Response to Dr Knowland: We thank Dr. Knowland for her comprehensive review of our manuscript and for providing a detailed guide towards proposed changes. This has led to significant updates with respect to domain definitions, selection of figures, and discussion of quantitative differences between modules, in the updated version of the manuscript. We provide detailed responses to the points raised below.

The manuscript is very long, with 9 tables and 18 figures within the main text. There are 6 tables and 10 figures in the supplemental, which are often referenced in the main text in a manner that should be included as "results" and not just available to the reader if they are curious to learn more. The authors need to take time to restructure the manuscript to bring out of the supplemental the tables and figures that are used in the paper for scientific discussion (Sections 4 and 5) and the conclusions (Section 6). This may require being more creative in the presentation of the figures. If there isn't new information, not all profiles or timeseries need to be shown; for example, instead make selections of only the ones that drive home the point the authors are trying to make, or maybe some of the months for a campaign can be combined.

We agree that the manuscript is rather long, but this paper will act as a future reference with respect to the nitrogen cycle in the chemical forecast model which has not received such attention before and requires co-located high quality datasets to assess performance, necessitating an in-depth assessment. We do not feel moving everything to the main manuscript will improve the readability and want to show as many comparisons as possible rather than removing information that may prove useful for the community in the future. To improve on the repetition aspect, we remove panels from the O_3 and NO_2 AirNow surface comparisons whilst keeping some form of regional representativity, whilst including bias and correlation statistics for all of the US regions defined in new tables.

I could not discern a clear metric for how the authors were measuring the improvements. Often there were statements that the differences were "significant" or "agreed well" that are subjective not quantitative. Is the main goal of the paper to identify if a configuration is suitable for the next version of CAMS?

The main purpose of this study is to quantify the ability of the respective atmospheric chemistry modules to describe air quality forecasts for a region outside Europe affected by anthropogenic pollution, without the application of data assimilation. This study therefore provides a benchmarking of such a configuration and provides a quantification of the uncertainty due to the model chemistry employed. Moreover, analysing model performance outside the European domain, where the model has not been tuned, provides more confidence with respect to CAMS forecasts which are used by US partners. The wealth of freely available observational data for the US (ozonesondes, surface measurements, aircraft campaigns) allows quantification of model performance for this region.

In this study the CAMS ReAnalysis (CAMSRA) is used as a reference, as it is a widely used and relatively well-known dataset. Nevertheless, we realize that interpretation of differences with respect to CAMSRA cannot be attributed to model chemistry updates alone and should be considered with care. Finally, the configuration tested here, though interesting from a chemical point of view, will not be identical to the one used for (future) operations where the horizontal resolution used in this study is lower and data-assimilation of tropospheric species is not applied. A separate, operational activity will take place to define, test, verify, validate and move over to a next version of the operational system, taking into account not only changes to model chemistry, but also emissions, meteorology, transport and data assimilation. We have now added a sentence regarding the use of CAMSRA, thus : "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." .

As I am based in the US now, I have observed the increased use of CAMS forecasts in the US, such as the inclusion in The Weather Channel for air quality forecasts. The reason this study focused on evaluating the updated chemistry in their three chemistry modules used in a global model produced by the European Center over the region of the USA is not strong (because it is a polluted region and most evaluation has been done on more background environments) and completely lacking from the abstract. Europe and Asia are also "affected by high anthropogenic emission sources" and could have been studied. If, for example, the US was chosen because of the vast amount of free, publicly-available observations from state and national ground-based and balloon-based networks and multiple field campaigns supported by national government agencies (NOAA and NASA), which the

magnitude and variety of publicly-available observations cannot be matched by other regions/countries, including in Europe, state that.

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. We wanted to focus on a corresponding NOy analysis in order to explain the simulated tropospheric O_3 in the simulations, for which aircraft measurements are a unique source of collocated data allowing this analysis. No such measurements are available with such a frequency for either Europe or China.

If the IFS configuration can only be run by ECMWF and its partners, what is the impact of this paper for the community?

The reviewer raises a valid concern regarding the availability of the IFS configuration beyond ECMWF and its partners. In response, there is more nuance than suggested by the reviewer. Firstly, the products from the CAMS system are widely used, including users in the US. A critical assessment of the model against surface observation networks of the US has not been previously documented in such detail and is therefore of interest to the wider modelling community. Regarding the availability of the exact model version to the external community, this is indeed currently restricted. We provide a comprehensive document of the details of model simulations, which in principle allows those interested to model a similar configuration in terms of emissions, meteorology, and atmospheric chemistry.

Here we repeat, and provide further details on the availability of model code: the IFS meteorological model is available to research institutes as part of OpenIFS (currently cycle 43r3, i.e. the version also used for the CAMS Reanalysis, while an update to provide a more recent cycle is planned). The CB05 and CB05-BASCOE atmospheric model chemistry used in cycle 43r3 will be made available soon (Huijnen et al., in preparation, 2022), while updated versions of model chemistry (particularly as reported here), are expected to also become available along with future releases of OpenIFS itself.

In order to help the quantification of current model performance we have now included more bias statistics in the revised manuscript associated with the AirNow surface comparisons and make additional reference to the bias tables provided in the supplement as related to the aircraft comparisons.

ln[24] *Why 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.

ln [55]: We have now corrected the sentence as : “Note that besides the global system, there is also an operational suite of regional-scale models for providing timely AQF for Europe (Marécal et al., 2015).”

ln [62] : *Is this then a companion paper to how these updates performed globally in the background regions? Why not Europe and Asia too which are heavily polluted too?.*

For assessing model performance in the pristine background away from strong emission sources the referee is referred to the evaluation presented in Huijnen et al. (2019), which pertains to this current study. The reviewer

correctly addresses the need for assessments of model performance for other pollution regions, as well as for background conditions. Analysis of model performance over other polluted regions is expected to take place in future studies. For the European domain, this is less relevant to the user community as the CAMS regional system is more appropriate. For other regions of the world, any analysis and evaluation is somewhat limited by the availability and subsequent interpretation of observational data for evaluation. We have now added the following sentence to the outlook: "Evaluation of model performance for other important polluted world regions, such as Europe and East Asia, will be the focus of future studies."

ln [70]: We have now rewritten the sentence as : "while stratospheric ozone can be well constrained by a linear model combined with (O_3) data assimilation"

ln [96]: Apostrophe removed : Volatile Organic Compounds (VOCs)

ln [101-102]: Removed double definition for O_3/CO and placed reference within brackets.

ln [154] We have now changed the sentence to "… for organic species by defining a single separate generic tracer species."

ln[159] Reference now corrected.

ln{169] Font in Table changed.

ln[215] Double definition corrected.

ln[236] Font corrected.

ln[264] 'NC' means Not Considered, where a definition has now been included in the table heading.

ln [257] *is the goal to be similar to CBA?*

Although all chemistry modules were expected to include parameterizations for heterogeneous scavenging on aquated aerosol, the implementation choices are left to the individual modeling teams depending on their insight. The teams developing the IFS(MOCAGE) and IFS(MOZART) chemistry versions have so far chosen to adopt the settings as developed for IFS(CBA) , although the IFS(MOZART) configuration does not account for any update on ice and cloud droplets. We now change the text to: "...fully updated. It was chosen to make this parameterization fully consistent with the IFS(CBA) configuration).

ln [272] capitals removed

ln[276] Paragraph break removed

ln[285] model top?

Model top is situated at 1 Pa, please see https://www.ecmwf.int/en/forecasts/documentation-and-support/137-model-levels (last access 29th December, 2021).

ln[286] *Are there still realistic biomass burning that are satellite based? Is there meteorological data assimilation or is the model nudge to analyzed meteorology?*

No data assimilation is employed in the current model experiments, neither from composition nor from meteorology. The system is initialized daily at 0 UTC with meteorology from ECMWF operational analyses.

ln[286] what does this mean? Is it a free running model each day? Does the model only do chemistry every 30 minutes? How long does it take to run these models for a year like this and how many resources does it use? Is it cheap or computationally expensive to initialize the model every day? That might be beyond the scope of the paper or what the authors need to share but I am curious.

The model is a free running model which uses a 30 minute time step for the chemistry calculations. The reason for choosing this configuration of daily forecasts is because this is how the operational system is configured and applied in most configurations. It is technically possible to run longer, (monthly) experiments using nudged meteorology, which helps to reduce the expense of any simulation but offers less flexibility in terms of system configuration, and is less supported.

ln[289] *why this period? Any large driving meteorological patterns to consider in 2014/2015 like ENSO?*

As we have discussed above, we chose the years 2014/2015 as based on the large availability of measurements within the US, rather than based on a large-scale phenomenon such as ENSO.

ln[307] Double definition removed

ln [312] Now corrected.

ln[314] We have now clarified our sentence to "The variability in the emission estimates compared to previous studies are due to both the use of different data in the derivation of fluxes and trends in the annual emission estimates (2014 versus 2019).

ln [317] Acronyms now added.

ln[325] Figure 1 is now referenced in the text associated with describing the observations.

ln[340] Figure 1 has been modified thus: Stars are used for the locations of the ozonesondes and triangles added for the GFDL surface sites. All Canadian surface sites have been removed from the Figure. The regions defined for the aircraft measurements have been renamed.

ln[377] We now correct to ".. in order to avoid .."

ln(381] We replace the abbreviation EC with an abbreviation representing the various states covered in the measurement campaign, as suggested by Dr. Knowland

ln[395] *Firstly, isn't this manuscript supposed to assess trace gas concentrations of O_3, CO and NO_2 because the IFS produces Air Quality Forecasts. Why not show the urban and suburban too?*

It is well known that data from suburban and urban sites include roadside measurement stations which invoke short scale gradients which cannot be captured by a global model, even when run at a high resolution. Here small scale meteorology is affected by buildings, concrete and traffic flow, which is why specific small scale models for cities are employed. The chemical emissions are likely to be strongly affected by traffic speed and density, which is not accounted for in the model. Therefore, we choose rural stations only to address this.

Secondly, the reason the US was chosen is because most of the model evaluation is done in "pristine environments". How does selecting only the clean background sites line up with this reason?

We only use the word pristine in association with previous studies. In addition to the points discussed above, the reason the US was chosen was due to the current provision of forecasts to existing users and the wealth of independent measurement data available for the year and region for evaluation. The environments sampled here are not considered pristine, which are typically located hundreds of kilometers away from strong sources. For the rural stations this is limited to tens of kilometers.

ln [400] *'how do you know this to be true? No reference to anyone doing this before'*

The process we describe is for removing zeros or values which persist for hours which cannot be true. We feel that data screening is an important step towards conducting a valid comparison. The technique is part of the established 'evaltools' software package which is readily available on the web via https://opensource.umr-cnrm.fr/projects/evaltools (last access 29th December 2021) and refer the referee to the description and information available online.

We add an additional sentence: " We use version v1.0.3 of the evaltools statistical package available online (https://opensource.umr-cnrm.fr/projects/evaltools)."

ln [404] *Comment on use of the year 2014/2015 for surface analysis.*

Our aim is to show a complete annual cycle therefore we limit our comparisons to Jan through to Dec 2014. Moreover, our simulations only ran through till June 2015, thus we do not have a complete simulation for 2015. We add an explanation in the text : "For brevity we do not show any comparisons for 2014".

ln[406] We have renamed the regions in Table 9 as suggested by Dr. Knowland.

ln[436] We have switched the order of JJA and DJF.

ln[438] *How does the PBL factor into this calculation. How many levels are likely included below 1km?*

The choice of 1km is to place the focus on the surface distributions rather than the polluted boundary layer. Increasing the height over which averages are calculated dampens the spatial variability due to mixing in the lower troposphere, therefore not directly relevant to showing a relevant surface distribution. The 137 vertical grid employed means averages are for 20 levels (please see https://www.ecmwf.int/en/forecasts/documentation-and-support/137-model-levels; last access 03-01-2022).

ln[442] .. NOx and VOC emissions. Changed Colorado to Kansas/Nebraska

ln[446] The statement ".. due to reduced transport" has been removed.

ln [451] Remove second definition of OH.

ln[453] Replaced with "ozonesonde"

ln[455] On the request of Dr. Knowland we have changed the sentence to : "For JJA there are positive biases near the surface for Yarmouth (east coast), whereas for Trinidad Head (west coast) biases are smaller therefore the agreement is more favorable."

ln[456] Thank you for highlighting, we have now changed Texas to Alabama

ln[465] We feel that the different ranges in mixing ratio used for each season provides more information with respect to the divergence across simulations.

ln[466] We now change the figure legend to "situated in the US/Canada"

ln[471] We now move the definition of station acronyms to the main text and out of the figure legend.

ln[481] ' Topography' has been included.

ln[485] *why do you think this was good and others less so?*

We no longer make direct comparisons against CAMSRA due to the lack of assimilation in the other simulations and the use of a different cycle of the IFS which introduce differences not due to the chemical mechanism. Sentence has been rewritten to state that all simulations capture the observations well.

ln[486] Sentence rewritten to state that the station is at an elevated site.

ln[490] The tick marks between months in this figure have been removed and the comparisons shifted such that each point occurs in the middle of a month

ln[496] Corrected table number

ln[514] *Did I miss something, I thought the urban sites were removed and only rural background stations were used? Where should I find this?*

We now remove all discussion related to scenarios which are not shown and refer Dr. Knowland to our response above as to why only rural background stations were selected.

ln[565] Second definition removed

ln[571] Included Eastern Coast in text

ln[572] For clarity we have rewritten the sentence to "The distribution of CO for this season in IFS(CBA) shows signatures of increased pollution visible over large urban centers in e.g. California, Washington state and New York state, in line with figures presented for LT O_3 related to regional NO_x emissions (See next section)."

ln[580] We now add a reference with information of biomass burning activity for 2014, where large fires occurred during summertime in Southern Canada near the observational site LEF.

ln[582] *June spikes low in range with obs and then the second half of the year seems positive again. Seems odd, do the authors have any ideas on this?*

Given that we sample the model to take account of the elevated station height (see revised text), thus limiting the influence of regional surface emissions from Denver (see response to Referee #2), we think that variability is governed by long-range transport and consecutive uplift of emissions as seen by other investigators (McDuffie et al., 2016).

ln[590] The x-axis of this figure has been changed to show months of the year

ln[600] the use of brackets to avoid verbose sentences is commonly used in the literature therefore we disagree that we overuse brackets as we maintain the naming culture used throughout the rest of the text.

ln[606] We now remove reference to urban comparisons of the AirNow data.

ln[613] We use 50hPa to bin the measurements resulting in 7-10 independent values depending on the pressure range per campaign. We have included details in the Section 3.2

ln[615] The metric for defining whether vertical profiles are captured well is both the gradients and the inflection in the gradient in the lower atmosphere representing the limit between the boundary layer and the free troposphere. We now change the text to : ".. the shape of the vertical profiles is captured well (i.e gradients and the inflection between the boundary layer and free troposphere) with .."

ln[617] A reference has been added related to the fracking emissions.

ln[629] Sentence corrected.

ln[630] *This reads as an entire paragraph about figures in the supplemental*

We have now moved the aircraft comparison of CH_2O out of the supplement and into the manuscript itself and subsequently renumbered the figures.

ln[634] Brackets removed.

ln[683] Changed definition of ratio to Capital F

ln[686] The stations are at similar locations for some

ln[704] *Are urban comparisons provided elsewhere?*

We now remove reference to urban comparisons of the AirNow data. See responses above related to this issue.

ln[715] We now make a reference to the Supplementary Information.

ln[726] We have now included concentration brackets in the definition of $[NO]/[NO_2]$ and changed the ratio to S.

ln[747] We have now changed Section 5 into a subsection of Section 4 as suggested.

ln[750] Colour legends now added

ln[770] We have now changed the sentence to : For IFS(MOZ) no N_2O_5 conversion on cloud particles and ice droplets is assumed, unlike the other chemistry modules.

ln[784] Figure 16 pertains to $[HNO_3]/[NO_x]$ and not$[HNO_3]/[N_2O_5]$, which is indeed Fig. S10.

ln[799] Indent removed. Brackets introduced for the definition of R'

ln[804] Pearson R used

ln[839] and [843] These values are calculated during the production of the deposition figure. For brevity we do not show all values for different regions but give a general description in the text..

ln[855] Reference brackets fixed

ln[867] Concentration brackets are introduced in the definition of the ratio.

ln[871] now corrected from 0.2 to 2.

ln[876] *So pressure levels between 1000 and 850 hPa?*

The height of the boundary layer is determined by both time of day and underlying terrain, therefore we change the sentence to read "... good agreement in F for the lower atmosphere."

ln[879] ND acronym now included.

ln[890] changed to " in the ECMWF IFS global model". Capitals removed.

ln[902] now changed to " good agreement for the Pacific NorthWest and an overestimation towards the Eastern Coast." using the new regional definitions suggested above.

ln[904] US removed from text

ln[934] Ratios now declared as $[HNO_3]/[NO_x]$

ln[954] $[PAN]/[NO_2]$ now included

ln[961] Comma added.

ln[990] We know replace with 'ozonesonde'.

ln[994] Thank you for highlighting this omission. We now acknowledge NASA for the DISCOVER-AQ campaign.

Refs:

V. Huijnen, P. Le Sager, M Köhler, G. Carver, S. Rémy, J. Flemming, S. Chabrillat, T .van Noije. OpenIFS/AC Atmospheric chemistry and aerosol in OpenIFS cycle 43r3. In preparation for Geosci. Model. Dev., 2022.

McDuffie, E. E., Edwards. P. M., Gilman, J. B., Lerner, B. M., Dubé, W. P., Trainer, M., Wolfe, D. E., Angevine, W. M., de Gouw, J., Williams, E. J., Tevlin, A. G., Murphy, J. G., Fischer, E. V., McKeen, S., Thomas B. Ryerson, T. B., Peischl, J., John S. Holloway, J. S., Aikin, K., Langford, A. O., Senff, C. J., Alvarez II, R. J., Hall, S. R., Ullmann, K., Lantz, K. O. and Brown, S. S.: Influence of oil and gas emissions on summertime ozone in the Colorado Northern Front Range, J. Geophys. Res. Atmos.,121,8712–8729, doi:10.1002/2016JD025265, 2016.