Response to Anonymous Referee #2 We thank the referee for the review of our manuscript and in making insightful comments which we feel improve on the final version of the manuscript. We provide detailed responses to the point raised below.

General Comments:

I was expecting that the authors will evaluate the impact of the updates to the homogeneous and heterogeneous chemistry against the older model version without these updates to demonstrate how important these updates here. However, such a comparison is not presented. Thus, the real value of these updates is not quantified.

The updates to reaction data and chemical mechanism presented in Tables 1 through 6 were implemented and tested in a stepwise fashion, with these steps being documented in publicly accessible deliverable reports written and submitted to the CAMS consortium. For brevity, and to also address the criticism from Dr. Knowland on the length of the paper, we felt the need to limit the number of simulations to those chosen. Rather than investigating the improvement of the CB05BASCOE scheme, the focus is primarily on the divergence seen across participating members of a proposed chemical ensemble initiative and to focus on seasonal differences and the main reasons for such differences when employed under identical conditions. We now also move the tables related to some of the detailed model revisions to the Appendix, as suggested by the referee.

The authors compared the three chemistry configurations against the CAMS reanalysis (CAMSRA). At lines 76-77, the authors state that any changes in the operational system should ensure that the change improves the model performance with respect to observations. In most cases (e.g., Figures 4- 6), the performance of CAMSRA is better than all the three individual configurations. Since any of the new configurations here do not outperform CAMSRA, does it mean that these configurations may not be suitable for operational adaptation? Furthermore, the assimilation of several species in CAMSRA makes a direct comparison of three chemistry simulations with CAMSRA difficult.

The intention of including the CAMSRA dataset was to provide a clear point of reference to a well-known reanalysis which is accessible to the community. We feel it provides some assessment with respect to the three chemical schemes employed in a more updated cycle of the IFS run at a higher horizontal resolution. Our aim was not to document each model improvement explicitly, but we agree with the reviewer confusion may have been raised throughout the discussion of model evaluations on this aspect. Clearly the current model performance is found occasionally to be of poorer quality as CAMSRA, but, as the reviewer also notes, this is at least in part caused by actual data assimilation, particularly for trace gases that are well-constrained (CO). Updates made to the IFS cycles are separate from updates made to the chemistry modeling aspect. A cycle update also changes to the data-assimilation configuration, the meteorological model. and emissions estimates that are applied. A rigorous assessment of the net changes is made and documented publicly in various system upgrade reports hosted at https://atmosphere.copernicus.eu/eqa-reports-global-services (last access: 21-12-21). We now clarify better the scope of our assessment by adding the following sentence: "However, attributing the differences in performance of chemical updates without data assimilation against CAMSRA is complicated by the simultaneous changes in IFS cycle and emission estimates since CAMSRA was produced."

The choice of US as a polluted region over Asia is not motivating enough given that anthropogenic emissions in Asia are much higher than the US.

The large number of independent measurements available for 2014/2015, especially for the lower troposphere from multiple aircraft campaigns, provides a unique opportunity to be able to investigate and quantify the seasonal variability over a significant area. Moreover, we wanted to focus on a corresponding NOy analysis for which aircraft measurements are a unique source. No such measurements are available with such a frequency for either Europe or China, which would have significantly hampered the analysis. Whilst CAMS provides a regional forecast for air quality using an ensemble of regional models for Europe, for other regions over the globe CAMS relies on global model simulations to provide such information due to the size of the domains. Therefore, we argue that quantification of the model performance in terms of air quality aspects for another region than Europe, with specific emphasis on uncertainty due to atmospheric chemistry modeling, is of interest to the CAMS consortium and wider community. This is the first time that such an in-depth analysis of the CAMS system adopting independent chemical schemes has been performed for the US.

57-58: Operational air quality forecasts over the US are not only provided by the CAMS system. The US has their National Air Quality Forecasting Capability (NAQFC) that provides air quality forecasts at 12 km grid spacing twice per day and the NAQFC forecasts are used by decision- makers to issue advance notices and warnings for air pollution episodes. Since lines 43-55 discuss aspects and benefits of air quality forecasting, it is relevant to discuss NAQFC here.

We agree that we should prevent confusion as to the status of the CAMS system to provide operational air quality forecasts on a global scale. Therefore, we have added the following text related to other forecasting systems:

"Forecasting services for the US are also provided by the National Air Quality Forecasting Capability (NAQFC,) for ozone (O_3), dust and smoke (e.g Chai et al., 2013; Lee et al., 2017), which utilizes the CMAQ model (Byun and Schere, 2006) driven by WRF-NMM meteorological model (e.g. Eder et al, 2009), provided at a 12 x 12km resolution. The model is regional meaning that prescription of boundary conditions is necessary to account for effects of long-range transport. One other notable system for AQF is the NASA GEOS Composition Forecast Modeling System (GEOS-CF v1.0; Keller et al, 2021), which is a global system similar to CAMS using the GEOS-chem chemical module coupled to the GEOS atmospheric model (v5) run at a global resolution of 25 x 25km."

Lines 63-64: Anthropogenic emissions in the US are much lower than Asia and thus it is hard to see this as a motivation for focusing on the US.

Please see our response to the second major point above.

Line 99: To what heterogeneous reactions you are referring here?

Principally the conversion of N_2O_5 to HNO_3, which has been improved on in the chemical schemes compared to previous studies and is considered critical for controlling the NOx regeneration cycle. We change the sentence to: "Huijnen et al. (2019) has highlighted the importance of heterogeneous reactions, especially the conversion of N_2O_5 to HNO_3, to explain differences".

Line 105: I suggest replacing the term "mini-ensemble" with "multiple chemistry simulations".

The reviewer is correct that we need to be careful with any referencing of our model configuration that is based on three model versions with independent chemistry as a 'mini-ensemble', because an actual ensemble should consist of a much more independent, well-chosen selection of members, and also requires assessment of ensemble statistics. The use of the referencing of three model versions as 'mini-ensemble' is essentially chosen from a practical perspective, in combination with an implicit future perspective where these chemistry versions could form a building block of constructing an actual global chemistry ensemble. We now update the sentence and write: "Furthermore, by applying all three model versions **with largely independent chemistry modules, for brevity here referred to as a chemistry mini-ensemble,** uncertainty ranges due to inaccuracies in the chemical component of the forecast can be estimated"

Line 144: Do all the chemistry configurations use the same quantum yield and cross sections for calculation of photolysis frequencies? Also, how are the effect of aerosols treated in photolysis rate calculation? If not, can the 5% differences in photolysis frequencies be attributed to these factors?

There is no homogenization across chemistry configurations regarding the calculation of photolysis frequencies, where differences have already been described in the previous literature. The Aerosol Optical Depth is used to attenuate the photolytic flux within each approach. We have added the sentence: "Different photolytic data is used for each of the chemistry configurations."

Tables 2-5: Please include a description of various symbols (Ko, R, T, and QY etc.) used in the rate expression column in the captions of these table.

We have added the following definition in the Table heading of Table 2: Key to abbreviations: T – Temperature (°K), K0 - Low Pressure Limiting Rate Constant, K_{inf} - High Pressure Limiting Rate Constant, QY – Quantum Yield and R – product yield.

No HONO emissions are introduced directly. We now state this in the header of Table 7 addressing the emissions applied in the simulations.

We thank the reviewer with this suggestion and will move some of the Tables associated with the model updates not attributed to Nitrogen Oxides.

We have now re-written the sentence to: "There is a minimum in Central rural regions around Kansas/Nebraska and North/South Dakota, with a variability of between 35-45 ppb in IFS(CBA)."

This is likely due to the lower PAN formation in IFS(MOZ) which reduces the fraction of NO$_x$ (c.f. Figure 18 and Figure S7) which can be transported away from the main source regions, which alters O$_3$ formation

We have modified the sentence accordingly: "Over the surrounding oceans, IFS(MOZ) has a decrease in mixing ratios of between 5-10 ppb due to reduced transport of pre-cursors such as PAN (see NO$_y$ discussion below)."

We agree that this is a somewhat surprising result which is difficult to explain being a consequence of the different reaction kinetics employed across the chemical schemes, which drive the OH production, recycling, and loss (e.g. Lelieveld et al., ACP, 2016). For example, primary production of OH from O$_3$ photolysis is strongly affected by the photolysis rate employed. Analysing the global budget terms shows that the net loss of O$_3$ via photolysis is around 12% higher in IFS(MOZ) compared with IFS(CBA), which will contribute towards the regional differences shown for the US. The extent of OH recycling is also influenced by the rate of oxidation of NO by HO$_2$, with higher summertime NO in IFS(MOZ) (Figure S2) introducing a greater recycling rate for the chemical scheme. Loss will then be the cumulative scavenging across reactants, where the kinetics of OH + CO exhibit a different pressure dependency without the inclusion of an intermediate in IFS(MOC) as for IFS(CBA) and IFS(MOZ) schemes. Due to the size and expense of the simulations it was not able to provide a full 3D budget analysis of the US region, which would provide insight in the changes in reaction cycles. We now add an additional sentence: "The OH primary production is globally higher in IFS(MOZ) than IFS(CBA), likely contributing to the higher summertime OH in this configuration."

We are somewhat puzzled by this point as when we revisited the original measurement data (hosted at: https://gml.noaa.gov/aftp/data/ozwv/SurfaceOzone/NWR/2014/ ; last access 27th December 2021) we do not find any significant gaps in the measurements for O$_3$ during 2014, with a (changing) value

being given every hour for each day and month allowing a valid observational monthly mean to be assembled. Therefore, we do not make any changes with respect to observational data frequency. We show a graphic of the dataset used below:
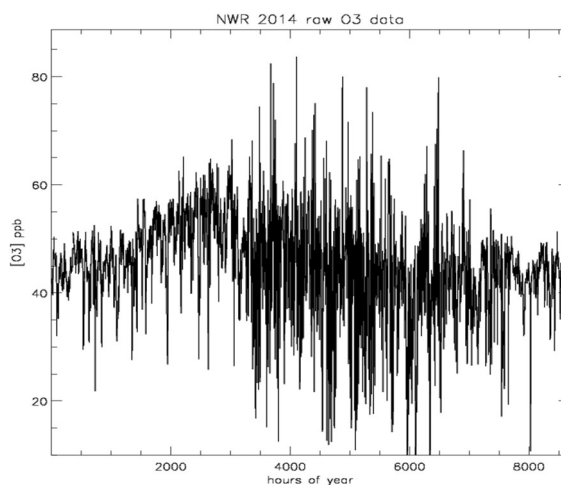

The hourly O_3 data used for the comparison presented for the NWR site, 2014.

Lines 535-540: Here you use ppb as ozone units while all the earlier discussion is in micrograms/m3. Please use consistent units.

We adopt the units provided by the various measurement networks therefore use these for our comparisons, where conversion of micrograms/m3 requires co-located pressure and temperature values associated with the measurements.

Line 581-582: Can you provide more insight into the different seasonal cycles of CO between the model and observations at NWR?

The observational site at Niwot Ridge is located approximately 35 km west of Boulder on top of a mountain range at around 3000m altitude. This site acts as an ideal location at which to sample background concentrations of CO, with no anthropogenic emission sources near the site. This means that this site is challenging when sampled by our global model with a resolution of 80 km that includes emission sources of the Denver area. To account for the fact that the measurement site is situated on a mountain, we sample the third (CAMSRA) and fifth (IFS(Chem)) model levels, respectively, to remove artifacts due to the different vertical and horizontal resolutions of the various datasets. In this way we make sure not to confuse the comparisons with signatures of local emissions – which indeed play an important role due to grid size. In future model configurations the model resolution should be increased to better capture more regional effects as encountered here. We now modify the text accordingly: "A signature exists at NWR from the chemical processing of polluted air masses from the Denver region during summertime (McDuffie et al., 2016), where all members show similar positive biases, where sampling accounts for the elevated station height."
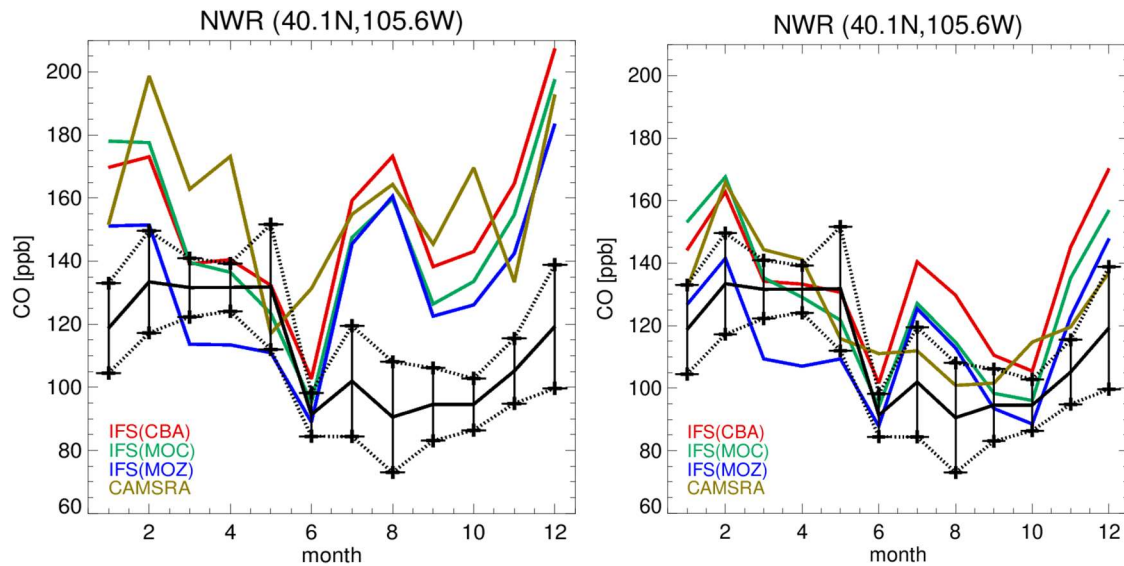
Left Panel: Adopting the surface level for interpolation Right Panel: Taking level 5 above surface for IFS and level 3 for CAMSRA.

Lines 775-783: This is a good discussion but it would have been great to perform a couple of sensitivity experiments using just one of the chemistry configurations to understand the impact of conversion efficiency on N_2O_5 hydrolysis.

This would be a point of interest to further investigate global variability, possibly for different years and locations where observations permit. But funding and resource limits mean that such studies are not currently feasible or planned.