Response to **Reviewer #1**

Thank you for the review.

Chen et al. provide an overview of VOC budgets as observed in a set of aircraft campaigns and as modeled by GEOS-Chem. The model shows a dominance of biogenic VOC emissions in terms of magnitude and reactivity, and anthropogenic VOCs in terms of VOC loading. Model skill is variable, and the authors highlight the importance of discrepancies caused by a small subset of VOCs. They show discrepancies in the FT are linked to an underestimate of BL ventilation. Overall, the paper is well written and the results are suitable for publication in ACP. I have only a few remarks, below:

Thank you for the positive comments. We have prepared a point by point response to the reviewer's comments below.

Major and Minor Comments:

Comment 1: A predictable concern is that a single model year (2013) is directly compared to C1 observations spanning from 2010 to 2014. This may disproportionately influence anthropogenic VOCs comparisons, as California is largely sampled in 2010, and DC is sampled in 2011. It needs to be shown that interannual variability in emissions and meteorology is unlikely to affect the overall magnitude and spatial pattern in biases shown here.

Reply to Comment 1: Thank you for the suggestion. To address this concern, we conducted a set of one-month (June) high-resolution (0.25˚×0.3125˚) simulations spanning multiple years (2012-2016) to assess the potential impact of interannual variability. This approach provides a compromise between sufficient temporal coverage and tractable computational burden (running the full set of simulated months for all years would not be feasible). While the utilized aircraft campaigns span 2010-2014, we use the 2012-2016 timeframe to assess variability over 5 years based on the availability of the GEOS-FP meteorological product. June was chosen based on the temporal overlap of two campaigns (SENEX and CalNex) sampling very different regions. Each of the five monthly simulations was initialized after a ~1-week nested spin-up of regridded concentration fields from a 1-year global spin-up.

As the reviewer notes, factors driving interannual variability include both emissions and meteorology. In our simulations, biogenic emissions are computed in each model grid square and temporal step based on GEOS-FP meteorological data for that location and time, while pyrogenic emissions are from the GFED4s database which likewise includes interannual variability. Anthropogenic emission over North America employ NEI 2011 with year-specific scale factors applied, with the default GEOS-Chem configuration used here employing 2013 scale factors also for later years. The latest EPA trends information (Table below) indicates that total US anthropogenic VOC emissions for each year in 2010-2014 (the years encompassed by all the aircraft campaigns used in this study) differ by less than 4% from those of 2013 (the year of our simulations). Inventory changes in the transportation sector are somewhat larger but still less than 15%. As a result, even if we were to carry out model simulations spanning the entire time range of the aircraft studies and using year-specific EPA emissions, the modeled anthropogenic VOC flux would only differ by a few percent compared to the values used for 2013. Therefore, changing meteorology, biogenic emissions, and pyrogenic emissions are the main factors expected to drive differences in the model results for total-VOC quantities between years.

Figure S7 below shows the resulting multi-year variability in modeled VOC-carbon and reactivity profiles when sampling the model in different years along the June SENEX and CalNex flight tracks. The observed total VOC-carbon and reactivity is higher during SENEX than CalNex due to elevated VOC

emissions in the US Southeast plus a more extensive instrument payload. These differences in VOC-carbon and reactivity between the two regions are well-captured by the model regardless of the simulation year.

The modeled CalNex VOC-C and reactivity profiles show only modest interannual variability in both the planetary boundary layer (PBL) and free troposphere (FT). The simulated SENEX profiles likewise show very little interannual variability in the FT. We see a higher degree of variability for SENEX in the PBL and in the PBL-FT transition region, but for both VOC-C and reactivity the general pattern of model-measurement difference remains robust between 2013 (the simulation year used in the paper) and the other four years : specifically, a model underestimate near-surface, an overly-flat vertical profile within the PBL, and a strong FT underestimate.

We have added this new figure to the paper (as Fig. S7) and now state in Sect. 5.3 that "Given the range in measurement years spanned by the aircraft measurements, we performed a set of one-month simulations spanning multiple years to assess the potential impact of interannual variability on these findings. Results (see Supplement) suggest that the key features of the model-measurement comparisons discussed here are robust across years."

The following text (plus Fig. S7 below) have been added to the Supplement: "Given the range in measurement years spanned by the aircraft measurements, we performed a set of one-month simulations spanning 5 different years (June for 2012-2016) to assess the potential impact of interannual variability on these findings. Results are shown in Fig. S7 for the CalNex and SENEX flight tracks. In both cases the model-measurement VOC-C differences are highly consistent across years. We see a higher degree of interannual variability for VOC-reactivity over the SENEX domain, reflecting year-to-year differences in biogenic VOC emissions over this region. However, the key features of the comparison (a model underestimate near-surface, overly-flat vertical profile within the PBL, and strong FT underestimate) are consistent between our simulation year (2013) and the other four years."

Comment 2: Observed VOC-loading is operationally defined by aircraft payload (section 5.1). The authors note that this is typically more comprehensive in the summer-time, BVOCcentric campaigns. Are there any biases that are expected to persist in a certain dataset? This information may be extractable from Table S1, but it should be directly stated. Similarly, are there any points where VOC loading as measured by each aircraft can be compared?

Reply to Comment 2: We wish to clarify here that the model is sampled according to the specific VOC payload for each campaign, so that the model-measurement analyses are comparing consistent quantities in each case. For example, 25 observed VOCs were summed to derive the total-VOC quantities during SEAC^{4}RS (Fig. S19), and those same 25 VOCs as simulated by GEOS-Chem were then summed to derive the corresponding total-VOC quantities for the model. Analogously, both the modeled and observed total-VOC quantities for DISCOVER-AQ DC are based on the same set of 9 VOCs (Fig. S12). Only those VOCs that are present in both the GEOS-Chem model and in the measurement payload for a given campaign are included in these comparisons. As a result we do not expect any payload differences between campaigns to affect the conclusions of the model-measurement comparisons.

Based on the same reasoning, we feel that the model-measurement comparisons are more informative than a measurement-measurement total-VOC comparison between campaigns with inherently different payloads.

Comment 3: What are the BOVOC/AOVOC tracer sources (section 7)? Are they formed from the oxidation of some VOC precursor with a specific lifetime? A more detailed explanation of the model setup is needed.

Reply to Comment 3: We have modified the wording in Section 7 slightly to make this point more clear: "a unique pair of biogenic ($\mathcal{B}_{OVOC}$) and anthropogenic ($\mathcal{A}_{OVOC}$) source tracers was developed for each OVOC based on the mixing ratio difference along the flight track for that species between the model base-case and simulations with either i) all biogenic VOC emissions perturbed by 10%, or ii) all anthropogenic VOC emissions perturbed by 10% (see Section 4.3)."

These tracers are thus derived based on the three model simulations described in Sect. 4.3: R1) base case; R2) all biogenic VOC emissions decreased by 10%; R3) all anthropogenic VOC emissions decreased by 10%. For a given OVOC, $\mathcal{B}_{OVOC}$ would be derived as the modeled mixing ratio difference along the aircraft flight track between simulations R1 and R2, and $\mathcal{A}_{OVOC}$ as the R1 – R3 difference. Analogous tracer pairs were derived for each of the OVOC in our analysis.

Comment 4: Yu et al. (2018) also diagnose weak PBL ventilation – this cited in the conclusions but perhaps should be specifically discussed in section 6. Is the error associated with using off-line meteorology enough to drive the model errors observed here?

Reply to Comment 4: Thank you for the suggestion. We have now added a discussion of this point to Sect. 6 as follows:

"These findings are consistent with those of Yu et al. (2018), who diagnosed inadequate vertical transport in the current off-line configuration of the GEOS-Chem CTM. Yu et al. (2018) identified as causes i) the off-line convective transport scheme (leading to a +10% bias in modeled ^{222}Rn at the surface, and a -5% bias in the upper troposphere), and ii) off-line archiving of the meteorological fields (+5% model surface bias and -20% upper troposphere bias). Fixing these issues would therefore reduce the errors found here for VOC in the free troposphere (~60% mean low bias) but worsen the aggregated model performance in the PBL (~30% mean low bias). In that case, we would likely see in the model a more consistent low VOC bias throughout the troposphere, which would then indicate errors in overall VOC emissions or other processes."

Specific Remarks:

Remark 1: Page 3, line 118: "The model simulation includes extensive new developments related to atmospheric VOCs. . ." Can you clarify what developments are new in this work? "Latest" may be more appropriate than "new".

Reply to Remark 1: We have modified the wording of that sentence as suggested and now refer the reader to the section where more details may be obtained: "The model simulation includes the latest updates related to atmospheric VOCs (Section 2) and provides a more comprehensive representation of atmospheric organics than has been available for prior model-measurement evaluations."

Remark 2: Page 5 line 180: ". . . implemented into GEOS-Chem as described by Hu et al. (2015) with updated 0.25◦×0.3125◦ distributions of plant functional types and base emission factors". Please clarify the source of the updated plant functional types and base emission factors if different than Hu et. al (2015).

Reply to Remark 2: Thank you for the comment. The plant functional types and emission factor files used here have higher horizontal resolution than originally implemented by Hu et al. (2015), but are derived from the same underlying data version. We have therefore removed the phrase "with updated 0.25˚×0.3125˚ distributions of plant functional types and base emission factors" from the paper.

<u>Remark 3:</u> Figure 2: Please label fluxes to indicate net directionality.

<u>Reply to Remark 3:</u> Thanks for the suggestion. We have clarified in the Figure 2 caption that "Positive fluxes denote sources and negative fluxes denote sinks.".

Reference:

Yu, K., Keller, C. A., Jacob, D. J., Molod, A. M., Eastham, S. D., and Long, M. S.: Errors and improvements in the use of archived meteorological data for chemical transport modeling: an analysis using GEOS-Chem v11-01 driven by GEOS-5 meteorology, Geosci Model Dev, 11, 305-319, https://doi.org/10.5194/gmd-11-305-2018, 2018.

Tables and Figures:

Table R1. EPA National Tier 1 VOC Emission Trends: 2010 - 2014

Source Category	2010	2011	2012	2013	2014
Total without wildfires	14,988	15,263	14,981	14,699	14,416
% change relative to 2013	2.0	3.8	1.9	0.0	-1.9
Transportation	5,061	5,029	4,732	4,435	4,138
% change relative to 2013	14.1	13.4	6.7	0.0	-6.7

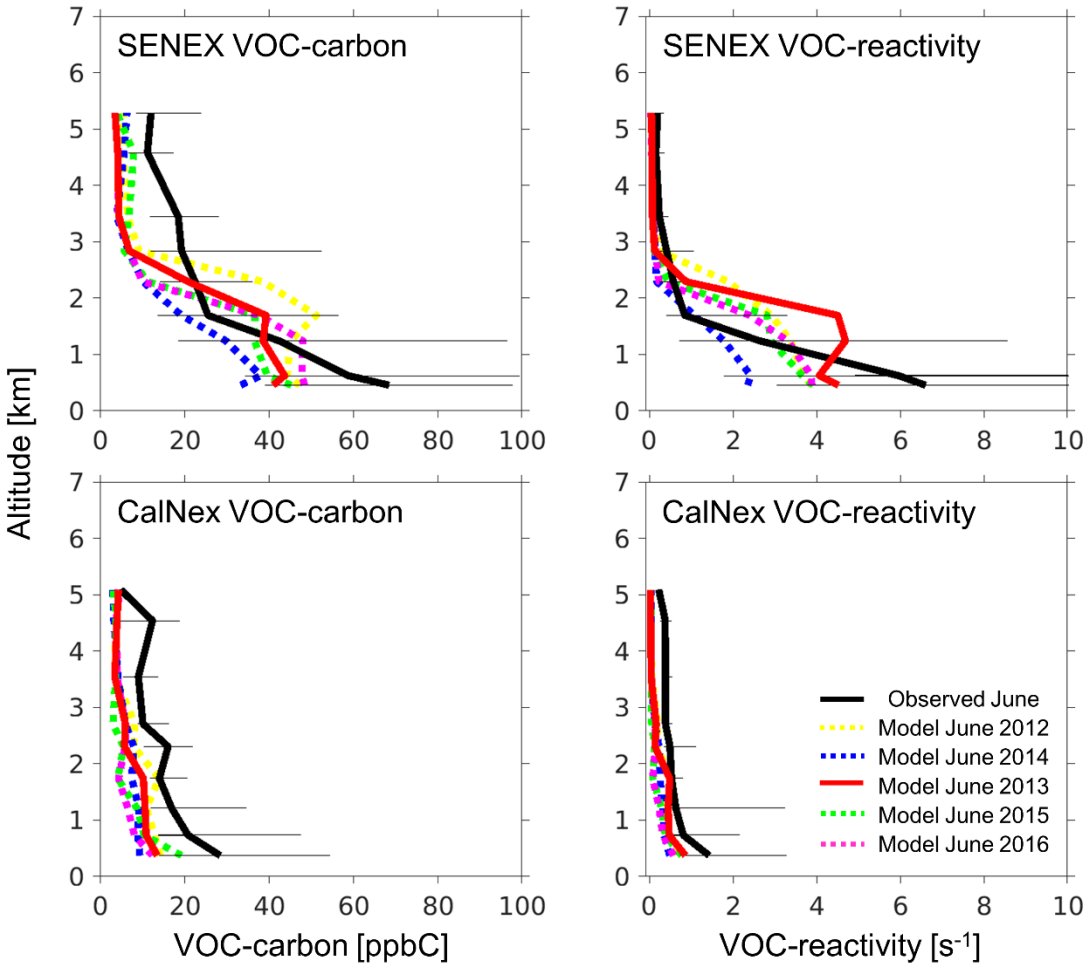

Figure S7. Vertical VOC-carbon and reactivity profiles as measured in 2013 and simulated by GEOS-Chem for June of 2012-2016 along the SENEX and CalNex aircraft flight tracks. Plotted are the observed (solid black lines) and predicted (2013: solid red lines; other years: dashed lines) median profiles, with horizontal bars indicating the 5th-95th percentiles measured for each vertical bin. Bin resolution is 0.5km below 3km and 1km above 3km. VOCs included in the profile correspond to the species shown in Fig. S17 (CalNex) and S18 (SENEX).