**Reviewer 1 Comments and Responses:**

**Cooper et al. "Effects of *a priori* shape assumptions on comparisons between satellite $NO_2$ columns and model simulations" presents a case study using synthetic OMI observations and the GEOS-Chem adjoint to show how different methods of accounting for the vertical sensitivity of satellite $NO_2$ measurements when comparing to model $NO_2$ fields affects emissions inferred from said comparison. This paper may have been a better fit for GMD rather than ACP, as it primarily touches on model comparison, but is also relevant to the remote sensing community as it informs what information must be contained in satellite products for effective model comparison, and to the broader atmospheric community for understanding possible sources of error in constrained emissions, so ACP is also appropriate.**

**I do have two concerns about the experimental design. First, I question whether using scattering weights computed for average OMI observing geometry in a $4^\circ \times 5^\circ$ grid cell is appropriate for creating synthetic observations. Second, some of the choices for prior test cases are not very relevant to current $NO_2$ retrievals. I will address these more below. If the authors can address these concerns, then this manuscript should be published in ACP.**

**Major concerns**

**Representativeness of average scattering weights**

**In sect. 3.2.1, line 226, the authors say: "To represent typical conditions, average scattering weight profiles for each grid box are found by averaging scattering weights for OMI observations during July 2010." I have two questions about this. First, it is ambiguous whether the mean scattering weights in question are found by averaging weights providing in the product (implied by "...average scattering weight profiles for each grid box are found by averaging scattering weights for OMI observations during July 2010...") or by using the average viewing conditions to compute the scattering weight vector for the average viewing angles, albedo, surface pressure, etc. (implied by "...OMI scattering weights are calculated using the LIDORT radiative transfer model (Spurr, 2002) by providing LIDORT with the observation geometry of the OMI observations and aerosol profiles from the GEOS-Chem base simulation..."). I'm assuming the latter, but this could be made clearer.**

**Assuming that the authors calculated their own scattering weights from the average viewing conditions, my second concern is that while this simplifies the problem of computing synthetic observations for each model step, it may not adequately represent the variation in OMI measurements during that time. Scattering weights depend nonlinearly on observation geometry, so:**

$$w(z|\overline{\theta}_s, \overline{\theta}_v, \overline{\phi}, \overline{a}_{\text{surf}}, \overline{p}_{\text{surf}}) \neq \frac{1}{c_{\text{obs}}} \sum_{i=1}^{c_{\text{obs}}} w(z|\theta_{s,i}, \theta_{v,i}, \phi_i, a_{\text{surf},i}, p_{\text{surf},i})$$

**where $\theta_s$ is the solar zenith angle, $\theta_v$ the viewing zenith angle, $\varphi$ the relative azimuth angle, $a$ the surface reflectivity, $p_{surf}$ the surface pressure, and overlined quantities represent grid cell averages. In other words, the vector of scattering weights corresponding to the average observation conditions is not guaranteed to be the same as the average of scattering weight vectors for all individual observations. Concretely, consider two observations, one with a viewing zenith angle of 0° and one at ~ 60°. The average of these two observations' scattering weights is not guaranteed to be the same as for an observation of 30°.**

**That being said, it may well be close enough, especially averaged over a 4°×5° grid cell. If the authors can show that the difference between using mean scattering weights and the mean of synthetic observations computed using individual OMI observation scattering weights is within the measurements' uncertainty for at least a few days of synthetic observations, then I think that would be adequate.**

Thank you for your comments. The method we use to calculate average scattering weight profiles is to provide LIDORT with observation geometry from individual OMI observations, and then average the resulting scattering weight profiles. We have edited the text to clarify this procedure (Line 224):

"Scattering weights are calculated using the LIDORT radiative transfer model (Spurr, 2002) by providing LIDORT with the observation conditions of OMI observations during July 2010, which are used to represent typical viewing conditions of low earth orbit satellite observations, and aerosol profiles from the GEOS-Chem base simulation. To represent typical conditions, these representative scattering weight profiles for each grid box are used to produce the synthetic slant columns."

The difference between using an average scattering weight profile and using individual observation scattering weights is indeed small when averaged over a 4°x5° grid. We now discuss this on Line 229:

"Tests performed for all 4°x5° grid boxes used here indicate that the mean relative difference between an air mass factor calculated using an average scattering weight profile and the average of air mass factors using observation-specific scattering weight profiles is less than 4%."

**Relevance of prior test cases**

**Of the shape factor test cases described in sect. 3.2.2, the $SF_{trop}$ and $SF_{BL}$ cases are not particularly relevant for satellite measurements. Both of the two main global OMI $NO_2$ retrievals (NASA SP3, Krotkov et al. 2017; QA4ECV, see Williams et al. 2016 for the $NO_2$ profiles) use ~ 1° resolution for their $NO_2$ profiles. Therefore, the $SF_{BL}$ and $SF_{trop}$ cases,**

**which assume one profile globally are not representative of any major satellite product. In fact, the more relevant question, assuming that 4°×5° is still a common resolution for adjoint modeling, is what happens if the satellite prior is *higher* resolution than the model profile.**

**In my opinion, a sixth test case similar to $SF_{prior}$ but using a set of priors from a 2°×2.5° or 1°×1.25° GEOS-Chem simulation would add value to the paper by studying the effect of having the satellite product's prior at higher spatial resolution than the adjoint model. Also, if the $SF_{BL}$ and $SF_{trop}$ cases are retained, it should be clearly stated that they represent extreme cases that do not represent any modern $NO_2$ product.**

Thank you for this suggestion. We now include a test $SF_{finer}$ that uses a set of priors based on a 2°x2.5° GEOS-Chem simulation. We also note in Section 3.2.2 that the $SF_{BL}$ and $SF_{trop}$ cases are extreme cases that do not represent any modern $NO_2$ product (Line 264):

"The $SF_{BL}$ and $SF_{trop}$ tests do not represent any modern retrieval algorithms, but are used as extreme examples of using an *a priori* that assumes no spatial variability"

**Other primary concerns**

- **In sect. 3.2, line 210, the authors say that they use one observation per grid box per hour. But OMI will only observe a given location twice per day, maximum, and usually only once per day at about 13:30 local standard time. Are you then filtering these once-per-hour observations down to the ones OMI would actually observe?**

  OMI viewing geometries are used here only as an example of typical viewing geometries of low earth orbit satellite instruments for the scattering weight calculations. The synthetic observations used here are not meant to be synthetic OMI observations or represent the spatial or temporal sampling of OMI. We have clarified this in the text at line 224 as previously noted:

  "…by providing LIDORT with the observation geometry of OMI observations during July 2010, which are used to represent typical viewing geometries of low earth orbit satellite observations, and aerosol profiles from the GEOS-Chem base simulation."

- **I don't follow Eq. (20). Specifically why $n_a$ shows up on the right hand side. Given that:**

$$M(\mathbf{n}) = \frac{\sum_i \mathbf{w}_i \mathbf{n}_i}{\sum_i \mathbf{n}_i}$$

**and**

$$\mathbf{A}_i(\mathbf{n}) = \frac{\mathbf{w}_i}{M(\mathbf{n})}$$

**then to compute $M(\mathbf{n}_m)$ given $\mathbf{A}(\mathbf{n}_a)$ you only need to multiply $\mathbf{A}(\mathbf{n}_a)$ by $M(\mathbf{n}_a)$ to extract the necessary scattering weights:**

$$M(\mathbf{n}_m) = \frac{\sum_i \mathbf{w}_i \mathbf{n}_{m,i}}{\sum_i \mathbf{n}_{m,i}}$$

$\mathbf{w}_i = \mathbf{A}_i(\mathbf{n}_a) \cdot M(\mathbf{n}_a)$

$$\therefore M(\mathbf{n}_m) = \frac{\sum_i \mathbf{A}_i(\mathbf{n}_a) M(\mathbf{n}_a) \mathbf{n}_{m,i}}{\sum_i \mathbf{n}_{m,i}}$$
$$= M(\mathbf{n}_a) \frac{\sum_i \mathbf{A}_i(\mathbf{n}_a) \mathbf{n}_{m,i}}{\sum_i \mathbf{n}_{m,i}}$$

**I don't think you need $\mathbf{n}_a$ to compute $M(\mathbf{n}_m)$ as long as $M(n_a)$ is included in the satellite data (both the NASA SP3 and QA4ECV OMI $NO_2$ products include the tropospheric AMFs), and given only the AKs and prior profile, it would be difficult if not impossible to compute $M(n_a)$. That means the statement on line 320 about needing the a priori profiles in the dataset is incorrect.**

Thank you for noticing the error in Equation 20. The $n_a$ term on the right hand side should indeed be $n_m$. The text at line 341 has been adjusted accordingly:

"This is most straightforward when scattering weights (rather than averaging kernels) are provided alongside retrieved column data, as scattering weights and shape factors are independently calculated, however simulation-based air mass factors can be calculated using the averaging kernel and *a priori*-based air mass factor via Eq. 19."

**Minor corrections**

- **For Eq. (8), it would be good to make clear that A(z) is the column averaging kernel and therefore a vector, since in Rodgers and Conner, the capital A is typically the full AK matrix. But I agree that following the convention in Eskes and Boersma (2003) is best.**
  We now clarify that A(z) is the column averaging kernel on Line 109.

- **Eq. (9) doesn't seem to be used anywhere else in the paper, and technically is inconsistent with the implicit definition of w(z) in Eq (8). Recommend removing Eq. (9).**

This equation has been removed.

- **In Eq. (17) and Eq. (18) it's unclear what is being summed. Recommend using *I* subscripts to make clear what terms are iterated in the sum.**

  This has been changed.

- **On line 198 in sect. 3.1, the authors say that Δ is computed using either Eq. (12) or (17). Given that much of sect. 2 was spent establishing that these two equations differ, this should be clarified. If I understood correctly, which equation is used effectively depends on which shape factor was used for a given test. If so, I recommend saying that explicitly.**

  This is now explicitly stated on line 193: "(Δ, from either Eq. (11) if using a simulation-based air mass factor or Eq. (16) if using the retrieval *a priori*-based air mass factor)"

- **For the different shape factors, have you considered the impact of profiles simulated by a model with systematically, rather than randomly (as in $SF_{n30}$), different emissions?**

  Thank you for this suggestion. We now include a test $SF_{diffem}$ that considers *a priori* profiles simulated by a model with systematically different emissions.

**References**


Eskes, H. J. and Boersma, K. F.: Averaging kernels for DOAS total-column satellite retrievals, Atmospheric Chemistry and Physics, 3, 1285–1291, doi:10.5194/acp-3-1285-2003, URL https://www.atmos-chem-phys.net/3/1285/2003/, 2003.

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 NO2 standard product, Atmospheric Measurement Techniques, 10, 3133–3149, doi:10.5194/amt-10-3133-2017, URL https://doi.org/10.5194/amt-10-3133-2017, 2017.

Williams, J. E., Boersma, K. F., Sager, P. L., and Verstraeten, W. W.: The high-resolution version of TM5-MP for optimised satellite retrievals: Description and Validation, doi: 10.5194/gmd-2016-125, URL https://doi.org/10.5194/gmd-2016-125, 2016.

**Reviewer 2 Comments and Responses:**

**This manuscript addresses an interesting feature of sensitivity of NOx emission inversions to a priori profile shape assumptions in AMF of satellite NO2 columns. Authors conclude "As the difference between the simulated profile shape and the a priori profile shape increases, so do the corresponding assimilated emission errors". In the discussion section, however, the authors indicate that the adjoint inversion is less sensitive to vertical representativeness errors in cases where emissions are poorly constrained. It is noted that choice of AMF will become increasingly important to adjoint inversions as emission inventories improve. The manuscript delivers some new and intriguing messages to satellite and air quality modeling community. I think the manuscript is well written. The introduction and revisit of AMF and averaging kernel are neat and helpful. There are some parts that need further investigation and explanation. Hope the authors revise and improve the manuscript before final publication.**

Thank you for your comments. We address the specific comments below.

**Major point Line 304-310: the manuscript deals with the "truth" emissions that deviate only 5% from the original anthropogenic NOx emissions. Here, it is written that the tests using random 15% or 30% perturbations to emissions were insensitive to the AMF. In real cases, NOx emission inventory errors are quite large (> 30%). Do the authors mean that choice of a priori shape factor is not important for most of real emission study cases? Please show the results from the random 15% or 30% perturbation tests (or other new cases if possible) and discuss more on applications to the real world problems (e.g., Qu et al., 2017).**

Results from the random 30% perturbation tests are now included in the Results section.

We now clarify in the conclusion that our results show that recent emission study cases are likely insensitive to the AMF (Line 330):

"This indicates that while the adjoint cost function is mathematically dependent on the AMF, the inversion is less sensitive to vertical representativeness errors in cases where emissions are poorly constrained, as is the case in recent adjoint inversion studies (e.g. Qu et al., 2017). However, choice of AMF will become increasingly important to adjoint inversions as emission inventories improve."

**Minor points 1. Examples of inconsistent a priori shape factor: I do not think these days retrieval groups use SF_BL, SF_Trop type a piori. In the abstract, up to 80% increased error is based on this choice. I am not sure if readers need to take this number seriously.**

We clarify in Section 3.2.2 that the SFBL and SFtrop tests are extreme cases that do not represent typical retrievals (Line 264):

"The $SF_{BL}$ and $SF_{trop}$ tests do not represent any modern retrieval algorithms, but are used as extreme examples of an *a priori* that assumes no spatial variability."

We also changed the value cited in the abstract to 30%, representing the error of the more realistic test cases.

**2. Line 274-276: I am not sure what these mean.**

We have replaced these statistics with root mean square differences, which more clearly describe the difference between the *a posteriori* emissions from the $SF_M$ and the other tests.

**3. Line 288-292: Examination of the mathematical frameworks behind two common methods for comparing simulated and retrieved columns highlights how the method introduced by Palmer et al. (2001) facilitates separation of observations sensitivity (scattering weights) from the profile shape (shape factor) enabling the model-retrieval comparison to be independent of a priori profile assumptions. In the last part, model-retrieval comparison to be independent of a priori. . .It is confusing because the main conclusion of the manuscript is that the model-retrieval comparison is not independent of a priori (in certain cases).**

The mathematical framework outlined in Section 2 shows that indeed any model-retrieval comparisons is dependent on the prior. One conclusion in the manuscript is that the 4D-Var assimilation process, while mathematically dependent on the prior, is not sensitive to the prior under some conditions. We have clarified this by adding Table 2 which indicates how the model-retrieval comparisons are affected by the *a priori* profile, and by adjusting the discussion to clarify this distinction (Line 326):

"Inversion tests performed using synthetic observations based on random 30% perturbations to emissions were insensitive to the AMF, despite large differences in *a priori* vertical column densities. In these tests, the adjoint cost function was more sensitive to the larger difference between the observed and simulated slant columns (i.e. $\Omega_{s,m} - \Omega_{s,o}$ in Eq. (13) and (19)) than to AMF. This indicates that while the adjoint cost function is mathematically dependent on the AMF, the inversion is less sensitive to vertical representativeness errors in cases where emissions are poorly constrained…"

**4. Line 307: Add "s" in the subscript.**

This has been corrected.

**5. Line 315-Line 320: It is good to emphasize these again. But I believe that retrieval groups are already doing this. It might be good to mention various data supported by the retrieval groups.**

Some retrieval data products do include scattering weights, but not all of them. However, we feel it is best to emphasize this as a general recommendation without singling out any particular retrieval groups.

**6. Can the posteriori NOx emission difference (%) for highly polluted cases be shown in Table 2? This type of results will also be useful for higher perturbation cases.**

Percent differences were not included in Table 2 for clarity. However, we now include mean "true" emission values in the captions to allow the reader to make such comparisons if desired.

**7. Is the number of iterations of 4D VAR assimilations for all the test cases the same? How many iterations are requited for the tests?**

We now mention the number of 4D Var iterations on Line 198: “Tests performed here required 20-30 iterations to minimize the cost function.”

**8. Is there a possibility that model spatial resolutions affect the results?**

The mathematical framework outlined in Section 2 presents the main point of the paper, which is that consistency between simulated profiles and shape factor profiles used to calculate AMFs is essential. This is true regardless of model resolution. We now include a test where the *a priori* profile is generated by a higher resolution model (GEOS-Chem run at 2x2.5 resolution). This test further supports the main conclusion of the paper. From Line 316:

“The $SF_{finer}$ test indicates that using a higher resolution model to generate *a priori* profiles does not provide an advantage in simulation-observation comparisons, as consistency between the simulation profile and the AMF shape factor is of greater importance.”

# Effects of *a priori* profile shape assumptions on comparisons between satellite $NO_2$ columns and model simulations

Matthew J. Cooper[1,2], Randall V. Martin[2,1,3], Daven K. Henze[4], Dylan B.A. Jones[5]

1. Department of Physics and Atmospheric Science, Dalhousie University, Halifax, Nova Scotia, Canada

2. Department of Energy, Environmental & Chemical Engineering, Washington University in St. Louis, St. Louis, Missouri, USA

3. Harvard-Smithsonian Center for Astrophysics, Cambridge, Massachusetts, USA

4. Department of Mechanical Engineering, University of Colorado, Boulder, Colorado, USA

5. Department of Physics, University of Toronto, Toronto, Ontario, Canada

## Abstract

A critical step in satellite retrievals of trace gas columns is the calculation of the air mass factor (AMF) used to convert observed slant columns to vertical columns. This calculation requires *a priori* information on the shape of the vertical profile. As a result, comparisons between satellite-retrieved and model-simulated column abundances are influenced by the *a priori* profile shape. We examine how differences between the shape of the simulated and *a priori* profile can impact the interpretation of satellite retrievals by performing an adjoint-based 4D-Var assimilation of synthetic $NO_2$ observations for constraining $NO_x$ emissions. We use the GEOS-Chem Adjoint model to perform assimilations using a variety of AMFs to examine how *a posteriori* emission estimates are affected if the AMF is calculated using an *a priori* shape factor that is inconsistent with the simulated profile. In these tests, an inconsistent *a priori* shape factor increased root mean square errors in *a posteriori* emissions estimates by up to 380% for realistic conditions over polluted regions. As the difference between the simulated profile shape and the *a priori* profile shape increases, so do the corresponding assimilated emission errors. This reveals the importance of using simulated profile information for AMF calculations when comparing that simulated output to satellite retrieved columns.

## 1. Introduction

Satellite observations provide a wealth of information on the abundance of trace gases in the troposphere (Fishman et al., 2008). The next generation of satellite instruments, including the

upcoming geostationary constellation of TEMPO (Chance et al., 2013; Zoogman et al., 2017), Sentinal-4 (Bazalgette Courrèges-Lacoste et al., 2011; Ingmann et al., 2012), and GEMS (Bak et al., 2013; Kim, 2012), will provide information on $NO_2$ and other air quality relevant pollutants on unprecedented spatial and temporal scales. Insight into processes that affect atmospheric composition, including emissions (Streets et al., 2013), lifetimes (Fioletov et al., 2015; de Foy et al., 2015; Laughner and Cohen, 2019), and deposition (Geddes and Martin, 2017; Kharol et al., 2018) can be gained by interpreting this information with atmospheric chemistry models.

There are three main stages in retrieving trace gas abundances from ultraviolet and visible solar backscatter radiance measurements: calculating a light-path "slant column" by fitting observed spectra to known spectral signatures of trace gases, removing the stratospheric portion of the column, and converting the slant column to a vertical column density using an air mass factor (AMF). AMFs are calculated using a radiative transfer model and are a function of viewing geometry, surface reflectance, clouds, and radiative transfer properties of the atmosphere. AMF calculations also require an *a priori* estimate of the trace gas vertical profile and are sensitive to the profile shape (Eskes and Boersma, 2003; Palmer et al., 2001). Uncertainties in AMF calculations are the dominant source of uncertainty in satellite $NO_2$ retrievals over polluted regions (Boersma et al., 2007; Martin et al., 2002) largely due to sensitivity to surface reflectance, clouds, aerosols, and *a priori* profile information (Lorente et al., 2017).

Boersma et al. (2016) highlighted the issue of representativeness errors in comparing model simulated values with UV-Vis satellite-retrieved columns. Vertical representativeness errors arise from the satellite's altitude-dependent sensitivity due to atmospheric scattering and can degrade the quality of model-measurement comparisons beyond errors that arise from either modeling or measurements alone. A consistent accounting of the altitude-dependent sensitivity is necessary to limit these errors.

Two common methods are used to account for vertical representativeness. In one method, observed slant columns are converted to vertical columns using an air mass factor calculated with scattering weights to represent instrument vertical sensitivity and shape factors to represent the vertical profile (Palmer et al., 2001). Another commonly used method employs an AMF provided with the retrieval to convert slant columns to vertical columns, and then applies an averaging kernel to the simulated profile to resample the simulated profile in a manner that

mimics the satellite vertical sensitivity (Eskes and Boersma, 2003). In this method both the averaging kernel and the retrieval AMF are calculated using an *a priori* $NO_2$ profile that may have a different shape than the simulated profile, which may introduce errors in the observation-simulation comparison (Zhu et al., 2016).

A common application of comparisons between satellite observed columns and model simulations is to constrain $NO_x$ emissions (e.g. Ding et al., 2018; Ghude et al., 2013; Lamsal et al., 2011; Martin et al., 2003; Vinken et al., 2014). One such approach is the use of four-dimensional variational (4D-Var) data assimilation, which seeks to minimize a cost function that accounts for the difference between simulated and retrieved values. As the cost function is a difference between observed and simulated $NO_2$ columns, it is susceptible to vertical representativeness errors resulting from inconsistent *a priori* vertical profile information. ~~Studies have shown that differences in retrieval processes between different $NO_2$ column products, including differences in *a priori* profile shape, can propagate into errors of up to 50% in adjoint inversions of $NO_x$ emissions (Qu et al., 2017). Studies have shown that shape factor errors can impact emission estimates from other inversion methods as well (Laughner et al., 2016).~~

In this work we examine how *a priori* profile assumptions impact satellite-model comparisons and use the GEOS-Chem adjoint as a case study to assess how this impact can affect the interpretation of satellite observations. Section 2 provides the mathematical framework for AMF calculations and satellite-model comparisons. Section 3 describes the adjoint model and synthetic observations for the case study. Section 4 discusses the results.

## 2. Mathematical frameworks

### 2.1 AMFs and averaging kernels

The air mass factor translates the line-of-sight slant column abundances ($\Omega_s$) retrieved from satellite observed radiances into vertical column abundances ($\Omega_v$). An air mass factor is the ratio of $\Omega_s$ to $\Omega_v$ and depends on the atmospheric path as determined by geometry, $NO_2$ vertical profile ($\boldsymbol{n}$), surface reflectance, and radiative transfer properties of the atmosphere. Here we use $M(\boldsymbol{n})$ to represent an air mass factor derived using the vertical number density profile $\boldsymbol{n}$:

$$M(\boldsymbol{n}) = \frac{\Omega_s}{\Omega_v} \quad (1)$$

In the method described by Palmer et al. (2001), a radiative transfer model is used calculate scattering weights $w$(z) (also known as box air mass factors) which characterize the sensitivity of backscattered radiance $I_B$ to the abundance of a trace gas at altitude $z$:

$$w(z) = -\frac{1}{M_g}\frac{\alpha_{a,z}}{\alpha_{eff}}\frac{\partial \ln (I_B)}{\partial \tau} \quad (2)$$

where $\alpha_{a,z}$ is the temperature-dependent absorption cross section ($m^2$ $molec^{-1}$), $\alpha_{eff}$ is the effective (weighted average) absorption cross section ($m^2$ $molec^{-1}$) and $\partial\tau$ is the incremental trace gas optical depth. $M_G$ represents a geometric path correction accounting for the satellite viewing geometry:

$$M_G = \sec\theta_o + \sec\theta \quad (3)$$

where $\theta$ is the solar zenith angle and $\theta_o$ is the satellite viewing angle. This information is then combined with an *a priori* $NO_2$ shape factor (i.e. normalized vertical profile)

$$\boldsymbol{S}(z) = \frac{\boldsymbol{n}(z)}{\Omega_v} \quad (4)$$

typically calculated with an atmospheric chemistry model to provide an air mass factor via:

$$M(\boldsymbol{n}) = \int_0^{tropopause} \boldsymbol{w}(z)\boldsymbol{S}(z)dz \quad (5)$$

where ***S(z)*** is calculated using vertical profile ***n(z)***. An attribute of the formulation of Palmer et al. (2001) is the independence of atmospheric radiative transfer properties ***w(z)*** and the vertical trace gas profile ***S(z)***. The AMF definition in Equation (1) combined with Eq. (4) indicates that a slant column can be calculated from a known vertical profile via:

$$\Omega_s = \int_0^{tropopause} \boldsymbol{w}(z)\boldsymbol{n}(z)dz \quad (6)$$

In an alternative formulation, the air mass factor is represented as part of an averaging kernel. As formulated by Rodgers and Connor (2003), the averaging kernel (***A***) provides the information needed to relate the retrieved quantity $\hat{\boldsymbol{n}}$ to the true atmospheric profile ***n***:

$$\hat{\boldsymbol{n}} - \boldsymbol{n_a} = \boldsymbol{A}(\boldsymbol{n} - \boldsymbol{n_a}) \quad (7)$$

where $\boldsymbol{n_a}$ is an assumed *a priori* profile of number density. The elements of the column

averaging kernel are related to the scattering weights by:

$$\boldsymbol{A}(z) = \frac{\boldsymbol{w}(z)}{M(\boldsymbol{n_a})} \quad (8)$$

where $M(\boldsymbol{n_a})$ is an air mass factor calculated using *a priori* vertical profile information. It is important to note that unlike scattering weights, averaging kernels depend on the *a priori* assumed vertical profile shape.

~~It is possible to decouple the radiative transfer information from the assumed vertical profile information in an averaging kernel by converting the supplied averaging kernels to scattering weights via:~~

$$~~w(z) = \frac{A(z)M(n_a)}{M_G}~~ \quad ~~(9)~~$$

A lexicon is given in Table 1 as notation used to describe these treatments has varied across the literature. We choose *M* for air mass factor as a single letter is clearer in equations, *w* for scattering weights to maintain the original formulation of Palmer et al. (2001), *n* for number density following IUPAC recommendations, and Ω for column densities as is common in radiative transfer literature.

Figure 1 shows examples of typical shape factor, scattering weight, and averaging kernel profiles for a range of atmospheric conditions. $NO_2$ shape factors have significant variability; Shape factors peak near the surface in urban regions due to local pollution sources, but peak in the upper troposphere in more remote regions due to lightning. The shape of a scattering weight profile depends strongly on surface reflectance and cloud conditions. Sensitivity in the lower troposphere increases over reflective surfaces. Clouds increase sensitivity above due to their reflectance but shield the satellite from observing the atmosphere below. Averaging kernels have similarities with scattering weights but depend on both the shape of the prior and the satellite sensitivity. As AMF calculations are a convolution of the shape factor and the scattering weight profiles, these shapes affect $NO_2$ retrievals. For these examples, the AMF for a clear sky observation with surface reflectance of 0.01 can range from 0.7 in an urban region to 1.7 in a remote region. This large difference demonstrates the importance of the assumed profile shape to the retrieval process.

## 2.2 Comparing satellite observations to simulated values

The following section expresses mathematically how satellite-model comparisons are made using various *a priori* profiles.

### 2.2.1 Using scattering weights

Following Palmer et al. (2001), a retrieved vertical column ($\widehat{\Omega}_{v,o}$) is estimated using an observed slant column $\Omega_{s,o}$ and a simulation-based air mass factor $M(\boldsymbol{n_m})$, which can be calculated with Eq. (5) using the model-simulated $NO_2$ profile ($\boldsymbol{n_m}$):

$$\widehat{\Omega}_{v,o} = \frac{\Omega_{s,o}}{M(\boldsymbol{n_m})} \quad (9\text{~~10~~})$$

The difference $\Delta_m$ between the estimated retrieved column and the model-simulated vertical column ($\Omega_{v,m}$) is:

$$\Delta_m = \Omega_{v,m} - \widehat{\Omega}_{v,o} \quad (10\text{~~1~~})$$

$$\Delta_m = \left(\sum_0^{tropopause} n_m\right) - \frac{\Omega_{s,o}}{M(\boldsymbol{n_m})} \quad (11\text{~~2~~})$$

Equation (11~~2~~) describes how this comparison is used in practice. However, we can rearrange this expression in terms of model ($\Omega_{s,m}$) and observed ($\Omega_{s,o}$) slant columns using the definition of air mass factor:

$$\Delta_m = \frac{\Omega_{s,m}}{M(\boldsymbol{n_m})} - \frac{\Omega_{s,o}}{M(\boldsymbol{n_m})} \quad (12\text{~~3~~})$$

$$\Delta_m = \frac{1}{M(\boldsymbol{n_m})}\left(\Omega_{s,m} - \Omega_{s,o}\right) \quad (13\text{~~4~~})$$

### 2.2.2 Using averaging kernels

Comparison of simulated and retrieved columns using the averaging kernel is described by Eskes and Boersma (2003) and in the retrieval documentation in Boersma et al. (2011). The averaging kernel is applied to the simulated profile in order to sample the simulated column in a manner that reflects the retrieval sensitivity:

$$\widehat{\Omega}_{v,m} = \sum_0^{tropopause} \boldsymbol{A n_m} \quad (14\text{~~5~~})$$

The resampled simulated column is then compared to the retrieved vertical column ($\Omega_{v,o}$) using the *a priori*-based air mass factor $M(\boldsymbol{n_a})$ supplied with the retrieval dataset:

$$\Delta_a = \widehat{\Omega}_{v,m} - \Omega_{v,o} \quad (15)$$

$$\Delta_a = \left(\sum_{i=0}^{tropopause} \boldsymbol{A_i n_{m,i}}\right) \sum_{0}^{tropopause} \boldsymbol{An_m} - \frac{\Omega_{s,o}}{M(\boldsymbol{n_a})} \quad (16)$$

Equation (16) describes how this method is used in practice. To facilitate the comparison with Eq. (13), Eq. (16) can be rewritten using an alternative formulation relating averaging kernels to scattering weights:

$$\Delta_a = \left(\sum_{i=0}^{tropopause} \frac{\boldsymbol{w_i n_{m,i}}}{M(\boldsymbol{n_a})}\right) \sum_{0}^{tropopause} \frac{\boldsymbol{wn_m}}{M(\boldsymbol{n_a})} - \frac{\Omega_{s,o}}{M(\boldsymbol{n_a})} \quad (17)$$

$$\Delta_a = \frac{1}{M(\boldsymbol{n_a})}\left(\Omega_{s,m} - \Omega_{s,o}\right) \quad (18)$$

By comparing Eq. (13) to Eq. (18), it is evident that the underlying difference between the two approaches is the choice of *a priori* profile information used to calculate the AMF, as the averaging kernel method is not independent of *a priori* profile assumptions. This bias could be addressed by replacing the *a priori* -based AMF in Eq. (18) with a simulation-based AMF using the following relationship (Boersma et al., 2016; Lamsal et al., 2010):

$$M(\boldsymbol{n_m}) = M(\boldsymbol{n_a}) \frac{\sum \boldsymbol{An_m}}{\sum \boldsymbol{n_m}} \quad (19)$$

It should be noted that both the averaging kernel and scattering weight methods are equivalent for comparisons that examine ratios of retrieved and modeled columns:

$$r_m = \frac{\widehat{\Omega_{v,o}}}{\Omega_{v,m}} = \frac{\Omega_{s,o} \Big/ M(\boldsymbol{n_m})}{\sum \boldsymbol{n_m}} = \frac{\Omega_{s,o}}{\sum \boldsymbol{n_m}} \frac{\sum \boldsymbol{n_m}}{\sum \boldsymbol{wn_m}} = \frac{\Omega_{s,o}}{\sum \boldsymbol{wn_m}} \quad (20)$$

$$r_a = \frac{\Omega_{v,o}}{\widehat{\Omega_{v,m}}} = \frac{\Omega_{s,o} \Big/ M(\boldsymbol{n_a})}{\sum \boldsymbol{An_m}} = \frac{\Omega_{s,o} \Big/ M(\boldsymbol{n_a})}{\sum \boldsymbol{wn_m}/M(\boldsymbol{n_a})} = \frac{\Omega_{s,o}}{\sum \boldsymbol{wn_m}} \quad (21)$$

For ratios, both methods are dependent on geophysical assumptions used to calculate scattering weights but are independent of *a priori* profile information. Lastly, some studies (e.g., Buscela et

al., 2013; Qu et al., 2017) may directly assimilate slant column densities rather than vertical column densities using

$$\Delta_{s,a} = \widehat{\Omega}_{s,m} - \Omega_{s,o} \quad (22)$$

$$= \left(\sum_{i=0}^{tropopause} \boldsymbol{w_i n_{m,i}}\right) - \Omega_{s,o} \quad (23)$$

This approach is also still dependent upon the scattering weights but not upon external *a priori* profile information. Overall, the choice of approach may be influenced by whether or not scattering weights are available from either the $NO_2$ retrieval product or radiative transfer calculations applied to the model. In contrast, use of Eq. (11) or (16) are applicable when these are not explicitly available or provided.

## 3. Tools and Methodology

### 3.1 GEOS-Chem and its adjoint

The GEOS-Chem chemical transport model (www.geos-chem.org) is used to create synthetic $NO_2$ observations and for their analysis. The GEOS-Chem version used here is version 35j of the GEOS-Chem Adjoint model. GEOS-Chem includes a detailed oxidant-aerosol chemical mechanism (Bey et al., 2001; Park et al., 2004) and uses assimilated meteorological fields from the Goddard Earth Observation System (GEOS-5), with 47 vertical levels up to 0.01 hPa and a horizontal resolution of 4°x5°. Global anthropogenic $NO_x$ emissions are provided by the Emission Database for Global Atmospheric Research (EDGAR) inventory (Olivier et al., 2005) with regional overwrites over North America (EPA/NEI99), Europe (EMEP), Canada (CAC), Mexico (BRAVO, (Kuhns et al., 2005)), and East Asia (Streets et al., 2006). Other $NO_x$ sources include biomass burning (GFED2 (Van der Werf et al., 2010)), lightning (Murray et al., 2012), and soils (Wang et al., 1998). This model has been used previously to constrain $NO_x$ emissions (Cooper et al., 2017; Henze et al., 2009; Qu et al., 2017, 2019; Xu et al., 2013; Zhang et al., 2016).

The GEOS-Chem adjoint (Henze et al., 2007, 2009) is used here to perform a 4D-Var data assimilation. The adjoint seeks to iteratively minimize a cost function generally defined by the difference between satellite retrieved and simulated columns ($\Delta$, from either Eq. (11~~2~~) if

using a simulation-based air mass factor or Eq. (167) if using the retrieval *a priori*-based air mass factor):

$$J = \frac{1}{2}\boldsymbol{\Delta}^{\boldsymbol{T}}\boldsymbol{S_o^{-1}}\boldsymbol{\Delta} + \frac{1}{2}\gamma_R(\boldsymbol{E}-\boldsymbol{E_a})^{\boldsymbol{T}}\boldsymbol{S_E^{-1}}(\boldsymbol{E}-\boldsymbol{E_a}) \quad (242)$$

where $E$ and $E_a$ are the *a posteriori* and *a priori* emissions, $\boldsymbol{S_o}$ and $\boldsymbol{S_E}$ are the retrieval and *a priori* emission error covariance matrices, and $\gamma_R$ is a regularization parameter that allows for weighting the cost function towards the retrieved columns or *a priori* emissions. Tests performed here required 20-30 iterations to minimize the cost function.

### 3.2 Experiment Outline

In this study we perform 4D-Var data assimilation experiments to infer surface $NO_x$ emissions using synthetic $NO_2$ observations. We use synthetic observations built from ~~a~~ known emission inventories~~y~~ to provide a "truth" that can be used to evaluate the inversion results. To demonstrate how *a priori* profile information can propagate in an assimilation, we use either the model profile ($\Delta_m$, Eq. (112)) or an *a priori* profile ($\Delta_a$, Eq. (167)) in the cost function. ~~For these tests, we use one observation per hour per 4°x5° grid box for a period of two weeks in July 2010.~~ A one-week spin-up window at the start of each adjoint iteration is used to allow $NO_x$ to reach steady state. Observation error covariances $S_o$ are described as a relative error of 30% of the slant column density, plus an absolute error of $10^{15}$ molecules $cm^{-2}$, which is representative of typical satellite retrieved $NO_2$ column uncertainties (Boersma et al., 2007; Martin et al., 2002). We omit the *a priori* emissions constraint in the cost function (i.e. set $\gamma_R$=0) to isolate the impact of the observations.

#### 3.2.1 Synthetic observations

Synthetic observations ($Obs_5$) are created using a GEOS-Chem simulation where random Gaussian noise with a standard deviation of 5% is added to the anthropogenic $NO_x$ emissions. Additional tests using observations where noise with a standard deviation of 30% is added ($Obs_{30}$) are also used. No additional noise is added to the individual observations to isolate the impact of AMF errors against~~-~~ additional sources of uncertainty. Figure 2 shows the standard (*a priori*) anthropogenic $NO_x$ emissions and the changes used to create the "true" emissions for the synthetic observations.

For these tests, we use one observation per hour per 4°x5° grid box for a period of two weeks in July 2010. Observations consist of synthetic slant columns ($\Omega_{s,o}$) created by applying scattering weights to the synthetic vertical profiles using Eq. (6). ~~To represent typical conditions, average scattering weight profiles for each grid box are found by averaging scattering weights for OMI observations during July 2010. OMI s~~Scattering weights are calculated using the LIDORT radiative transfer model (Spurr, 2002) by providing LIDORT with the observation ~~geometry~~ conditions of ~~the~~ OMI observations during July 2010, which are used to represent typical viewing ~~geometries~~conditions of low earth orbit satellite observations, and aerosol profiles from the GEOS-Chem base simulation. To represent typical conditions, ~~average~~these representative scattering weight profiles for each grid box are used to produce the synthetic slant columns. Tests performed for all 4°x5° grid boxes used here indicate that the mean relative difference between an air mass factor calculated using an average scattering weight profile and the average of air mass factors using observation-specific scattering weight profiles is less than 4%.

### 3.2.2 Shape Factors

To test the impact of *a priori* profile information, ~~five~~ seven different tests are performed using ~~five~~ seven different $NO_2$ profile shapes for AMF calculations:

- Case $SF_M$ : The GEOS-Chem model simulated profile ($\boldsymbol{n_m}$), updated at each iteration of the adjoint run
- Case $SF_{prior}$: The *a priori* GEOS-Chem simulated profile, without updating.
- Case $SF_{n30}$: An *a priori* profile created by a GEOS-Chem simulation where global anthropogenic $NO_x$ emissions were perturbed with random Gaussian noise with a standard deviation of 30%. In cases where this results in negative emissions, a value of zero is used.
- Case $SF_{diffem}$: An *a priori* profile created by a GEOS-Chem simulation where regional emission overwrites are turned off.
- Case $SF_{finer}$: An *a priori* profile created by a GEOS-Chem simulation run at finer (2°x2.5°) resolution.
- Case $SF_{trop}$: An *a priori* profile that assumes the $NO_2$ profile shape is uniform from the surface to the tropopause (~200 hPa).

- Case $SF_{BL}$: An *a priori* profile that assumes the $NO_2$ profile shape is uniform from the surface to the boundary layer (~800 hPa).

An advantage of using scattering weights and the simulated shape factor in a 4D-Var framework is that it allows for the shape factor, and thus the AMF, to be updated at each iteration. When *a priori* profiles from an external source are used it is not possible for them to update during the inversion. The $SF_M$ and $SF_{prior}$ cases ~~will~~ test the impact that iterative updates to the AMF ~~will~~ have on *a posteriori* estimates. The additional cases ~~will~~ test for the impact of using an averaging kernel based on *a priori* profile assumptions that are inconsistent with the model. In practice, averaging kernels and *a priori* profiles included in retrieval data sets are generally derived from chemical transport models that have different physical processes, emissions, or spatial resolutions. The $SF_{n30}$ and $SF_{diffem}$ tests ~~is~~ are representative of ~~an~~ inversions that use~~s~~ *a priori* profile information from a different chemical transport model with similar resolution but different emissions. The $SF_{finer}$ test represents an inversion that uses *a priori* profiles ~~with~~ from a chemical transport model with a different horizontal resolution. The $SF_{BL}$ and $SF_{trop}$ tests do not represent any modern retrieval algorithms, but are used as~~are~~ extreme examples of ~~using~~ an *a priori* that~~based on a coarser resolution model, as both tests~~ assume~~s~~ no spatial variability. The $SF_{BL}$ profile is representative of polluted regions as indicated by the typical urban profile in Fig. 1, while the $SF_{trop}$ profile is representative of a typical rural profile. Table 2 provides global mean AMFs for these test cases, which range from 1.3-2.1, and the resulting global mean observed vertical columns, which range from ~~8.5~~0.9-1~~4.~~.5 x $10^{15}$ molec/$cm^2$. Global mean 'observed' vertical columns are 33% higher for $SF_{2x25}$ than for $SF_M$, and up to 66% higher for $SF_{BL}$. Global mean 'observed' vertical columns for $SF_{n30}$ and $SF_{diffem}$ are similar to $SF_M$, although individual observations may differ by up to 18% for $SF_{n30}$ and 28% for $SF_{diffem}$.

## 4. Results

Figure 3 shows root mean square errors (RMSE) for the *a posteriori* emissions estimated by the 4D-Var assimilations of $Obs_5$ synthetic observations ~~tests~~. All tests successfully reduce the *a priori* emission error by an order of magnitude or more. The $SF_M$ has the lowest RMSE indicating that it can best estimate the "true" emissions. The next lowest RMSE is for the $SF_{prior}$

test, which uses the same initial model shape factor but does not update during the adjoint iterations, followed by the $SF_{finer2x25}$, $SF_{diffem}$, $SF_{n30}$, $SF_{trop}$, and $SF_{BL}$ tests.

Figure 4 shows maps of the difference in RMSE between the $SF_M$ test and the other tests for $Obs_5$ observations. The $SF_M$ test has a lower RMSE than the other tests in 65-72% grid boxes where the difference is nonzero. Again, the $SF_{prior}$ test is closest to the $SF_M$ test with a ~~mean absolute~~root mean square difference of ~~6x10$^6$~~ 2.9x10$^7$ molec/cm$^2$/s, followed by ~~$SF_{n30}$ (7x10$^6$ molec/cm$^2$/s),~~ $SF_{finer2x25}$ (3.6x10$^7$ molec/cm$^2$/s), $SF_{n30}$ (3.8x10$^7$ molec/cm$^2$/s), $SF_{diffem}$ (4.0x10$^7$ molec/cm$^2$/s), $SF_{trop}$, (7.8~~13~~x10$^7$~~6~~ molec/cm$^2$/s), and $SF_{BL}$ (9.0~~16~~x10$^7$~~6~~ molec/cm$^2$/s).

Table 3~~2~~ summarizes additional error statistics focused on grid boxes with significant emission sources. Errors in *a posteriori* emission estimates are correlated with the "true" emissions in the $SF_{trop}$ and $SF_{n30}$ tests, and weakly correlated in the $SF_{BL}$, $SF_{prior}$, and $SF_{diffem}$ tests, indicating that these tests are not well constraining the emissions. Differences between tests are more significant over polluted regions where AMF errors are more influential; For example, in the regions with the highest $NO_x$ emissions, RMSE values indicate $SF_M$ outperforms $SF_{n30}$ by 30% and $SF_{trop}$ by >80%. Another sign of adjoint inversion quality is a low variance in errors. While the posterior error is reduced relative to the *a priori* error in all tests, error standard deviations are 30% higher for $SF_{n30}$ and 90% higher for $SF_{trop}$ compared to $SF_M$. The global maximum error for the $SF_{trop}$ test is 30% higher than for the $SF_M$ test. All metrics indicate that the $SF_M$ test best represents the "true" emissions.

Tests using $Obs_{30}$ observations and the $SF_M$ and $SF_{trop}$ shape factors were also performed. Despite the difference between *a priori* observed vertical columns using these shape factors as indicated by Table 2, these assimilations produced similar *a posteriori* results, with RMSE of 2.9x10$^8$ molec/cm$^2$/s for $SF_M$ and 2.8x10$^8$ molec/cm$^2$/s for $SF_{trop}$.

## 5. Discussion & Conclusions

Accounting for the vertical profile dependence of satellite observations is essential to accurately interpret those observations. This work examines how the choice of shape factor affects differences between simulated and satellite-retrieved quantities in a ~~4D-Var~~data assimilation framework. Examination of the mathematical frameworks behind two common methods for comparing simulated and retrieved columns highlights how the method introduced by Palmer et al. (2001) facilitates separation of observation sensitivity (scattering weights) from

the profile shape (shape factor) enabling the model-retrieval comparison to be independent of *a priori* profile assumptions.

In these case studies, vertical representativeness errors were best reduced by using a shape factor that was consistent with the model simulation. This was especially true in polluted regions where the AMF errors dominate observation uncertainties, as deviations between the tests were largest in these regions. The further the shape factor deviated from the model state the larger the inversion errors became, as indicated by Fig. 5. The $SF_{finer2x25}$ test indicates that using a higher resolution model to generate *a priori* profiles does not provide an advantage in simulation-observation comparisons, as consistency between the simulation profile and the ~~*AMF*~~*a priori* shape factor is of greater importance. Comparing the $SF_M$ and $SF_{prior}$ tests shows that allowing for the shape factor to update during the iterative adjoint process further reduces the RMSE by 10%. However, even without allowing for shape factor updates, using a shape factor that is consistent with the initial model state produces a more accurate inversion result than using other assumed profile shapes.

The case study presented here demonstrates that the shape factor source can have a strong influence on adjoint inversion results. However, the magnitude of this influence can vary. ~~Additional~~ Inversion tests performed using synthetic observations ~~built~~ based on~~using~~ random ~~15% or~~ 30% perturbations to emissions ~~(rather than the 5% perturbation used here)~~ were insensitive to the AMF, despite large differences in *a priori* vertical column densities. In these tests, the ~~adjoint~~ cost function was more sensitive to the larger difference between the observed and simulated slant columns (i.e. $\Omega_{s,m}$ - ~~$\Omega_o$~~ $\Omega_{s,o}$ in Eq. (13) and (1~~9~~8)) than to the AMF. This indicates that while the ~~adjoint~~ cost function is mathematically dependent on the AMF, the inversion is less sensitive to vertical representativeness errors in cases where emissions are poorly constrained, as is the case in recent adjoint inversion studies (e.g. Qu et al., 2017).~~;~~ However~~Conversely~~, choice of AMF will become increasingly important to adjoint inversions as emission inventories improve. Furthermore, omitting the *a priori* emissions constraint in the cost function and omitting noise in the observations in these tests to isolate the impact of the AMF effectively assumes poorly constrained *a priori* emissions and ideal observations. In practice, cost function sensitivity to AMF choice may be buffered when *a priori* emissions uncertainties and observational noise are considered.

As it is beneficial for a consistent shape factor to be used when comparing satellite

retrieved values to model simulated results, it will be useful for data products to provide the information required for this method to the user community. This is most straightforward when scattering weights (rather than averaging kernels) are provided alongside retrieved column data, as scattering weights and shape factors are independently calculated, however ~~averaging kernels can be converted to scattering weights if the *a priori* profiles used are included in the dataset~~simulation-based air mass factors can be calculated using the averaging kernel and *a priori*-based air mass factor via Eq. 19.

In summary, when comparing a model simulation to a satellite retrieved $NO_2$ column in a ~~4D-Var~~data assimilation environment utilizing column differences, calculating the AMF using the simulated shape factor allows for better accuracy in inversion results. This demonstration can provide general guidance for other methods of interpreting satellite observations with models, as using the simulated shape factor assures consistency in the vertical representativeness between model and retrieval.

## 6~~7~~. Author Contributions

MJC and RVM designed the overall study. MJC designed and carried out the case studies and their analysis. All co-authors provided guidance in analyzing results. MJC prepared the manuscript with contributions from all co-authors.

## 7~~8~~. Competing interests

The authors declare that they have no conflict of interest.


## 8~~9~~. Acknowledgements

This work was supported by the Canadian Space Agency. DH acknowledges support from NASA NNX17AF63G.


## 9~~10~~. Data Availability

The GEOS-Chem chemical transport model and its adjoint are available at www.geos-chem.org (last access: 20 August 2017). OMI $NO_2$ data used in this study is available from the NASA Goddard Earth Sciences Data and Information Services Center (https://disc.sci.gsfc.nasa.gov; last access: 14 March 2019). AMF code (Spurr, 2002; Martin et al., 2002) used to calculate scattering weights and air mass factors is available at http://fizz.phys.dal.ca/~atmos (last access: 19 June 2017).

## 11. References


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

Bazalgette Courrèges-Lacoste, G., Ahlers, B., Guldimann, B., Short, A., Veihelmann, B. and Stark, H.: The Sentinel-4/UVN instrument on-board MTG-S, in EUMETSAT Meteorological Satellite Conference, Oslo, Norway., 2011.

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: Model description and evaluation, J. Geophys. Res. Atmos., 106(D19), 23073–23095, 2001.

Boersma, K., Braak, R. and van der A, R. J.: Dutch OMI NO2 (DOMINO) data product v2. 0, Tropospheric Emissions Monitoring Internet Service on-line documentation, [online] Available from: http://www.temis.nl/docs/OMI_NO2_HE5_2.0_2011.pdf, 2011.

Boersma, K. F., Eskes, H. J., Veefkind, J. P., Brinksma, E. J., van der A, R. J., Sneep, M., van den Oord, G. H. J., Levelt, P. F., Stammes, P., Gleason, J. F. and Bucsela, E. J.: Near-real time retrieval of tropospheric NO2 from OMI, Atmos. Chem. Phys., 7(8), 2103–2118, doi:10.5194/acp-7-2103-2007, 2007.

Boersma, K. F., Vinken, G. C. M. and Eskes, H. J.: Representativeness errors in comparing chemistry transport and chemistry climate models with satellite UV–Vis tropospheric column retrievals, Geosci. Model Dev, 9, 875–898, doi:10.5194/gmd-9-875-2016, 2016.

Chance, K., Liu, X., Suleiman, R. M., Flittner, D. E., Al-Saadi, J. and Janz, S. J.: Tropospheric emissions: monitoring of pollution (TEMPO), edited by J. J. Butler, X. (Jack) Xiong, and X. Gu, p. 88660D, International Society for Optics and Photonics., 2013.

Cooper, M., Martin, R. V., Padmanabhan, A. and Henze, D. K.: Comparing mass balance and adjoint methods for inverse modeling of nitrogen dioxide columns for global nitrogen oxide emissions, J. Geophys. Res. Atmos., doi:10.1002/2016JD025985, 2017.

Ding, J., van der A, R. J., Mijling, B., Jalkanen, J.-P., Johansson, L. and Levelt, P. F.: Maritime NOx Emissions Over Chinese Seas Derived From Satellite Observations, Geophys. Res. Lett., 45(4), 2031–2037, doi:10.1002/2017GL076788, 2018.

Eskes, H. J. and Boersma, K. F.: Averaging kernels for DOAS total-column satellite retrievals, Atmos. Chem. Phys., 3(5), 1285–1291, 2003.

Fioletov, V. E., McLinden, C. A., Krotkov, N. and Li, C.: Lifetimes and emissions of SO2 from point sources estimated from OMI, Geophys. Res. Lett., 42(6), 1969–1976, doi:10.1002/2015GL063148, 2015.

Fishman, J., Al-Saadi, J. A., Creilson, J. K., Bowman, K. W., Burrows, J. P., Richter, A., Chance, K. V., Edwards, D. P., Martin, R. V., Morris, G. A., Pierce, R. B., Ziemke, J. R., Schaack, T. K., Thompson, A. M., Fishman, J., Al-Saadi, J. A., Creilson, J. K., Bowman, K. W., Burrows, J. P., Richter, A., Chance, K. V., Edwards, D. P., Martin, R. V., Morris, G. A., Pierce, R. B., Ziemke, J. R., Schaack, T. K. and Thompson, A. M.: Remote Sensing of Tropospheric Pollution from Space, Bull. Am. Meteorol. Soc., 89(6), 805–821, doi:10.1175/2008BAMS2526.1, 2008.

de Foy, B., Lu, Z., Streets, D. G., Lamsal, L. N. and Duncan, B. N.: Estimates of power plant NO x emissions and lifetimes from OMI NO 2 satellite retrievals, Atmos. Environ., 116, 1–11, 2015.

Geddes, J. A. and Martin, R. V.: Global deposition of total reactive nitrogen oxides from 1996 to 2014 constrained with satellite observations of NO2 columns, Atmos. Chem. Phys., 17(16), 10071–10091, doi:10.5194/acp-17-10071-2017, 2017.

Ghude, S. D., Pfister, G. G., Jena, C., van der A, R. J., Emmons, L. K. and Kumar, R.: Satellite constraints of nitrogen oxide (NOx) emissions from India based on OMI observations and WRF-Chem simulations, Geophys. Res. Lett., 40(2), 423–428, 2013.

Henze, D. K., Hakami, A. and Seinfeld, J. H.: Development of the adjoint of GEOS-Chem, Atmos. Chem. Phys., 7(9), 2413–2433, doi:10.5194/acp-7-2413-2007, 2007.

Henze, D. K., Seinfeld, J. H. and Shindell, D. T.: Inverse modeling and mapping US air quality influences of inorganic PM 2.5 precursor emissions using the adjoint of GEOS-Chem, Atmos. Chem. Phys., 9(16), 5877–5903, 2009.

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. Environ., 120, 58–69, doi:10.1016/j.rse.2012.01.023, 2012.

Kharol, S. K., Shephard, M. W., McLinden, C. A., Zhang, L., Sioris, C. E., O'Brien, J. M., Vet, R., Cady-Pereira, K. E., Hare, E., Siemons, J. and Krotkov, N. A.: Dry Deposition of Reactive Nitrogen From Satellite Observations of Ammonia and Nitrogen Dioxide Over North America,

Geophys. Res. Lett., 45(2), 1157–1166, doi:10.1002/2017GL075832, 2018.
Kim, J.: GEMS (Geostationary Environment Monitoring Spectrometer) onboard the GeoKOMPSAT to monitor air quality in high temporal and spatial resolution over Asia-Pacific Region, in EGU General Assembly Conference Abstracts, vol. 14, p. 4051., 2012.
Kuhns, H., Knipping, E. M. and Vukovich, J. M.: Development of a United States–Mexico emissions inventory for the big bend regional aerosol and visibility observational (BRAVO) study, J. Air Waste Manage. Assoc., 55(5), 677–692, 2005.
Lamsal, L. N., Martin, R. V., Van Donkelaar, A., Celarier, E. A., Bucsela, E. J., Boersma, K. F., Dirksen, R., Luo, C. and Wang, Y.: Indirect validation of tropospheric nitrogen dioxide retrieved from the OMI satellite instrument: Insight into the seasonal variation of nitrogen oxides at northern midlatitudes, J. Geophys. Res. Atmos., doi:10.1029/2009JD013351, 2010.
Lamsal, L. N., Martin, R. V, Padmanabhan, A., van Donkelaar, A., Zhang, Q., Sioris, C. E., Chance, K., Kurosu, T. P. and Newchurch, M. J.: Application of satellite observations for timely updates to global anthropogenic NOx emission inventories, Geophys. Res. Lett., 38(5), 2011.
Laughner, J. L. and Cohen, R. C.: Direct observation of changing NOx lifetime in North American cities, Science (80-. )., 366(6466), 723–727, doi:10.1126/science.aax6832, 2019.
Laughner, J. L., Zare, A. and Cohen, R. C.: Effects of daily meteorology on the interpretation of space-based remote sensing of NO 2, Atmos. Chem. Phys, 16, 15247–15264, doi:10.5194/acp-16-15247-2016, 2016.
Lorente, A., Folkert Boersma, K., Yu, H., Dörner, S., Hilboll, A., Richter, A., Liu, M., Lamsal, L. N., Barkley, M., De Smedt, I., Van Roozendael, M., Wang, Y., Wagner, T., Beirle, S., Lin, J.-T., Krotkov, N., Stammes, P., Wang, P., Eskes, H. J. and Krol, M.: Structural uncertainty in air mass factor calculation for NO2 and HCHO satellite retrievals, Atmos. Meas. Tech., 10(3), 759–782, doi:10.5194/amt-10-759-2017, 2017.
Martin, R. V., Jacob, D. J., Chance, K., Kurosu, T. P., Palmer, P. I. and Evans, M. J.: Global inventory of nitrogen oxide emissions constrained by space-based observations of NO2 columns, J. Geophys. Res., 108(D17), 4537, doi:10.1029/2003JD003453, 2003.
Martin, R. V, Chance, K., Jacob, D. J., Kurosu, T. P., Spurr, R. J. D., Bucsela, E., Gleason, J. F., Palmer, P. I., Bey, I. and Fiore, A. M.: An improved retrieval of tropospheric nitrogen dioxide from GOME, J. Geophys. Res. Atmos., 107(D20), 4437, doi:10.1029/2001JD001027, 2002a.
Martin, R. V, Jacob, D. J., Logan, J. A., Bey, I., Yantosca, R. M., Staudt, A. C., Li, Q., Fiore, A.

M., Duncan, B. N. and Liu, H.: Interpretation of TOMS observations of tropical tropospheric ozone with a global model and in situ observations, J. Geophys. Res. Atmos., 107(D18), 2002b.
Murray, L. T., Jacob, D. J., Logan, J. A., Hudman, R. C. and Koshak, W. J.: Optimized regional and interannual variability of lightning in a global chemical transport model constrained by LIS/OTD satellite data, J. Geophys. Res. Atmos., 117(D20), 2012.
Olivier, J. G. J., Van Aardenne, J. A., Dentener, F. J., Pagliari, V., Ganzeveld, L. N. and Peters, J. A. H. W.: Recent trends in global greenhouse gas emissions: regional trends 1970–2000 and spatial distributionof key sources in 2000, Environ. Sci., 2(2–3), 81–99, 2005.
Palmer, P. I., Jacob, D. J., Chance, K. and Martin, R. V: Air mass factor formulation for spectroscopic measurements from satellites' Application to formaldehyde retrievals from the Global Ozone Monitoring Experiment, J. Geophys. Res., 106(D13), 14,539-14550, doi:10.1029/2000JD900772, 2001.
Park, R. J., Jacob, D. J., Field, B. D., Yantosca, R. M. and Chin, M.: Natural and transboundary pollution influences on sulfate-nitrate-ammonium aerosols in the United States: Implications for policy, J. Geophys. Res. Atmos., 109(D15), 2004.
Qu, Z., Henze, D. K., Capps, S. L., Wang, Y., Xu, X., Wang, J. and Keller, M.: Monthly top-down NOx emissions for China (2005-2012): A hybrid inversion method and trend analysis, J. Geophys. Res. Atmos., doi:10.1002/2016JD025852, 2017.
Qu, Z., Henze, D. K., Theys, N., Wang, J. and Wang, W.: Hybrid mass balance/4D-Var joint inversion of NOx and SO2 emissions in East Asia, J. Geophys. Res. Atmos., 2018JD030240, doi:10.1029/2018JD030240, 2019.
Rodgers, C. D. and Connor, B. J.: Intercomparison of remote sounding instruments, J. Geophys. Res. Atmos., 108(D3), 2003.
Spurr, R. J. D.: Simultaneous derivation of intensities and weighting functions in a general pseudo-spherical discrete ordinate radiative transfer treatment, J. Quant. Spectrosc. Radiat. Transf., 75(2), 129–175, doi:10.1016/S0022-4073(01)00245-X, 2002.
Streets, D. G., Zhang, Q., Wang, L., He, K., Hao, J., Wu, Y., Tang, Y. and Carmichael, G. R.: Revisiting China's CO emissions after the transport and chemical evolution over the Pacific (TRACE-P) mission: synthesis of inventories, atmospheric modeling, and observations, J. Geophys. Res. Atmos., 111(D14), 2006.
Streets, D. G., Canty, T., Carmichael, G. R., De Foy, B., Dickerson, R. R., Duncan, B. N.,

Edwards, D. P., Haynes, J. A., Henze, D. K., Houyoux, M. R., Jacob, D. J., Krotkov, N. A., Lamsal, L. N., Liu, Y., Lu, Z., Martin, R. V., Pfister, G. G., Pinder, R. W., Salawitch, R. J. and Wecht, K. J.: Emissions estimation from satellite retrievals: A review of current capability, Atmos. Environ., 77, 1011–1042, doi:10.1016/j.atmosenv.2013.05.051, 2013.

Vinken, G. C. M., Boersma, K. F., Maasakkers, J. D., Adon, M. and Martin, R. V.: Worldwide biogenic soil NOx emissions inferred from OMI NO2 observations, Atmos. Chem. Phys., doi:10.5194/acp-14-10363-2014, 2014.

Wang, Y., Jacob, D. J. and Logan, J. A.: Global simulation of tropospheric O3-NO x-hydrocarbon chemistry: 1. Model formulation, J. Geophys. Res. Atmos., 103(D9), 10713–10725, 1998.

Van der Werf, G. R., Randerson, J. T., Giglio, L., Collatz, G. J., Mu, M., Kasibhatla, P. S., Morton, D. C., DeFries, R. S., Jin, Y. van and van Leeuwen, T. T.: Global fire emissions and the contribution of deforestation, savanna, forest, agricultural, and peat fires (1997–2009), Atmos. Chem. Phys., 10(23), 11707–11735, 2010.

Xu, X., Wang, J., Henze, D. K., Qu, W. and Kopacz, M.: Constraints on aerosol sources using GEOS-Chem adjoint and MODIS radiances, and evaluation with multisensor (OMI, MISR) data, J. Geophys. Res. Atmos., 118(12), 6396–6413, doi:10.1002/jgrd.50515, 2013.

Zhang, L., Shao, J., Lu, X., Zhao, Y., Hu, Y., Henze, D. K., Liao, H., Gong, S. and Zhang, Q.: Sources and Processes Affecting Fine Particulate Matter Pollution over North China: An Adjoint Analysis of the Beijing APEC Period, Environ. Sci. Technol., 50(16), 8731–8740, doi:10.1021/acs.est.6b03010, 2016.

Zhu, L., Jacob, D. J., Kim, P. S., Fisher, J. A., Yu, K., Travis, K. R., Mickley, L. J., Yantosca, R. M., Sulprizio, M. P., De Smedt, I., González Abad, G., Chance, K., Li, C., Ferrare, R., Fried, A., Hair, J. W., Hanisco, T. F., Richter, D., Jo Scarino, A., Walega, J., Weibring, P. and Wolfe, G. M.: Observing atmospheric formaldehyde (HCHO) from space: validation and intercomparison of six retrievals from four satellites (OMI, GOME2A, GOME2B, OMPS) with SEAC4RS aircraft observations over the southeast US, Atmos. Chem. Phys., 16(21), 13477–13490, doi:10.5194/acp-16-13477-2016, 2016.

Zoogman, P., Liu, X., Suleiman, R. M., Pennington, W. F., Flittner, D. E., Al-Saadi, J. A., Hilton, B. B., Nicks, D. K., Newchurch, M. J., Carr, J. L., Janz, S. J., Andraschko, M. R., Arola, A., Baker, B. D., Canova, B. P., Chan Miller, C., Cohen, R. C., Davis, J. E., Dussault, M. E.,

549 Edwards, D. P., Fishman, J., Ghulam, A., González Abad, G., Grutter, M., Herman, J. R., Houck,
550 J., Jacob, D. J., Joiner, J., Kerridge, B. J., Kim, J., Krotkov, N. A., Lamsal, L., Li, C., Lindfors,
551 A., Martin, R. V., McElroy, C. T., McLinden, C., Natraj, V., Neil, D. O., Nowlan, C. R.,
552 O'Sullivan, E. J., Palmer, P. I., Pierce, R. B., Pippin, M. R., Saiz-Lopez, A., Spurr, R. J. D.,
553 Szykman, J. J., Torres, O., Veefkind, J. P., Veihelmann, B., Wang, H., Wang, J. and Chance, K.:
554 Tropospheric emissions: Monitoring of pollution (TEMPO), J. Quant. Spectrosc. Radiat. Transf.,
555 186, 17–39, doi:10.1016/j.jqsrt.2016.05.008, 2017.
556
557

Figures:

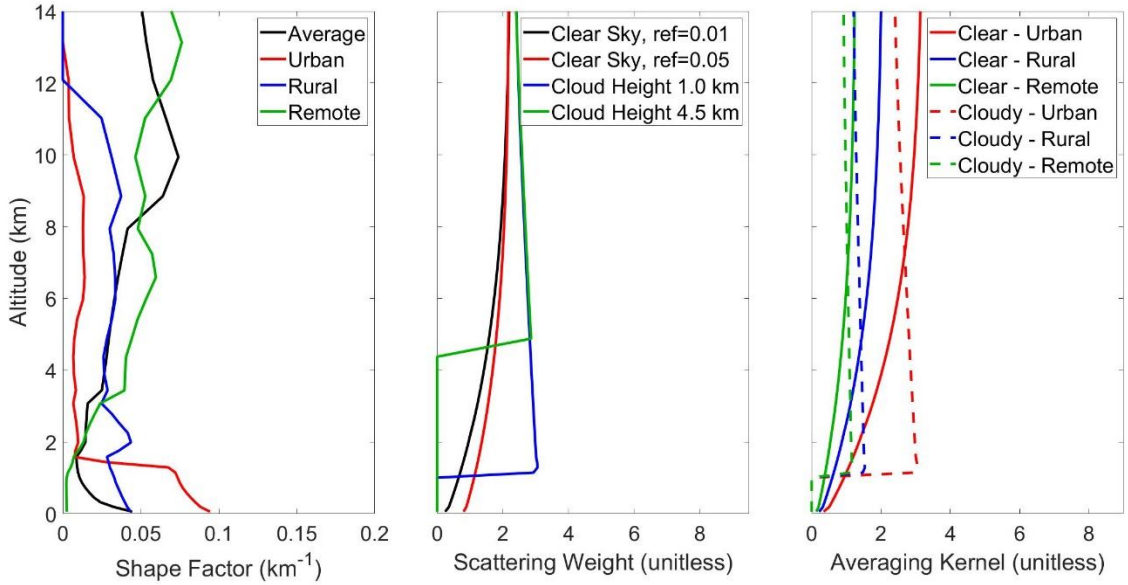



Figure 1: (Left) Shape factor profiles from a GEOS-Chem simulation for July 2010. Shown are a global average, and typical urban (Beijing), rural (Midwest USA), and remote (Tropical Pacific) profiles. (Middle) Typical OMI scattering weight profiles for varying surface reflectance and cloud height. (Right) Averaging kernels calculated using the same shape factors and scattering weights ("Clear Sky" surface reflectance is 0.01, "Cloudy" uses cloud height of 1 km).

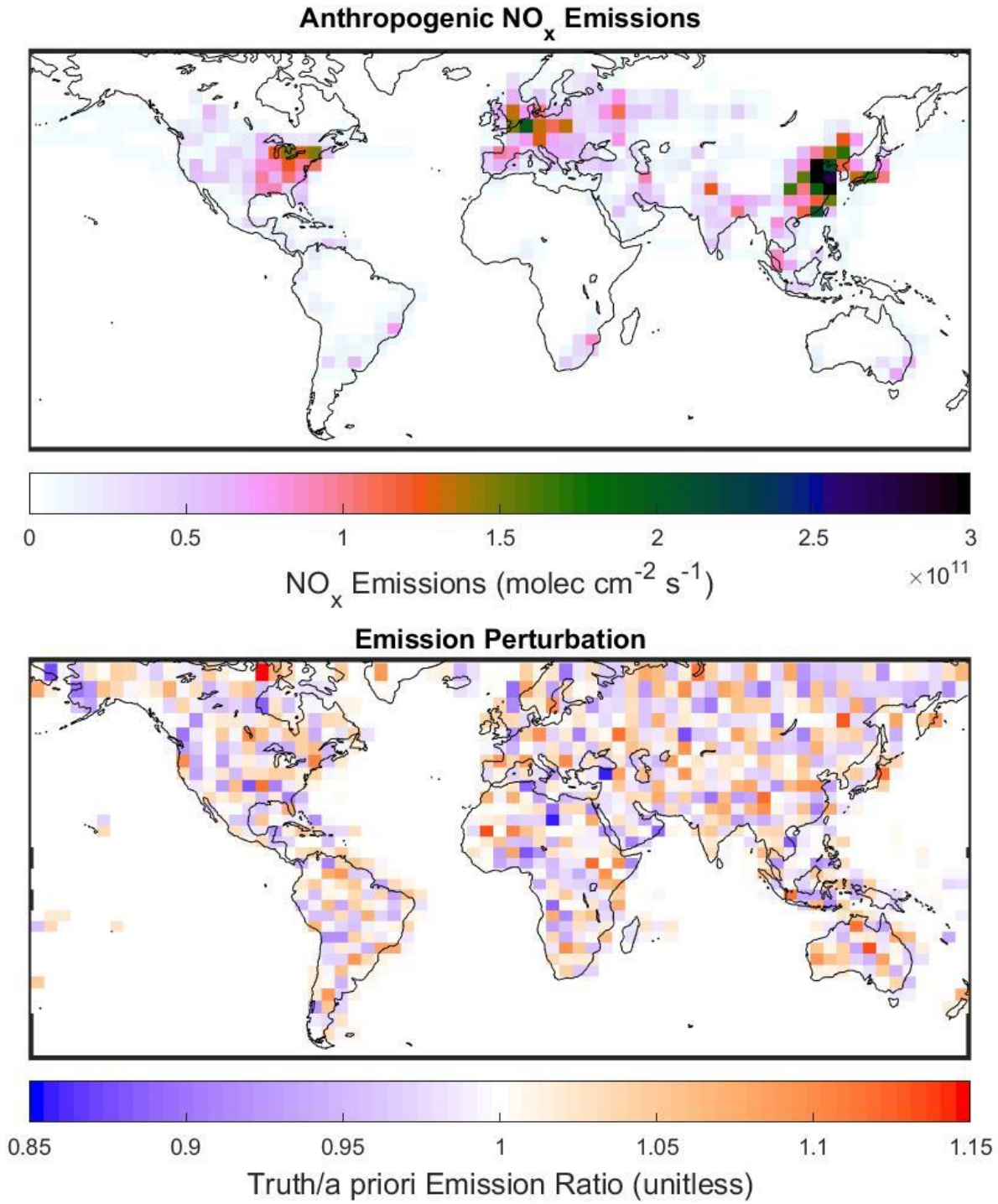
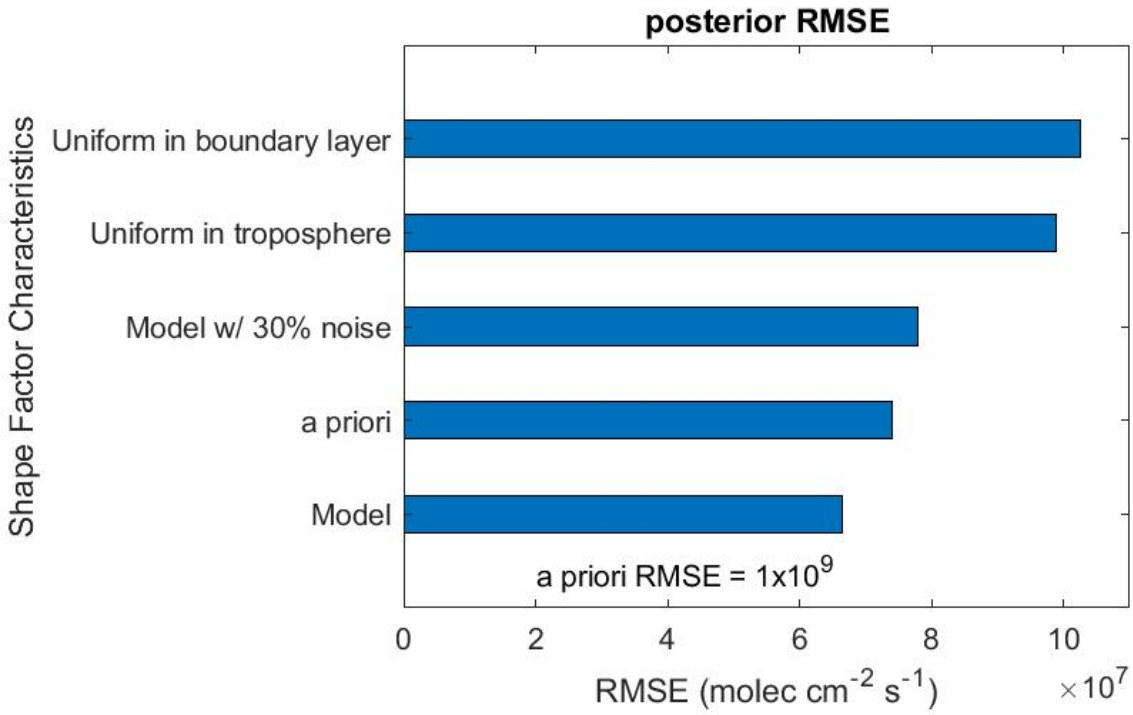
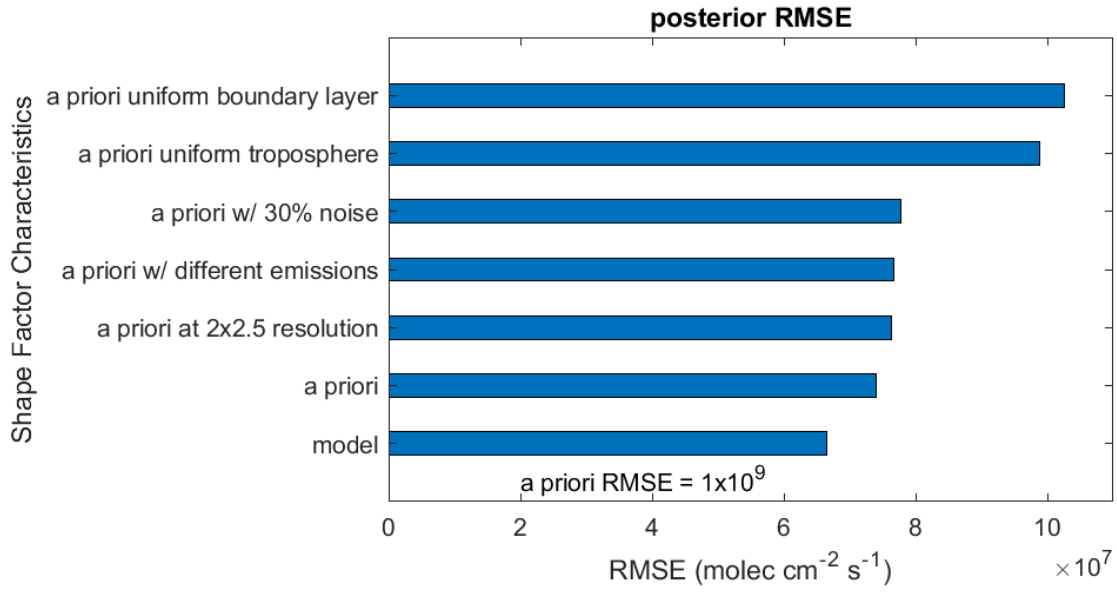



Figure 2 (top) Anthropogenic $NO_x$ emissions for July 2010 used in GEOS-Chem. (bottom) Ratio of "true" emissions used to create *Obs*$_5$ synthetic observations to a priori $NO_x$ emissions.

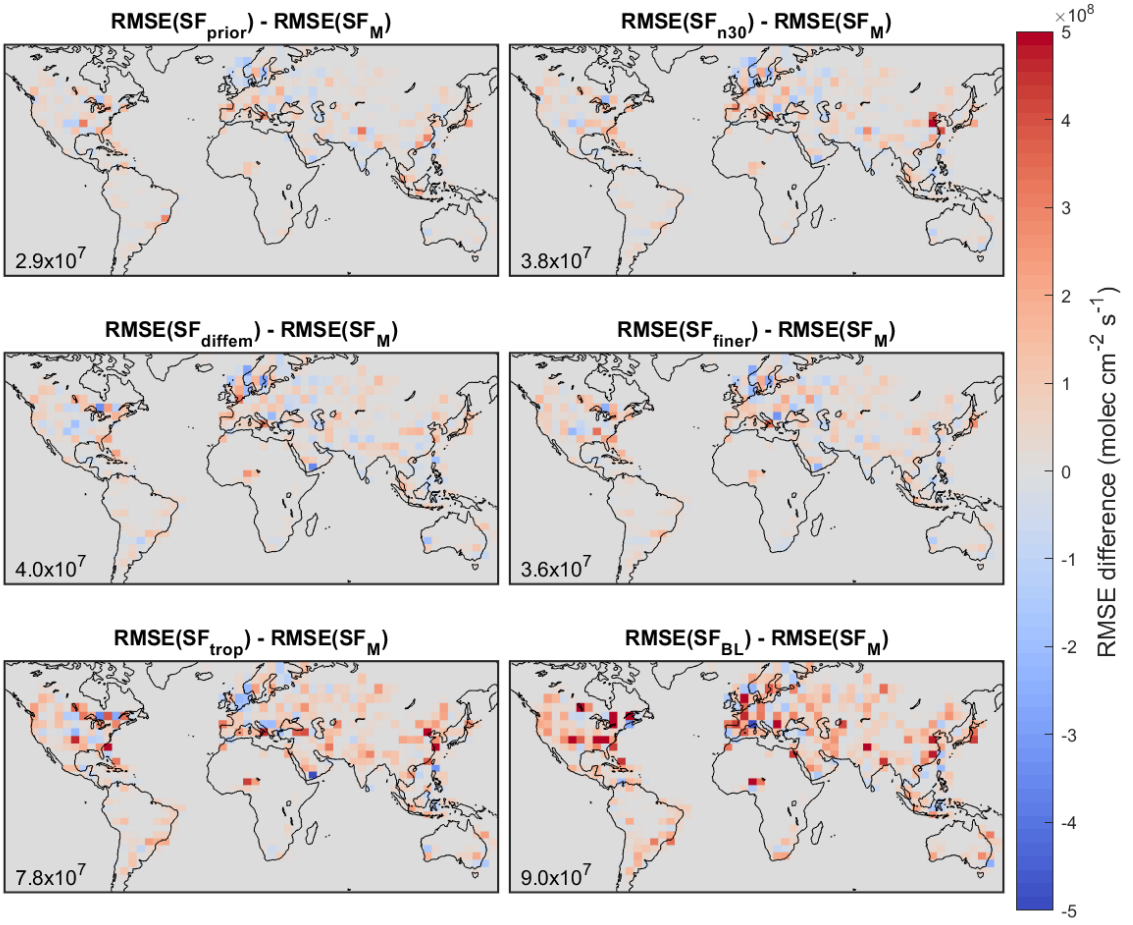

posterior RMSE
Shape Factor Characteristics
Uniform in boundary layer
Uniform in troposphere
Model w/ 30% noise
a priori
Model
a priori RMSE = 1x10^9
0
2
4
6
8
10
RMSE (molec cm^-2 s^-1)
×10^7

571


posterior RMSE
Shape Factor Characteristics
a priori uniform boundary layer
a priori uniform troposphere
a priori w/ 30% noise
a priori w/ different emissions
a priori at 2x2.5 resolution
a priori
model
a priori RMSE = 1x10^9
0
2
4
6
8
10
RMSE (molec cm^-2 s^-1)
×10^7

572

573 Figure 3: Global root mean square error (RMSE) values for 4D-Var estimates of $NO_x$ emissions
574 for tests using various shape factors in AMF calculations.
575




576

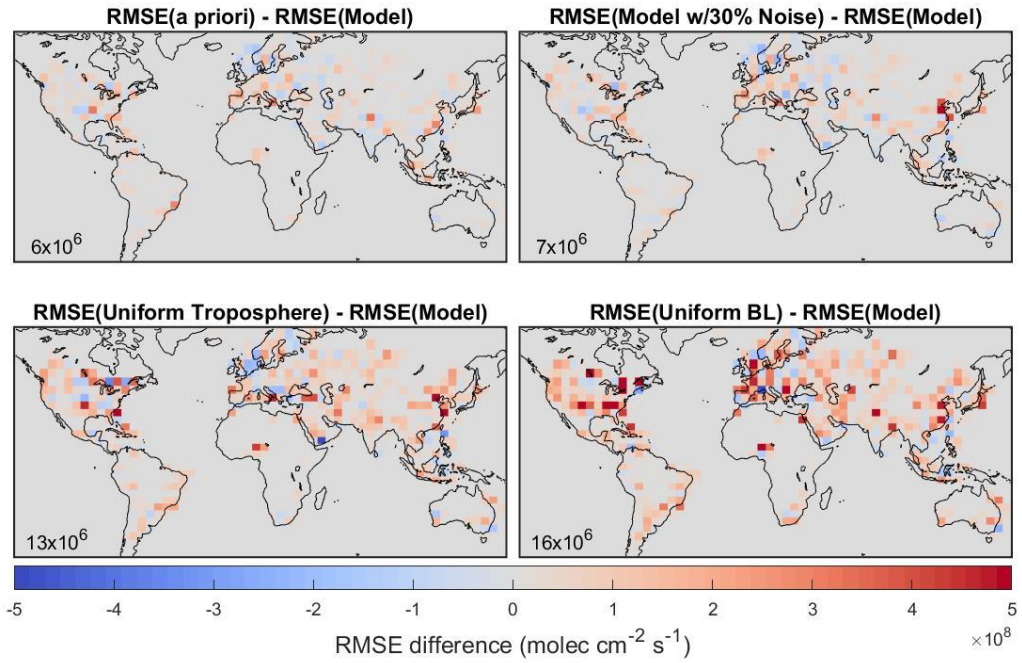
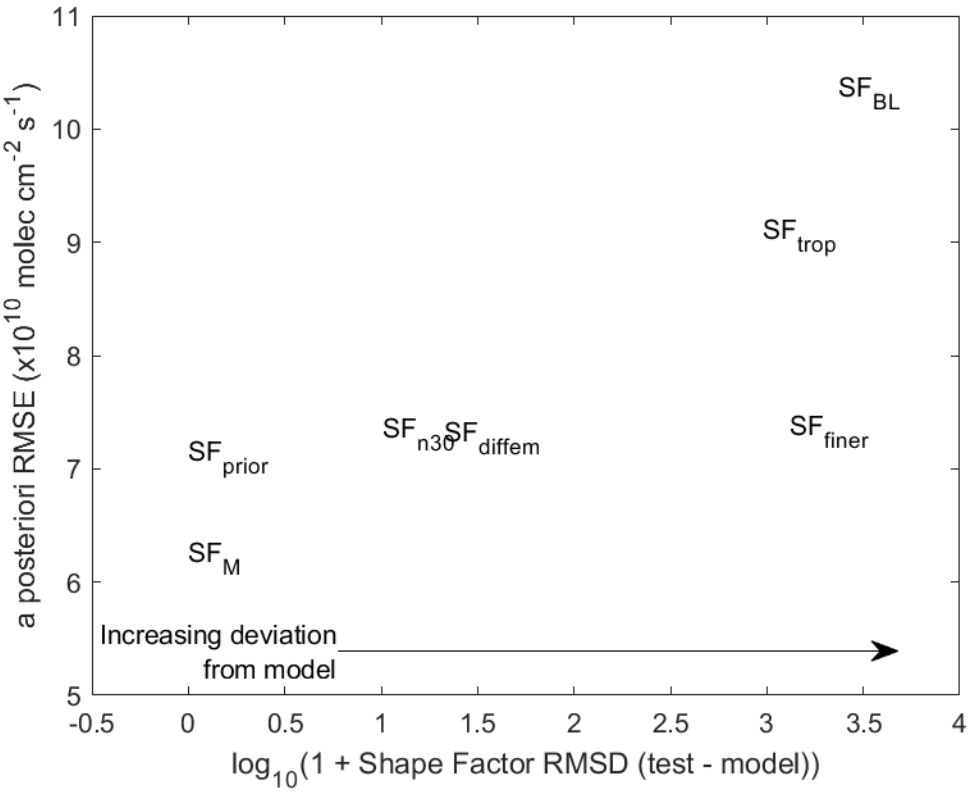



Figure 4: Difference between root mean square error (RMSE) of adjoint tests for $Obs_5$ synthetic observations. Root mean square differences between the *a posteriori* emissions ~~Mean absolute difference~~ estimates (molec/cm$^2$/s) are ~~values~~ inset.


11
10
9
8
7
6
5
-0.5
0
0.5
1
1.5
2
2.5
3
3.5
4
a posteriori RMSE (x10^10 molec cm^-2 s^-1)
log10(1 + Shape Factor RMSD (test - model))
SF_BL
SF_trop
SF_n30
SF_diffem
SF_finer
SF_prior
SF_M
Increasing deviation from model

582

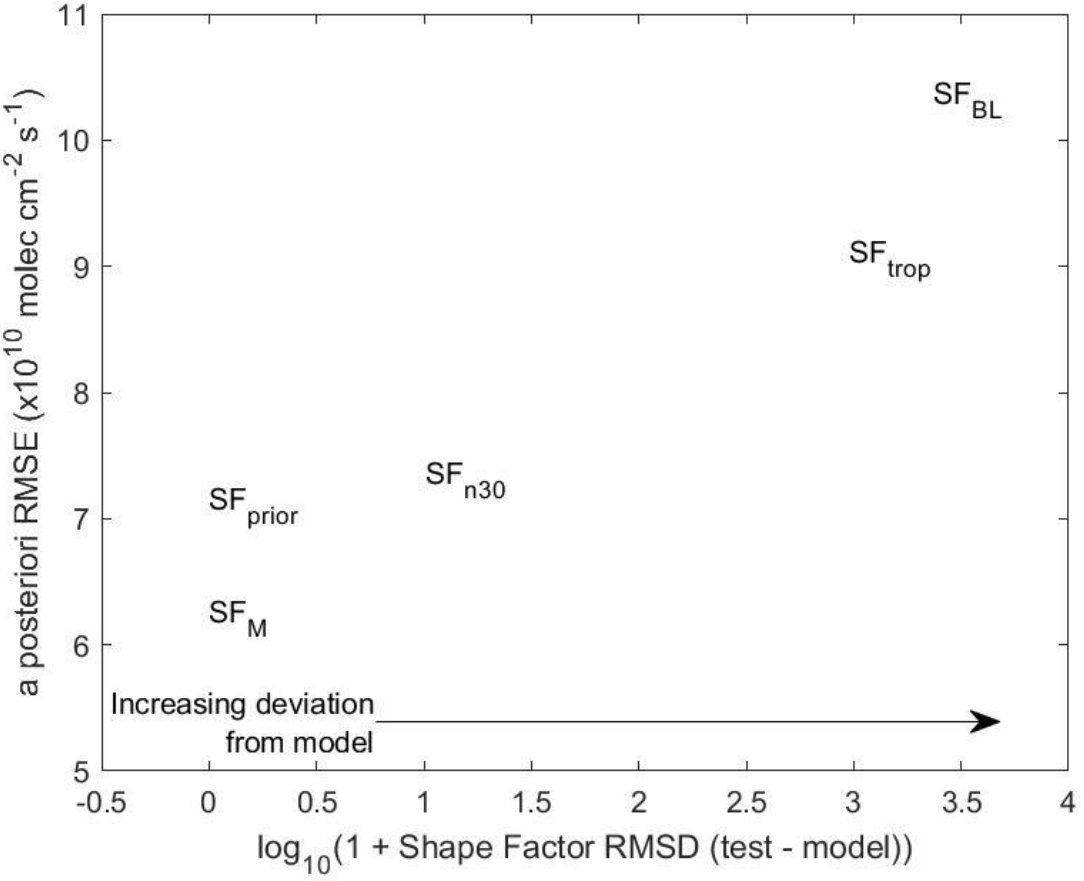



Figure 5: Scatterplot of adjoint test results. X-axis represents the deviation of the shape factor from the model simulated shape factor (root mean square difference). Y-axis represents the *a posteriori* emissions error from the adjoint inversion.

| Variable | *Palmer et al., 2001* | *Eskes & Boersma, 2003* | *Boersma et al., 2016* | Notation used here |
|---|---|---|---|---|
| Air mass factor | AMF | M | M | M |
| Slant Column | $\Omega_S$ | S | $N_S$ | $\Omega_s$ |
| Vertical Column | $\Omega_V$ | V | $N_V$ | $\Omega_v$ |
| Scattering Weight | w(z) | $C_l$ | $m_l$ | W |
| Shape Factor | $S_z(z)$ | | | S(z) |
| Averaging Kernel | | A | A | A |
| Number density | n(z) | X | $x_l$ | n(z) |
| Geometric AMF | $AMF_G$ | | | $M_G$ |

589 Table 1: Lexicon comparing notation used in this paper to that used in previous studies.

590

591

592

593

| Test name | Shape factor source | Air Mass Factor (unitless) | Synthetic observation ($Obs_5$) vertical column density (x$10^{15}$ molec/cm$^2$) | |
|---|---|---|---|---|
| | | Global Mean | Global Mean (x$10^{15}$ molec/cm$^2$) | Maximum difference from $SF_M$ (%) |
| $SF_M$ | Model | 2.1 | 0.9 | - |
| $SF_{n30}$ | Model w/ 30% noise | 2.1 | 0.9 | 19 |
| $SF_{diffem}$ | Model w/ different emissions | 2.1 | 0.9 | 28 |
| $SF_{finer2x25}$ | Model at finer (2°x2.5°) resolution | 1.6 | 1.2 | 23 |
| $SF_{trop}$ | Uniform in troposphere | 1.8 | 1.0 | 57 |
| $SF_{BL}$ | Uniform in boundary layer | 1.3 | 1.5 | 27 |

594 Table 2: Global mean air mass factors and synthetic observation vertical column density for

595 shape factors tested here.

596

| Test Name | Shape Factor Source | Correlation (r) of *a posteriori* ~~RMSE~~ error and "true" emissions | *a posteriori* RMSE ($x10^8$ molec/cm$^2$/s) | | Error standard deviation ($x10^8$ molec/cm$^2$/s) | | Maximum error ($x10^9$ molec/cm$^2$/s) |
|---|---|---|---|---|---|---|---|
| | | if "true" emissions > $10^{10}$ molec/cm$^2$/s | "true" emissions > $10^{10}$ molec/cm$^2$/s | "true" emissions > $10^{11}$ molec/cm$^2$/s | "true" emissions > $10^{10}$ molec/cm$^2$/s | "true" emissions > $10^{11}$ molec/cm$^2$/s | |
| $SF_M$ | Model | 0.06~~3~~* | 1.8 | 3.0 | 1.8 | 2.9 | 1.6 |
| $SF_{prior}$ | a priori | 0.11~~03~~* | 2.0 | 3.2 | 2.0 | 3.3 | 1.6 |
| $SF_{n30}$ | Model w/ 30% noise | 0.2~~1~~6 | 2.1 | 3.9 | 2.1 | 3.8 | 1.8 |
| $SF_{diffem}$ | Model w/ different emissions | 0.13 | 2.0 | 3.6 | 2.0 | 3.7 | 1.9 |
| $SF_{finer}$ | Model at finer (2°x2.5°) resolution | 0.05* | 2.1 | 3.2 | 2.1 | 3.2 | 1.8 |
| $SF_{trop}$ | Uniform in troposphere | 0.~~68~~39 | 2.8 | 5.6 | 2.8 | 5.5 | 2.1 |
| $SF_{BL}$ | Uniform in boundary layer | 0.17~~08~~* | 2.8 | 4.6 | 2.8 | 4.6 | 1.9 |

597 Table 3~~2~~: Summary of error statistics for adjoint tests. Values marked * indicate that correlation
598 is not statistically significant (p>0.05). For comparisons, mean "true" emissions for grid boxes
599 with emissions>$10^{10}$ molec/cm$^2$/s is 4.9x$10^{10}$, and mean "true" emissions for boxes with
600 emissions>$10^{11}$ molec/cm$^2$/s is 1.6x$10^{11}$ molec/cm$^2$/s.