

Reviewer 1:

Thank you for reviewing the paper and identifying places for improvement.

Following this, (1. Introduction, top of p. 20128) the following statement is incorrect: “Recent chemistry evaluations suggest that chemical representation could contribute to underpredictions of NO_x via overpredicting HNO₃ formation rates (Olson et al., 2006; Bertram et al., 2007; Ren et al., 2008). These studies do not support this statement.

As you point out, we incorrectly cited a *high*-bias for HO_x. These papers, however, do emphasize chemistry as a potential source of observed-to-model high-biases for HO₂. The Ren paper concludes, “at altitudes above 8 km suggests the presence of an unknown HO_x source or an error in the model’s chemistry involving some of the other atmospheric constituents.” The Olson paper shows that “changes in precursor inputs to the model and/or uncertainties in their observations and reaction rates have a more pronounced impact on HO_x predictions compared to conditions with lower NO_x levels.” Also, the Bertram 2007 paper, in the supplementary on-line material, cites high-bias in modeled OH values as the reason for using only observed HO_x values. To correct the manuscript, we adjusted the abstract statement below.

p 20126 “Recent observation-based studies, in the upper troposphere, identify chemical rate coefficients as a potential source of error (Olson et al., 2006; Ren et al., 2008).”

p 20128 “Zero dimensional modeling studies have suggested either missing observations or errors in chemical transformation of radical precursors in the upper troposphere (Olson et al., 2006; Ren et al., 2008).”

The authors need to state what value they use for NO. Table 2 does mention that calculated NO is used, rather than observation (and this is a good choice), but this should be brought out in the text as well. Additionally, Table 2 lists the HNO₃ measurement used as from CIT (similar to NO, I would also suggest this be mentioned in the text).

To address calculated NO and adjustments to HNO₃, both of which follow Bertram 2007 Supplementary Online Material, have been added to the methodology section as follows:

p 20131 line 24 “Observations of NO and HNO₃ in the INTEX-NA dataset have known uncertainties and limitations that require adjustment. The NO chemiluminescence measurement has a 1-minute integration time, is most reliable for mixing ratios greater than 100 ppt (Singh et al., 2007) and, during the INTEX-NA campaign, has a strong bias compared to steady-state NO ($[NO]_{ss} = j_{NO_2}[NO_2] / (k_{O_3+NO}[O_3] + k_{HO_2+NO}[HO_2])$). For this analysis, we require finer time resolution and detection of low NO (17% of $[NO]_{ss}$ observations are below 100 ppt), so we use the steady-state calculated mixing ratio. During the INTEX-NA study, the University of New Hampshire (UNH) and the California Institute of Technology

(CIT) took two separate measurements of HNO_3 . When observations are available simultaneously between 8 and 10 km, the linear least squares fit of UNH as a function of CIT is 61.7%. The reason for the discrepancy is currently unknown, so we and other researchers (Bertram, 2007) adjust UNH by a factor of 1.2 and CIT by a factor of 0.8. The CIT measurement has better time resolution, but less temporal coverage. We use the adjusted CIT measurement when available and fill in measurement gaps with the adjusted UNH measurement.”

The measurement of HNO_3 from the UNH instrument is quite different than values from the CIT during INTEX-NA (median ratio between the two for 8-10 km is ~ 0.68). This would result in NO_x/HNO_3 ratios nearly 50% larger than those derived using CIT- HNO_3 . How would these differences impact your calculated mean air parcel lifetimes and conclusions? A discussion of the uncertainty in your results due to the choice of HNO_3 measurement needs to be included.

To address the influence of HNO_3 measurement uncertainty on our analysis, we have re-run the simulations and evaluations for GEOS-Chem using only CIT and only UNH measurements. In the manuscript, we will address this uncertainty by adding a short description in the methods section and a detailed appendix section.

Methods: “The large discrepancy in the HNO_3 measurement could bias our chemical surrogate of age. As a result, we have performed this analysis using CIT, UNH, and the adjusted HNO_3 values (see Appendix). The conclusions of this study are robust to the choice of measurement. Since the cause of the discrepancy is unknown, we use the adjusted values in the rest of this analysis.”

Appendix: “The disagreement between HNO_3 measured by CIT and UNH has the potential to influence our mean air parcel lifetime. The CIT and UNH measurements are not always coincident in time, so analysis using either has fewer total observations. When using just the CIT measurements, there are only 507 total observations and only 65 initial conditions. When using the UNH measurements, there are 842 total observations and 107 initial conditions. The CIT and UNH measurements both have log-normal distributions. The CIT distribution is broader and less smooth than the UNH distribution (see Fig. 11).

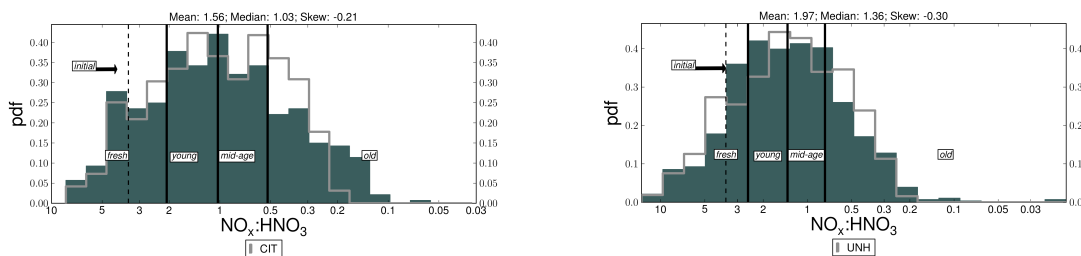


Fig. 11: Observed $\text{NO}_x:\text{HNO}_3$ (bars; left: CIT, right: UNH) compared to simulated (lines) from the GEOS-Chem chemical mechanism using the optimized, bias-corrected statistical model.

Our modeling framework is able to capture $\text{NO}_x:\text{HNO}_3$ distribution using the UNH and CIT measurements. Both the UNH and CIT mean air parcel lifetimes (τ_{UNH} and τ_{CIT}) are shorter than predicted with adjusted values (see 12). Using the lower UNH measurements increases the $\text{NO}_x:\text{HNO}_3$ values and shifts the $\text{NO}_x:\text{HNO}_3$ age bin cut-points to higher values. Because the chemistry model tends to overpredict conversion of NO_x to HNO_3 , the higher $\text{NO}_x:\text{HNO}_3$ observations and higher cut-points cause τ_{UNH} to be shorter than the standard air. Using the higher CIT measurements increases $\text{NO}_x:\text{HNO}_3$ values and shifts the $\text{NO}_x:\text{HNO}_3$ age bin cut-points to lower values. With lower cut-points, we might expect an increased τ_{CIT} relative to the adjusted values. The CIT measurements, however, extend the distributions low-value tail that the model cannot capture without over predicting values from 0.2 to 0.5. Because our goodness-of-fit statistic is sensitive to the whole distribution, τ_{CIT} is 1 hour shorter than when using adjusted HNO_3 values.

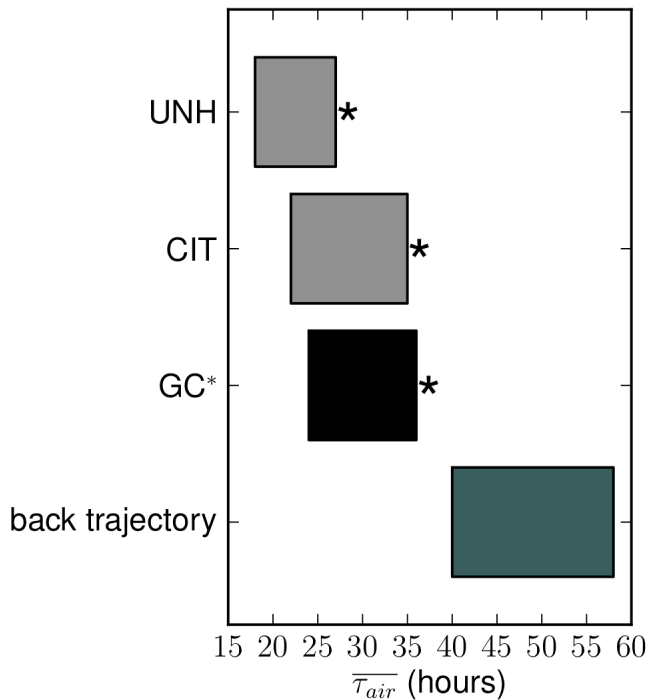


Fig. 12: Estimated mean air parcel lifetimes (τ_{air}) derived from back trajectory and chemical simulation. Chemical simulations use our updated GEOS-Chem mechanism with the adjusted HNO_3 , the unadjusted CIT measurement, or the UNH unadjusted measurement. Asterisks indicate whether chemically simulated $\text{NO}_x:\text{HNO}_3$ is statistically consistent with observations ($\alpha < 0.01$) when using the exponential (left, Eq. 3) and bias-corrected (right, Eq. 4) statistical models.

Using either measurement by itself increases the discrepancy between τ_{air} inferred

from back trajectories and τ_{air} inferred from chemical simulation. By increasing the discrepancy, using either measurement by itself increases our estimation of the chemistry-based NO_2 low-bias.

Figure 5: In calculating the statistical differences for these species, did the authors include instrument uncertainty? For instance, though the statistics identified O3 from SAPRC99 as statistically distinct from observations for the mid-age and old subsets, I have a hard time believing that such a small difference (few ppb) is significant.

For the purposes of fully addressing the authors comment, we separately address the uncertainty as “repeated measurement variance” and “systematic measurement bias.” For repeated measurement variance, the Mann-Whitney-Wilcoxon rank sum test accounts for uncertainty if the uncertainty distribution is symmetric. The rank sum test is calculated using rank paired observations, and will only reject the null hypothesis if observations throughout the distribution (i.e., not just the median) are consistently higher or lower. For systematic measurement bias, we attempt to account for small uncertainties by setting an extremely low alpha value ($\alpha = 0.0001$). If there is a systematic measurement bias, we expect it to uniformly affect the evaluation of all models.

Line 16 (H2O<200 ppb) typo? Should this be ppm? Whichever the case, this discrimination of stratospheric air confuses me. 200 ppm H2O is much too high to use to discriminate stratospheric air, but 200 ppb is extraordinarily low (i.e., when I obtain the 10s merged INTEX-A data between 8-10 km, I see no points that have water vapor (from the DLH instrument) less than 18 ppm). On the other hand, I find that 37% have H2O < 200 ppb, and of those points, 91% have O3 < 100 ppb (as low as 33 ppb), which is certainly not indicative of stratospheric influence. Why not use a combination of high O3 and low CO, along with the radioisotope ratios?

We thank the reviewer for catching this typo, H2O<200ppb should have been H2O<200ppm.

The combined requirement of $\text{H}_2\text{O} < 200$ ppm and $\text{Be}/\text{Pb} > 1000$ was based on an a priori analysis using 60 second unfiltered data. We chose not to filter using O_3 and CO because of the potential for the filtering process to influence the analysis of target species O_3 and CO . In this dataset, Be/Pb is never above 1000 when H_2O is above 200 ppm and so the H_2O filter is redundant and will be removed from the paper.

p 20131 line 16 “We also removed air parcels that might have been influenced by stratospheric intrusion (${}^7\text{Be}$: ${}^{210}\text{Pb}>1000$) or biomass burning ($\text{CH}_3\text{CN}>200$ ppt).”

Annotated Bibliography:

Bertram 2007 SOM Section 2.1 pgs 2,3: “As in other model descriptions of the UT during INTEX-NA (5, 6), our unconstrained model over-estimates OH by nearly a factor of two in the UT and under-estimates HO₂ by a similar amount.”