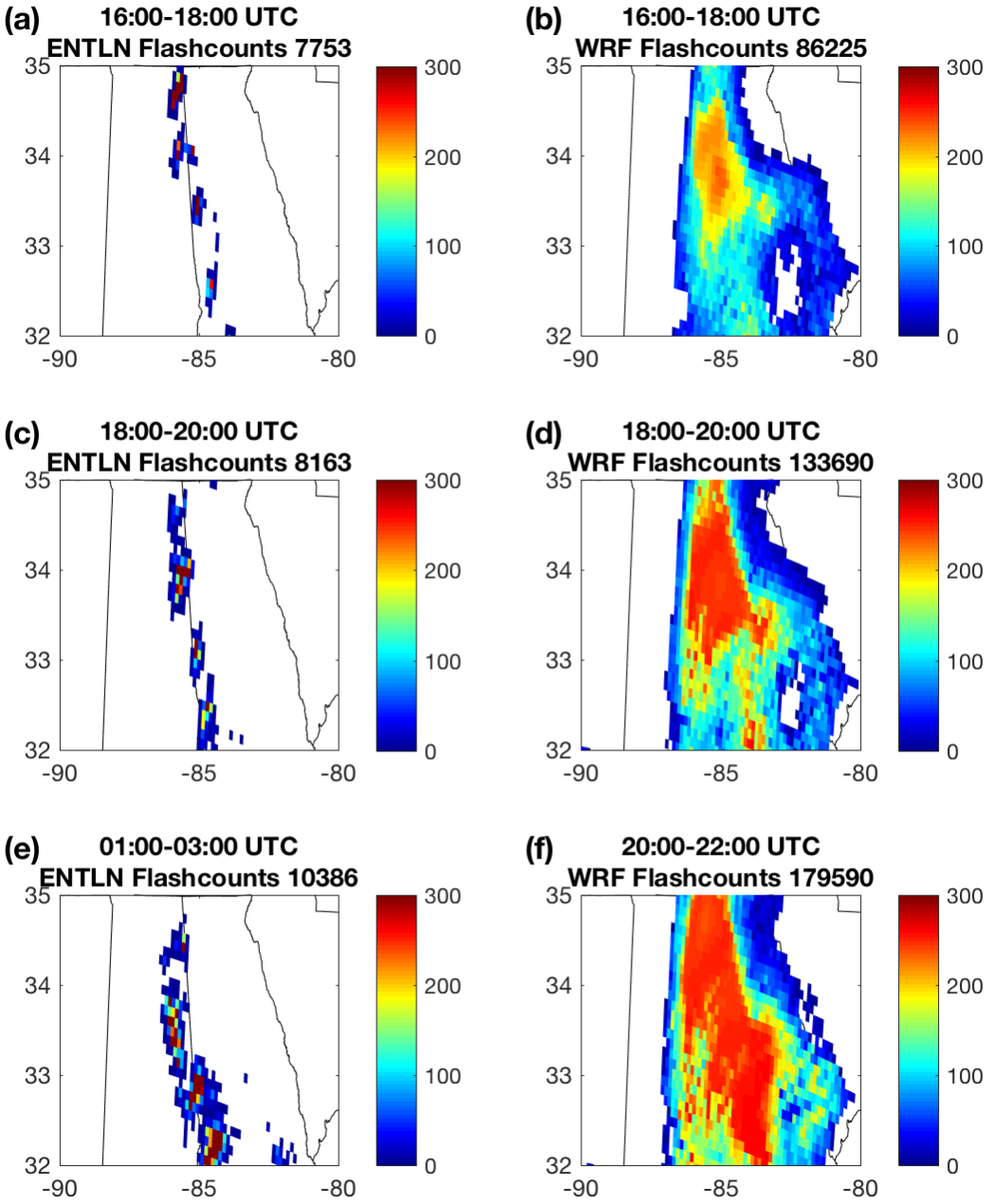

Figure S8: Time-evolved development of storm in the southeast US illustrated by lightning flashes observed by ENTLN **(a, c, e)** and simulated by WRF-Chem **(b, d, f)** on June 14 2012. The number of flashes occurring within the time range is denoted.

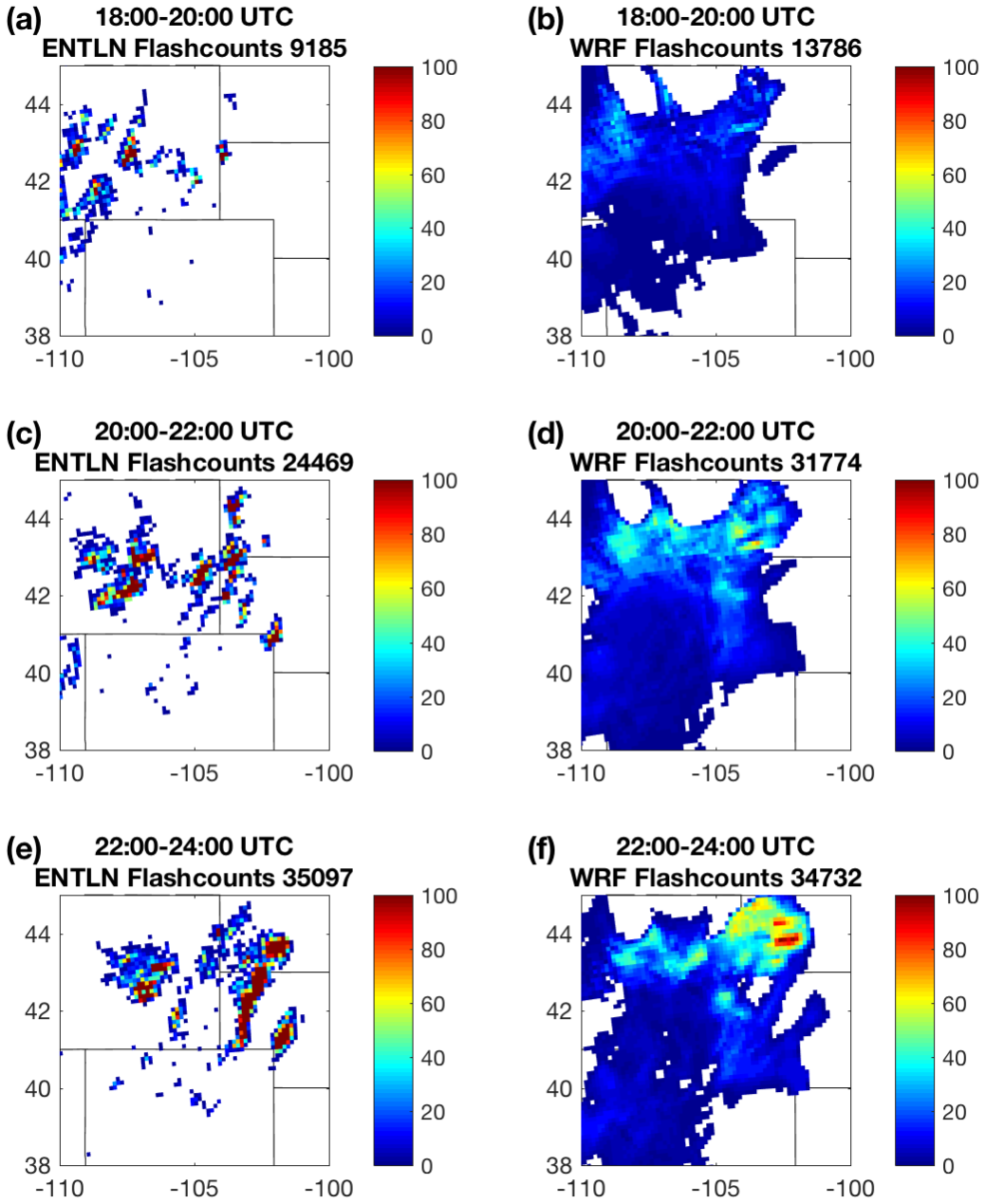

Figure S9: Time-evolved development of storm in the central US illustrated by lightning flashes observed by ENTLN **(a, c, e)** and simulated by WRF-Chem **(b, d, f)** on May 18 2012. The number of flashes occurring within the time range is denoted.

S4 Change in anthropogenic emissions in the a priori profiles, v2 to v3

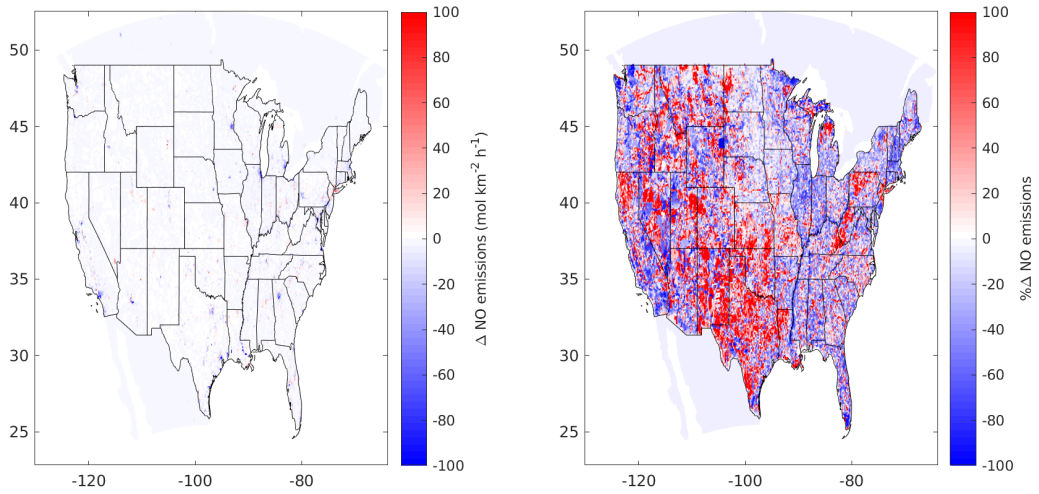

Figure S10: Absolute **(a)** and percent **(b)** change in WRF-Chem NO emissions between the EPA NEI 2005 inventory, unscaled, and the EPA NEI 2011 inventory, scaled to 2012 levels.

Figure S10 shows the absolute and percent difference between the NO emissions in the 2005 and 2011 (scaled to 2012 levels by multiplying by 0.94) EPA NEI inventories. This difference approximates the difference between the emissions used in the a priori profile simulations for BEHR v2.1C and v3.0B. (The emissions for v3.0B are exactly those used; the emissions for v2.1C were reproduced based on the description in Russell et al. 2011).

For BEHR v3.0B with daily profiles, the difference varies from year to year, but the difference will be constant over the domain, as the emissions are simply scaled by a scalar factor (Laughner et al., 2018). The difference shown in Fig. S10 will be $\sim 5\%$ and $\sim 10\%$ greater in 2013 and 2014, respectively, than in 2012. Urban areas see consistently large absolute and percent decreases. Rural areas can have large percent increases or decreases; however, the absolute difference is always very small.

S5 Surface reflectivity evaluation

In BEHR v3, we use the MODIS combined Band 3 MCD43D BRDF coefficients along with the Ross-Thick Li-Sparse kernels to compute a surface reflectivity. This is a computationally simpler approach than Vasilkov et al. (2017), who combined MODIS BRDF coefficients with the VLIDORT radiative transfer model to calculate a modified Lambertian Equivalent Reflectivity (m-LER) that assumes a uniform Lambertian surface under a scattering atmosphere. Here, we evaluate the difference resulting from using the MODIS BRDF directly.

We do so using the MODIS BRDF values using the SCIATRAN radiative transfer model (Rozanov et al., 2005). SCIATRAN is run in plane-parallel scattering mode, including polarization effect. The incident beam is assumed to be unpolarized (Stokes vector [1 0 0 0]). The aerosol profile uses the included WMO aerosol scenario, with 4 layers with upper boundaries of 2 km, 12 km, 30 km, and 100 km above the ground elevation, and aerosol types of continental, continental, background, and background, respectively.

We follow Vasilkov et al. (2017) to calculate modified-LERs (mLERs) from:

$$I(\lambda, \theta, \theta_0, \phi, P_s, R_{\text{BRDF}}) = I_0(\lambda, \theta, \theta_0, \phi, P_s) + \frac{R_{\text{LER}} \cdot T(\lambda, \theta, \theta_0, P_s)}{1 - R_{\text{LER}} \cdot S_b(\lambda, P_s)} \tag{S1}$$

for R_{LER}, which is the mLER. The other variables are:

- I: the top-of-atmosphere intensity at wavelength λ for the given viewing zenith angle (VZA, θ), solar zenith angle (SZA, θ_0), relative azimuth angle (RAA, ϕ), surface pressure (P_s), and BRDF function (R_{BRDF})
- I_0: the top-of-atmosphere intensity for the same wavelength, geometry, and surface pressure as I, but with a 0-reflectivity (i.e. perfectly black) surface
- T: the intensity of light transmitted through the atmosphere; specifically, it represents the solar irradiance that reaches the surface, divided by π (to account for isotropic scattering from the assumed Lambertian surface, which reduces the intensity in a given solid angle), and multiplied by the transmittance of the atmosphere along the viewing direction.
- S_b: the spherical albedo of the atmosphere under the condition of illumination from below. This accounts for additional light incident on the surface due to downward scattering by the atmosphere of light already reflected from the surface.

As in Vasilkov et al. (2017), look up tables (LUTs) are created for I_0, T, and S_b. The I_0 LUT is created from the intensity output of SCIATRAN iterated over three wavelengths (450, 460, and 470 nm) and the same SZAs, VZAs, RAAs, and surface pressures used in the

scattering weight LUT, with surface reflectivity set to 0. The LUTs for T and S_b are created by solving a system of linear equations obtained by rearranging Eq. (S1):

$$\begin{bmatrix} (I - I_0)_{R=0.05} & 1 \\ (I - I_0)_{R=0.1} & 1 \end{bmatrix} \begin{bmatrix} S_b \\ T \end{bmatrix} = \begin{bmatrix} \left(\frac{I-I_0}{R}\right)_{R=0.05} \\ \left(\frac{I-I_0}{R}\right)_{R=0.1} \end{bmatrix} \tag{S2}$$

where the subscripts $R = 0.05$ and $R = 0.1$ indicate that I was calculated with a Lambertian surface reflectivity of 0.05 and 0.1, respectively. These are computed for the same SZAs, VZAs, RAAs, surface pressures, and wavelengths as the I_0 table, although S_b and T are theoretically invariant with respect to some of those parameters. This holds in practice, except for S_b when both the SZA and VZA are very large.

The m-LER is then calculated at 85 sites throughout the continental United States for 189 geometries per site using MCD43D07, MCD43D08, and MCD43D09 coefficients from the first day of each month in 2005 by inputting those coefficients into SCIATRAN to calculate I in Eq. (S1) at 466 nm. Using the previously discussed LUTs for I_0, T, and S_b, we calculate the m-LER from Eq. (S1).

Finally, we calculate the BRDF albedo as in Laughner et al. (2018) for each geometry and month at each site, noting that the RAA definition for SCIATRAN is reversed from that for the BRDF kernels (i.e. $\phi_{\mathrm{SCIA}} = 180 - \phi_{\mathrm{MODIS}}$). We match each m-LER to the corresponding BRDF albedo for the comparison below.

We compare surface reflectances calculated directly from the MODIS coefficients and BRDF kernels with m-LERs calculated with the SCIATRAN radiative transfer model. Figure S11 shows the results for 85 sites (a combination of urban, power plant, and rural sites) with 189 geometry combinations for each site. Figure S11a shows only a 3% variation from a 1:1 line in the regression, and Fig. S11b and c shows a median difference of only 0.005 (8%), with the 75th percentile difference of 0.007 (14%). We retrieved 1 June 2012 with a 14% increase in surface reflectance and found, on average, only a $1.5 \pm 4\%$ (1σ) decrease in the NO_2 column. Since the overall effect of including the radiative transfer calculations on the retrieved columns is small, we choose to use the BRDF coefficients directly to account for the directional dependence of surface reflectance.

We do note that for surface reflectances < 0.3, larger differences in the surface reflectance obtained with radiative transfer calculations compared to the raw BRDF coefficients are associated with large solar zenith angles ($\sim 70°$). This indicates that the uncertainty in individual pixels due to the choice of surface reflectance will be greater during the winter months. However, when individual months are fit, the slope does not change significantly (range 1.011 ± 0.001 to 1.0395 ± 0.0005), indicating that the average uncertainty does not vary significantly with season. This is explored in more detail in Sect. S6 of the main paper.

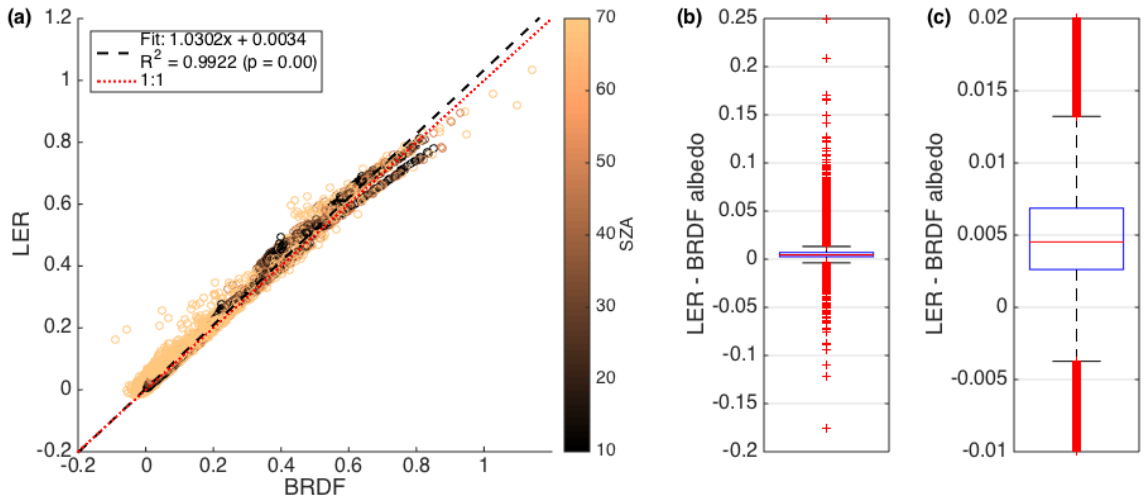

Figure S11: Comparison of a m-LER calculated with MODIS BRDF coefficients to the surface reflectance calculated directly from MODIS BRDF coefficients and kernels. **(a)** scatter plot of the m-LER on the y-axis and direct BRDF on the x-axis, colored by solar zenith angle; a reduced major axis regression is used to fit the data. **(b)** box plot of the difference between the two quantities. The red line is the median, the blue box the 25th and 75th percentiles, the black lines are the furthest non-outlier values, and the red crosses are outliers. **(c)** same as **(b)**, but zoomed in on the interquartile range.

S6 Uncertainty analysis

We determine the uncertainty in the AMF due to surface reflectance, surface pressure, tropopause pressure, cloud pressure, cloud radiance fraction, and profile shape numerically by perturbing each parameter in turn and re-retrieving the NO_2 VCDs with the perturbed values (Table S4). For each perturbation, we reretrieved all of 2012 with the varied parameter.

Surface reflectivity, surface pressure and tropopause pressure are varied by fixed percentages (surface and tropopause pressure are explicitly limited to the range 1020 to 60 hPa). The error in cloud pressure is given as a function of cloud pressure and fraction by Acarreta et al. (2004); we add and subtract the given error for each pixel. Acarreta et al. (2004) also indicates that the error in cloud fraction is < 0.05; to transform that to an error in cloud radiance fraction, we use:

$$\sigma_{f_r} = 0.05 \cdot \left. \frac{\partial f_r}{\partial f_g} \right|_{f_{g,\mathrm{pix}}} \tag{S3}$$

where f_r is the cloud radiance fraction and f_g the cloud fraction. We determine $\partial f_r / \partial f_g$ at $f_{g,\mathrm{pix}}$ (the cloud fraction of a specific pixel) by binning all f_r and f_g for the current OMI

Quantity	Perturbation	Reasoning
Surface reflectivity	±17%	Quadrature sum of 14% LER error and 10% from Schaaf et al. (2010)
Surface pressure	±1.5%	Comparing WRF and BEHR surface pressure
Tropopause presssure	Replace w/NASA tropopause	Alternate method
Cloud pressure	Variable	Fig. 3 of Acarreta et al. (2004)
Cloud radiance fraction	Cloud fraction ±0.05	Acarreta et al. (2004) with correlation of cloud frac. and cloud rad. frac.
Profiles	Quasi-Monte Carlo	Assume variability of model profiles is a reasonable metric

Table S4: Perturbation of input parameters to the AMF calculation used in the uncertainty analysis.

orbit in increments of 0.05 and using that relationship to numerically convert the error in cloud fraction to an error in cloud radiance fraction.

To determine the error due to profile uncertainty, we take advantage of the high spatial and temporal resolution of our WRF-Chem profiles, akin to Boersma et al. (2004). We run two sensitivity retrievals, first allowing the profile to be taken from any day of the same month as the satellite observation, and second allowing each pixel to shift by -0.2, 0, or $+0.2$ degrees in the longitudinal and latitudinal directions for the purpose of matching it with the corresponding NO_2 and temperature profiles. The first is a very conservative simulation of the possible error due to erroneous meteorological drivers (especially wind speed and direction); the second effectively simulates errors in emissions location, chemical kinetics, and transport by moving the pixel so that its profile reflects different aging time since emission.

Figure S12 shows the summed uncertainty for the four seasons as well as the individual contributions to the uncertainty. In all seasons, the a priori NO_2 profiles dominate the uncertainty. ProfileTime is the largest component in all seasons; the represents the uncertainty due primarily to errors in wind direction and speed, since it is calculated by randomly choosing profiles from a different day of the same month as the OMI data. It is unsurprising that this is the greatest contributor to uncertainty, since errors in meteorology may completely change the NO_2 profile of any given pixel, i.e. is it downwind of a source or not, as well as the impact of lightning in the SE US. This is a conservative upper bound, as we saw in Sect. 3.2, WRF captures the plume direction well $\sim 70\%$ of the time, whereas the uncertainty analysis essentially assumes that the WRF winds are uncorrelated with the real winds. Reducing the uncertainty by 70% as a rough correction would make it of similar magnitude to the other contributions and significantly reduce the total uncertainty.

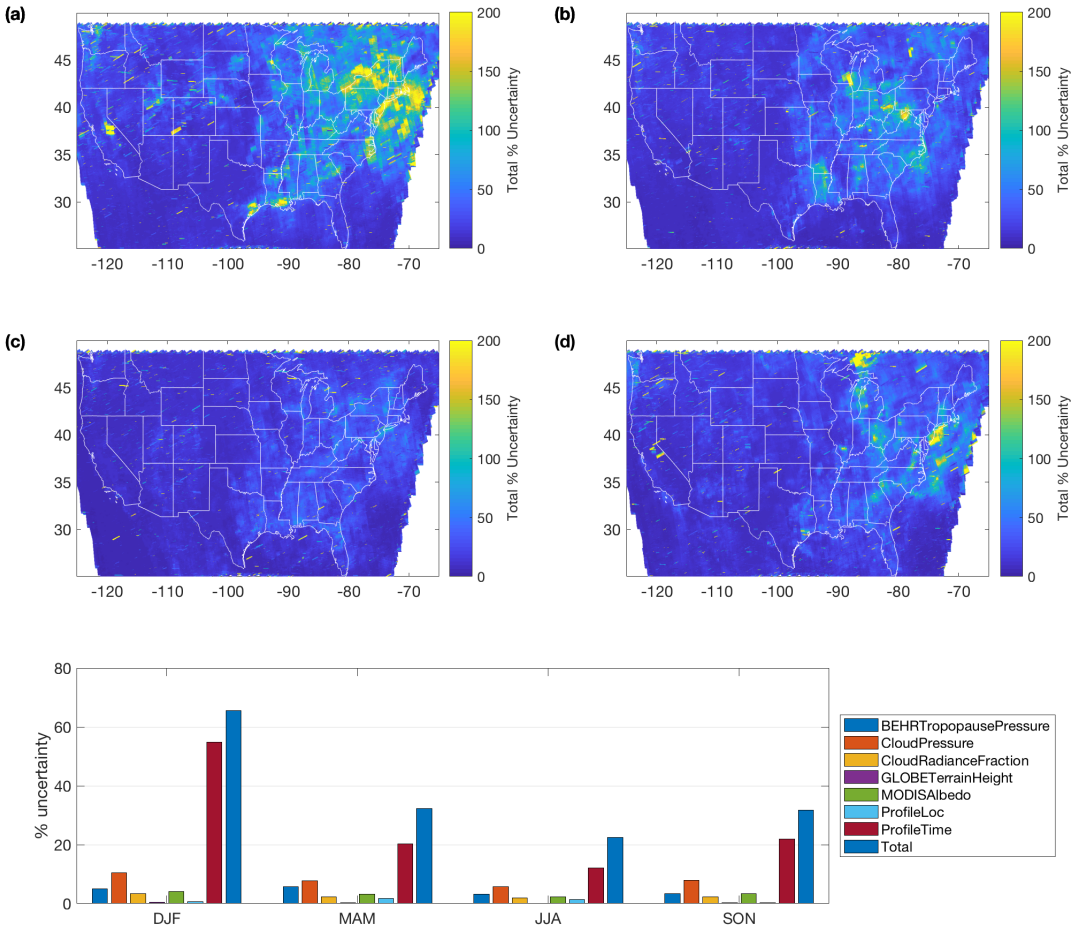

Figure S12: **(a–d)** Total percent uncertainty in tropospheric NO_2 VCDs for **(a)** Jan, Feb, Dec; **(b)** Mar.–May, **(c)** June–Aug., and **(d)** Sept.–Nov., 2012. **(e)** The domain average effect of each varied parameter and the domain average total uncertainty for the same four time periods.

The tropopause and cloud pressures are the next two largest contributors to uncertainty in most seasons. Of the non-profile contributors, the retrieval is most sensitive to cloud pressure. The retrieval sensitivity to the other four non-profile parameters is similar (~ 5 to -10%) in all seasons. The tropopause pressure is generally the second largest non-profile contributor to the uncertainty as one of the integration limits in the AMF calculation; the sensitivity of the NO_2 columns to it and the terrain height ($\%\Delta V_{NO_2}/\%\Delta p_{trop}$) are generally similar ($\sim 0.25\%/\%$, not shown), but the greater uncertainty in the tropopause pressure calculation causes it to have the greater impact on the retrieved VCDs.

Overall, the uncertainty due to the AMF calculation is $\sim 70\%$ in the winter, but much smaller ($\leq 30\%$) during the remainder of the year. The 30% uncertainty is similar to that calculated for polluted conditions in Boersma et al. (2004). This seems reasonable, as in winter, longer NO_x lifetime means that more pixels will have high levels of surface NO_2, and getting the wind direction wrong (i.e. what is tested with ProfileTime) will have effects over larger ares. In the summer the error in urban plumes is still important, but over smaller areas. The highest uncertainties are found in the northeast US, which has a significant number of urban areas. Our greater average uncertainty compared to Boersma et al. (2004) likely follows from the greater variability of our 12 km a priori profiles than the $5^\circ \times 3.75^\circ$ used in Boersma et al. (2004).

Data files containing the seasonal average uncertainties may be downloaded at `behr.cchem.berkeley.edu` for users who require spatially varying uncertainty information for their applications. It is also included in the data repository for this paper.

References

Acarreta, J. R., J. F. De Haan, and P. Stammes (2004). "Cloud pressure retrieval using the O2-O2 absorption band at 477 nm". In: *J. Geophys. Res. Atmos.* 109.D5, p. D05204. ISSN: 2156-2202. DOI: 10.1029/2003JD003915. URL: http://dx.doi.org/10.1029/2003JD003915.

Boersma, K. F., H. J. Eskes, and E. J. Brinksma (2004). "Error analysis for tropospheric NO2retrieval from space". In: *J. Geophys. Res. Atmos.* 109.D4, n/a–n/a. DOI: 10.1029/2003jd003962. URL: https://doi.org/10.1029/2003jd003962.

Hastings, D.A. and P.K. Dunbar (1999). *Global Land One-kilometer Base Elevation (GLOBE) Digital Elevation Model, Documentation, Volume 1.0. National Oceanic and Atmospheric Administration, National Geophysical Data Center, 325 Broadway, Boulder, Colorado 80303, U.S.A.* Key to Geophysical Records Documentation (KGRD) 34.

Laughner, J. L., Q. Zhu, and R. C. Cohen (2018). "The Berkeley High Resolution Tropospheric NO_2 Product". In: *Earth System Science Data Discussions* 2018, pp. 1–33. DOI: 10.5194/essd-2018-66. URL: https://www.earth-syst-sci-data-discuss.net/essd-2018-66/.

Rozanov, A. et al. (2005). "SCIATRAN 2.0 – A new radiative transfer model for geophysical applications in the 175–2400nm spectral region". In: *Advances in Space Research* 36.5,

pp. 1015–1019. DOI: 10.1016/j.asr.2005.03.012. URL: https://doi.org/10.1016/j.asr.2005.03.012.

Russell, A. R. et al. (2011). "A high spatial resolution retrieval of NO_2 column densities from OMI: method and evaluation". In: *Atmos. Chem. Phys.* 11.16, pp. 8543–8554. DOI: 10.5194/acp-11-8543-2011. URL: https://doi.org/10.5194/acp-11-8543-2011.

Schaaf, Crystal Barker et al. (2010). "Aqua and Terra MODIS Albedo and Reflectance Anisotropy Products". In: *Land Remote Sensing and Global Environmental Change*. Springer New York, pp. 549–561. DOI: 10.1007/978-1-4419-6749-7_24. URL: https://doi.org/10.1007/978-1-4419-6749-7_24.

Vasilkov, Alexander et al. (2017). "Accounting for the effects of surface BRDF on satellite cloud and trace-gas retrievals: a new approach based on geometry-dependent Lambertian equivalent reflectivity applied to OMI algorithms". In: *Atmospheric Measurement Techniques* 10.1, pp. 333–349. DOI: 10.5194/amt-10-333-2017.

Zhou, Y. et al. (2009). "An improved tropospheric NO_2 retrieval for OMI observations in the vicinity of mountainous terrain". In: *Atmos. Meas. Tech.* 2.2, pp. 401–416. DOI: 10.5194/amt-2-401-2009. URL: https://www.atmos-meas-tech.net/2/401/2009/.