We thank the reviewer for his/her comments. Here they are reported (in bold) with our replies.

1. No comparisons are made to previous model versions. The authors have made no arguments for which conditions such a complex mechanism is needed. This work ought to show what difference, and particularly improvement, is achieved with this complex chemistry scheme over standard, more reduced schemes.

We thank the reviewer for pointing this out. We indeed have discussed this issue in the manuscript (line 46-54) : "A comparison for VOCs with the MIM chemical mechanism [...] that was used previously [...] is not shown here as in such mechanism (i) most of organics are either lumped [...] or missing; [...] (ii) primary species common to MIM and MOM would be influenced only by the different sinks (mainly OH), and a detailed description of OH budget is presented here; (iii) the model bias with respect to secondary species, e.g. oxygenated VOCs, has been linked to a mis-representation or lack thereof of processes like in-cloud chemistry." Therefore we believe that most of the comparison would not yield any additional information. It must be also stressed that we do not claim that this chemical mechanism is better than those previously used; rather the MOM mechanism is more complete, which allows studies which are not possible with simplified chemistry. In the few cases of specific trace gases (e.g. ozone or formaldehyde), which are present in both, MIM and MOM,a comparison could indeed be useful. This, however, would require a different model set-up, specifically aimed at such a comparison, and therefore is beyond the scope of this manuscript.

2. The VOC measurements used for evaluation of the model are really not adequate and numerous additional observations could have been used to evaluate the VOCs. The Emmons compilation of aircraft observations only includes data through 2000, and the quality and quantity of in-situ (aircraft) measurements of VOCs have increased dramatically since then. For example, the ATom missions provide global coverage of the compounds presented here (HCs, CH3CHO, CH3COCH3, etc.). Numerous other aircraft campaigns led by NASA, NOAA and NSF in the U.S. provide measurements over polluted regions (e.g., INTEX-NA, DISCOVER-AQ, SEAC4RS, SENEX, KORUS-AQ, FIREX-AQ, etc., etc.). These datasets are all freely available and in standard, similar formats which making them fairly easy to use, though it does require a bit more effort than taking the single profiles provided by the Emmons climatology.

We believe that an evaluation of a global model of such complexity, as the one presented in this study, should show the capability of the model to represent background/climatological conditions, rather than polluted/episodic ones. For this reason we adopted the so called "Emmons database", as each profile is based on multiple flights in the same regions, mostly for background conditions. As mentioned in the manuscript (line 138), we consider the used aircraft campaigns as representative for the period and the regions, i.e., equivalent to an observational-based climatology, although we acknowledge its limitations by using additional satellite observations. In addition, as these data have been used in many other evaluations, we can compare our results with previous studies.

A comparison with single flight measurements would give us a more "episodic" comparison, and should be accompanied with detailed process studies, as done before with this mechanism (e. g. Lelieveld et al., 2018; Tadic et al., 2021; Wang et al., 2020). Therefore, we restrain from using the campaign data suggested by the referee in such a general evaluation as the one presented here, as we would miss an overall overview.

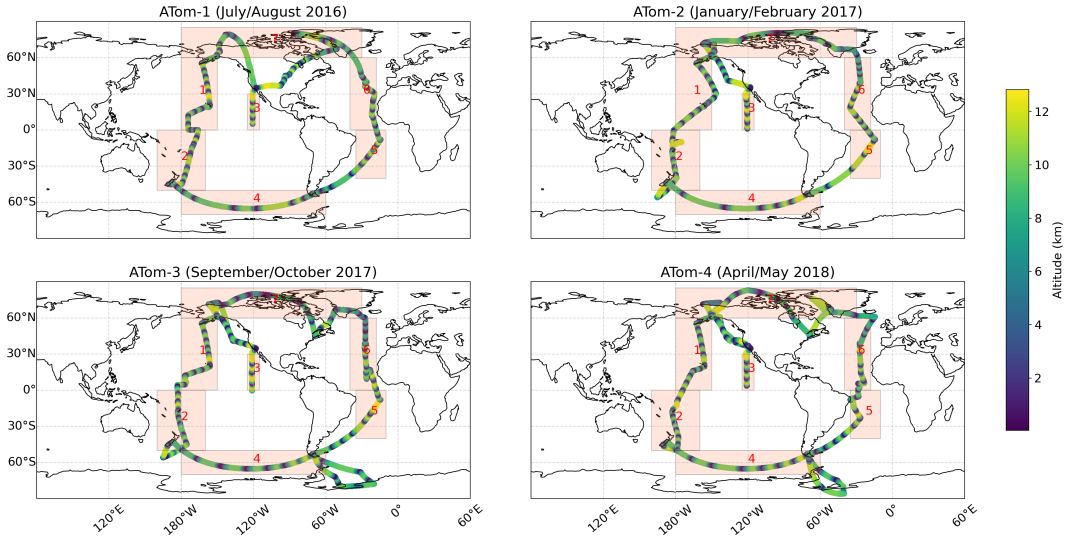

Figure 1: Flight paths of the 4 ATom campaigns. The respective altitude is colored and the paths are sub-divided into 7 different remote regions, namely Northwest Pacific (1), Southwest Pacific (2), East Pacific (3), Southern Ocean (4), South Atlantic (5), North Atlantic (6), and North Canada/Alaska/Greenland (7).

Nevertheless, we believe that the Atmospheric Tomography Mission (ATom campaign), does fit perfectly our needs, being over background regions, for different seasons, and with an extensive set of observed trace gases, and we thank the referee for pointing that out.

The ATom campaign took place from July 2016 to May 2018 and included measurements in four different seasons conducting 6 flights in each season. They roughly covered the same region in each season, including the Pacific, Atlantic, and Southern Ocean, as well as the continental region of North America. During the flights with the NASA DC-8 aircraft, they recorded numerous profiles (see Figure 1), which makes the dataset perfect for the evaluation of atmospheric model simulations.

The data used in this evaluation was taken from the "Merged Atmospheric Chemistry, Trace Gases, and Aerosols, Version 2 dataset" (Wofsy et al., 2021) and is based on various instruments: UC-Irvine Whole Air Sampler (WAS; Barletta et al., 2019), Trace Organic Gas Analyzer (TOGA; Apel et al., 2021), California Institute of Technology Chemical Ionization Mass Spectrometer (CIT-CIMS; Allen et al., 2019), NOAA Chemical Ionization Mass Spectrometer (NOAACIMS; Veres et al., 2021), Georgia Tech Ionization Mass Spectrometer (GTCIMS; Huey et al., 2019), NOAA Nitrogen Oxides and Ozone Instrument (NOyO3; Ryerson et al., 2019) and NOAA Picarro G2401 spectrometer (NOAA-Picarro; McKain and Sweeney, 2021).

Similar to the Emmons database, we divided the ATom observations into data for different regions (Figure 1; namely Northwest Pacific, Southwest Pacific, East Pacific, Southern Ocean, South Atlantic, North Atlantic, and North Canada/Alaska/Greenland) based on expected homogeneous properties of organic compounds in those region.

Analogously to the evaluation using the Emmons database, we want to make a climatological comparison between the simulated year for the evaluation (2010, see manuscript) and the data from the ATom campaign (2016-2018). With this comparison we can further evaluate, if the MOM mechanism is able to reproduce remote conditions. For that purpose we divided the altitude range

Species (Instrument)	N.points	MM	MSTD	OM	OSTD	$\overline{MM/OM}$	PF2	RMSE	CORR
CO (NOAA-Picarro)	332	76.86	18.98	81.22	23.99	0.98	100.0	10.56	0.82
C_2H_4 (WAS)	143	7.63	9.33	17.65	27.79	0.64	32.2	12.61	0.26
C_2H_6 (WAS)	334	369.47	191.54	550.41	375.48	0.79	86.5	195.09	0.87
C_3H_6 (WAS)	30	0.74	0.82	7.77	7.33	0.11	0.0	7.03	0.54
C_3H_8 (WAS)	330	48.38	65.63	83.64	130.04	0.97	61.5	44.92	0.87
CH_3COCH_3 (TOGA)	331	417.38	127.29	370.67	259.18	2.37	74.3	163.93	0.6
CH_3OH (TOGA)	333	372.0	186.19	601.5	393.94	1.02	64.0	292.98	0.5
CH_3OOH (CIT-CIMS)	306	161.03	143.64	419.17	412.65	0.47	33.0	259.92	0.89
HCHO (TOGA)	321	102.31	82.41	156.88	106.11	0.63	73.2	55.81	0.92
HCOOH (NOAACIMS)	168	39.44	21.43	188.59	359.94	0.61	35.7	155.87	0.3
HNO_3 (CIT-CIMS)	334	207.64	350.83	121.83	259.36	2.98	42.5	127.17	0.69
H_2O_2 (CIT-CIMS)	334	470.86	444.63	363.72	397.68	2.11	68.9	153.47	0.87
PAN (GTCIMS)	241	30.71	34.7	75.43	51.33	0.54	36.5	50.13	0.38
O_3 (NOyO3)	334	88.26	113.49	66.01	80.95	1.37	92.5	26.5	0.89
iC_4H_{10} (TOGA)	319	6.88	13.72	9.71	18.63	0.83	52.7	4.94	0.89
nC_4H_{10} (TOGA)	316	15.79	31.81	17.14	33.62	1.17	58.5	8.41	0.89

Table 1: Summary of simulated and observed mixing ratios of different tracers from the ATom campaign. N.points is the number of points used for the statistical estimation. The arithmetic mean of the model simulated mixing ratios (MM), the observed mixing ratios (OM), and the corresponding standard deviations (MSTD, OSTD) in pmol/mol are listed in the subsequent columns. For CO and O_3 the units are in nmol/mol. PF2 denotes the percentage of simulated points within a factor of 2 with respect to the observations, and RMSE represents the root mean square error between simulated and observed points. CORR is the Pearson correlation coefficient.

into 12 bins (linearly between 0.5 and 12.5 km) and defined a maximum of 336 (7 regions $\times$ 4 seasons $\times$ 12 altitude bins) points of interest (POI). For each POI we compared the mean value of all observations in the specific region, altitude bin and season to the mean value of all accordingly simulated values in the range of $\pm$ 1 day in 2010 around the flights conducted in that region.

Table 1 summarises the comparison between the ATom observations and simulated values (analogously to Table 2 in the manuscript). Unfortunately, there are no measurements of acetic acid (CH_3COOH) and the observations of propene (C_3H_6) are very limited. Thus, they will not be subject to further discussion.

The very good agreement of CO, C_2H_6 and O_3 (PF2 $> 85\%$) seen in the comparisons of the Emmons database is confirmed by the ATom observations. Also the mean values of the ratios $\overline{MM/OM}$ are in a comparable range for these species. In general, the RMSE is smaller for the comparison to the ATom campaign than for the Emmons database, increasing the confidence in the results of this evaluation. The overestimation of O_3 is already stated for the Emmons database.

However, the deviations for formic acid (HCOOH), nitric acid (HNO_3) and acetone (CH_3COCH_3) are larger. Formic acid is largely underestimated by our simulation (observations only exist for ATom-3 and ATom-4). The underestimation is present mainly at low altitudes, while there is a better agreement in the free troposphere. HNO_3 is substantially overestimated by our simulation, which is also confirmed by an additional measurement of another instrument. The Emmons database shows an overestimation of a smaller magnitude as well.

Measured CH_3COCH_3 is lower in the ATom dataset compared to the Emmons dataset. Thus, the model overstimates the mixing ratios, especially at high altitudes. However, there are still 74.3% of the simulated values within a factor of 2 of the observations. The comparison of simulated values to IASI observations showed an underestimation of factor 2 of the total column in the Northern Hemisphere. Figure 2 depicts the vertical columns of observations and simula-

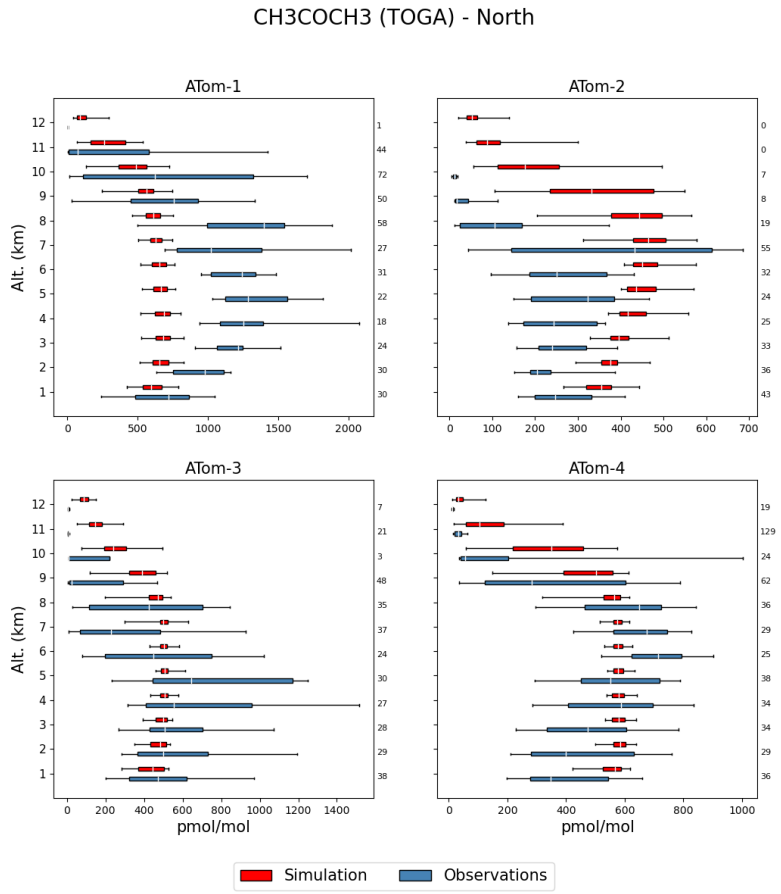

Figure 2: Vertical column of CH_3COCH_3 over Canada, Alaska and Greenland for the different ATom campaigns / seasons. The number on the right side of each plot is the number of observations in the respective altitude bin.

tions over the northern part of North America and Greenland. In Northern Hemisphere summer (July/August; ATom-1) this strong underestimation can be observed as well. However, for other seasons, this is not the case. In overall, there is only a small underestimation for this region.

Methyl hydroperoxide (CH_3OOH) shows insufficient agreement, between our simulations and the ATom observations (as it does for the Emmons database as well). However, this seems to be based on a systematic underestimation, while observations and simulation show a very high correlation (> 0.85) and the shape and vertical distribution is well reproduced. The same holds for formaldehyde (HCHO) that is in general underestimated by a factor of 0.63, but still has a high PF2 value and is strongly correlated (> 0.9). These features of CH_3OOH and HCHO are presented in Figure 3. PAN is underestimated in the same order of magnitude as in the Emmons database. There is in general a very bad agreement and low correlation between observations and simulation for PAN.

Additionally, we included an aircraft evaluation of i-butane (iC_4H_{10}) and n-butane (nC_4H_{10}) from the ATom observations. The mean ratio between simulation and observations is close to 1 and the simulated and observed mean values are very close and well within the standard deviations. The low PF2 value results from the high relative standard deviation. At the same time, both simulated iC_4H_{10} and nC_4H_{10} are very well correlated to the measurements (0.89).

Overall we conclude that thanks to the comment of the referee, we successfully extended our

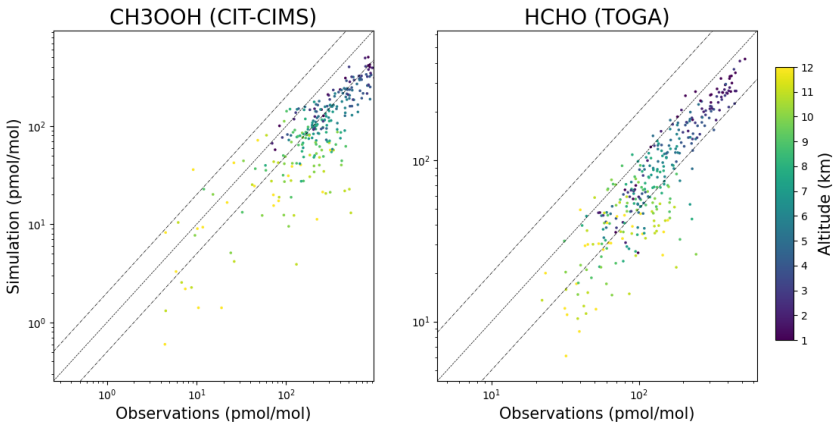

Figure 3: Scattered simulated values of CH_3OOH and HCHO against the observations of the ATom measurements. The different colors represent the altitude. Both species are generally underestimated by our simulations, but at the same time highly correlated between simulation and observations.

aircraft evaluation by including measurements from the ATom campaign. The results from the Emmons database are mostly confirmed, and thus our confidence in the simulations is strengthened. However, not all species could be properly evaluated due to a limited number of measurements for some species. Differences between simulation and observations could partly also be explained by differing meteorology in the considered year (2010 for the simulations, 2016-2018 for the observations). The reported results will be included in the revised manuscript.

The IASI retrievals are used extensively, but much more should be included about their accuracy and the results of published validation results. Perhaps they have much greater uncertainty at high latitudes, or remote regions - this should be discussed here.

Random and systematic components affect the uncertainties of the VOC measurements from IASI. Every IASI product comes with its own estimate of the random uncertainty associated with an individual retrieved column (see, e.g., Franco et al., 2018). For a non-background abundance of the considered species, the relative uncertainty on an individually retrieved column ranges typically between 20% and 50%, with the highest uncertainties found for the low columns. This single-pixel uncertainty increases for lower and background columns as the lower abundance of the target species approaches the IASI detection threshold. Nevertheless, these random uncertainties become negligible for the column averages calculated here for the comparisons with EMAC, because of the total number of measurements used per model grid cell. Indeed, after data filtering, 17 IASI measurements per day on average fall into each model grid box close to the Equator, i.e., more than 6,000 over the year 2010. Owing to the satellite polar orbits, this number increases with latitude and with the higher spatial sampling of IASI.

With respect to the systematic uncertainties, comparisons with independent measurements were performed to identify any potential bias in the IASI VOC columns. For the species that are also retrieved from ground-based FTIR measurements, namely CH3OH, HCOOH and PAN, column comparisons at various latitudes and environments were performed (Franco et al., 2020; Mahieu et al., 2021). Since no column measurements of acetone and acetic acid are currently available, the IASI data were compared to independent columns derived from tropospheric aircraft profiles

taken from various campaigns, complemented in the lower stratosphere by model data (Franco et al., 2019, 2020). These comparisons confirmed the absence of any large systematic biases of the IASI data, and that there is no noticeable latitudinal impact on the discrepancies. However, an underestimation (locally up to 30%) of the highest columns over tropical source regions (e.g., the Amazon Basin) during the dry season has been identified for all the species, except for PAN. Such underestimation of the elevated VOC columns affects the nadir-viewing sounders in general. For instance, it has also been observed in an extensive comparison of the TROPOMI formaldehyde product with over 20 FTIR measurement sites (Vigouroux et al., 2020).

Despite the underestimation of the highest columns locally, in the context of our study, the accuracy of the IASI products is relatively stable and sufficient to provide a global evaluation of the EMAC performance, considering the large uncertainties that still affect the emissions and atmospheric modelling of these VOCs.

References

Allen, H., J. Crounse, M. Kim, A. Teng, and P. Wennberg (2019). *ATom: L2 In Situ Data from Caltech Chemical Ionization Mass Spectrometer (CIT-CIMS)*. en. DOI: 10.3334/ORNLDAAC/1713.

Apel, E., E. Asher, A. Hills, and R. Hornbrook (2021). *ATom: Volatile Organic Compounds (VOCs) from the TOGA instrument, Version 2*. en. DOI: 10.3334/ORNLDAAC/1936.

Barletta, B., B. Biggs, D. Blake, N. Blake, A. Hoffman, S. Hughes, S. Meinardi, N. Vieznor, and C. Woods (2019). *ATom: L2 Halocarbons and Hydrocarbons from the UC-Irvine Whole Air Sampler (WAS)*. en. DOI: 10.3334/ORNLDAAC/1751.

Franco, B., L. Clarisse, T. Stavrakou, J.-F. Müller, A. Pozzer, J. Hadji-Lazaro, D. Hurtmans, C. Clerbaux, and P.-F. Coheur (2019). "Acetone Atmospheric Distribution Retrieved From Space". In: *Geophysical Research Letters* 46.5, pp. 2884–2893. DOI: 10.1029/2019gl082052.

Franco, B., L. Clarisse, T. Stavrakou, J.-F. Müller, D. Taraborrelli, J. Hadji-Lazaro, J. W. Hannigan, F. Hase, D. Hurtmans, N. Jones, E. Lutsch, E. Mahieu, I. Ortega, M. Schneider, K. Strong, C. Vigouroux, C. Clerbaux, and P.-F. Coheur (2020). "Spaceborne Measurements of Formic and Acetic Acids: A Global View of the Regional Sources". In: *Geophysical Research Letters* 47.4, e2019GL086239. DOI: 10.1029/2019gl086239.

Franco, B., L. Clarisse, T. Stavrakou, J.-F. Müller, M. Van Damme, S. Whitburn, J. Hadji-Lazaro, D. Hurtmans, D. Taraborrelli, C. Clerbaux, and P.-F. Coheur (2018). "A General Framework for Global Retrievals of Trace Gases From IASI: Application to Methanol, Formic Acid, and PAN". In: *Journal of Geophysical Research: Atmospheres* 123.24, pp. 13, 963–13, 984. DOI: 10.1029/2018jd029633.

Huey, L., J. Nowak, D. Tanner, and S. Kim (2019). *ATom: L2 In Situ Peroxyacetyl Nitrate (PAN) Measurements from Georgia Tech CIMS*. en. DOI: 10.3334/ORNLDAAC/1715.

Lelieveld, J., E. Bourtsoukidis, C. Brühl, H. Fischer, H. Fuchs, H. Harder, A. Hofzumahaus, F. Holland, D. Marno, M. Neumaier, et al. (2018). "The South Asian monsoon—pollution pump and purifier". In: *Science* 361.6399, pp. 270–273.

Mahieu, E., E. V. Fischer, B. Franco, M. Palm, T. Wizenberg, D. Smale, L. Clarisse, C. Clerbaux, P.-F. Coheur, J. W. Hannigan, et al. (2021). "First retrievals of peroxyacetyl nitrate (PAN) from ground-based FTIR solar spectra recorded at remote sites, comparison with model and satellite data". In: *Elem Sci Anth* 9.1, p. 00027.

McKain, K. and C. Sweeney (2021). *ATom: CO2, CH4, and CO Measurements from Picarro, 2016-2018*. en. DOI: 10.3334/ORNLDAAC/1732.

Ryerson, T., C. Thompson, J. Peischl, and I. Bourgeois (2019). *ATom: L2 In Situ Measurements from NOAA Nitrogen Oxides and Ozone (NOyO3) Instrument*. en. DOI: 10.3334/ORNLDAAC/1734.

Tadic, I., C. M. Nussbaumer, B. Bohn, H. Harder, D. Marno, M. Martinez, F. Obersteiner, U. Parchatka, A. Pozzer, R. Rohloff, M. Zöger, J. Lelieveld, and H. Fischer (2021). "Central role of nitric oxide in ozone production in the upper tropical troposphere over the Atlantic Ocean and western Africa". In: *Atmospheric Chemistry and Physics* 21.10, pp. 8195–8211. DOI: 10.5194/acp-21-8195-2021.

Veres, P., J. Neuman, and T. Ryerson (2021). *ATom: L2 Measurements from NOAA ToF Chemical Ionization Mass Spectrometer, Version 2*. en. DOI: 10.3334/ORNLDAAC/1921.

Vigouroux, C., B. Langerock, C. A. Bauer Aquino, T. Blumenstock, Z. Cheng, M. De Mazière, I. De Smedt, M. Grutter, J. W. Hannigan, N. Jones, et al. (2020). "TROPOMI–Sentinel-5 Precursor formaldehyde validation using an extensive network of ground-based Fourier-transform infrared stations". In: *Atmospheric Measurement Techniques* 13.7, pp. 3751–3767.

Wang, N., A. Edtbauer, C. Stönner, A. Pozzer, E. Bourtsoukidis, L. Ernle, D. Dienhart, B. Hottmann, H. Fischer, J. Schuladen, J. N. Crowley, J.-D. Paris, J. Lelieveld, and J. Williams (2020). "Measurements of carbonyl compounds around the Arabian Peninsula: overview and model comparison". In: *Atmospheric Chemistry and Physics* 20.18, pp. 10807–10829. DOI: 10.5194/acp-20-10807-2020.

Wofsy, S. et al. (2021). *ATom: Merged Atmospheric Chemistry, Trace Gases, and Aerosols, Version 2*. en. DOI: 10.3334/ORNLDAAC/1925.