The authors thank the reviewers for the opportunity to revise the manuscript. The comments of the reviewers are indicated point-by-point in the following text where we explain how we have carefully addressed each of them (our answers in blue text).

Anonymous Referee #1

Received and published: 26 August 2022

The authors compare the NO2 simulated by CAMS and observed by TROPOMI. The comparison shows better agreement in summer than that in winter. The finding about the vertical profile is very informative. The methodology and conclusions are sound. However, the authors seem to favor super long sentences, which makes it difficult for readers sometimes. I recommend rephrasing the long sentences thoroughly to make them more reader-friendly.

The authors thank the reviewer for reviewing the manuscript and for the useful suggestions for improving it. Efforts have been made in particular for improving the readability of the text.

General comments:

1. Section 3. The authors discussed a lot of details about the ensemble database. It is not very clear to me what has been used for comparison in this study, what will be upgraded in the near future, and what has been done by previous studies since all information was mixed. I recommend reorganizing this section.

   The text of section 4 in the revised manuscript is improved to describe better what the CAMS ensemble product is, how it has evolved in time and which of the CAMS products were used in this study.

2. section 5.5. It will be useful to compare the differences between ensemble vs tropomi and ensemble vs individual models.

   Elements of the comparison between the CAMS ENSEMBLE and TROPOMI as well as between the ENSEMBLE and the individual models are presented in figure 9 and table 3, as well as through figures 7 & 8, which include a representation of the model spread. By spread, we refer to the range of values provided by all individual models, i.e. the distance between the minimum and the maximum values. We consider however that a comparison between the CAMS ENSEMBLE and the individual models as such is beyond the scope of this work. Elements of such a comparison can be found (in interactive form) in: https://regional.atmosphere.copernicus.eu/evaluation.php?interactive=cdf

   or in the form of quarterly reports in the "validation of CAMS regional services" reports in https://atmosphere.copernicus.eu/publications

3. I recommend adding a table listing all products used for comparison in the manuscript and adding a brief description of those products.

   Quantities (and their nomenclature) used in the comparisons in the paper are presented in figure 1 and their description can be found in section 4.1. The text of the revised manuscript has been improved to clarify how each product is used. Moreover, the caption of figure 1 now points to the relevant section that contains the definitions of those quantities.

Specific comments:

1. line 30. The grammar seems incorrect for the 2nd Please check.

   Text has been improved in the revised manuscript.

2. line 40. Line 50. Those sentences are too long to read.

   The sentences have been improved in the revised manuscript to make them more readable.

3. Line 70. I don't see the reason to separate items 2 & 3 as two angles. Additionally, it is useful to point out that the vertical profiles are replaced in item 2. Otherwise, it is confusing for the readers why TM5 is mentioned here.

   The authors see the point of the reviewer, the text has been modified and now refers to two different "angles" instead of three.

4. Line 145. What is "compo"?

   "Compo" stands for composition to differentiate between the ECMWF meteorological model (IFS) and the online atmospheric composition model. The term is now explained in section 3 of the revised manuscript.

5. Line 149. Is it operational now?

   Yes, it became operational in October 2021. It has been corrected in the revised manuscript.

6. Line 204. I suppose the R in S5P-R represents regional? I suggest putting the name after the description directly. It is easier for the reader to link the name with the product.

   Indeed, R stands for regional, G for global and RG for their combination. The text in the revised manuscript has been revised to clarify this.

7. Line 275. I suggest commenting on the potential reason why TROPOMI cannot detect ship lanes here.

   The text in the manuscript actually mentions that "ship tracks are generally more prominent in the CAMS fields", thus does not imply that they are completely absent in the TROPOMI fields. It is now documented that TROPOMI can in fact not only detect ship lanes but also individual ship tracks (Georgoulias et al 2020 Environ. Res. Lett. 15 124037,http://dx.doi.org /10.1088/1748-9326/abc445) under certain favourable conditions i.e. stable, calm wind conditions with limited dispersion of ship plumes. It would however be beyond the scope of this work to investigate whether the prevailing conditions during the days shown in figures 4 and 5 were favourable in this respect. Potential reasons for the apparent difference between modelled and TROPOMI fields in this particular respect

include the inherent noise in the TROPOMI fields and unrealistically low dispersion characteristics of the modelled plumes.
The text in the revised manuscript now reads "Ship tracks can be seen in both TROPOMI and CAMS fields but are generally more prominent in the CAMS fields (e.g., in the golf of Biscay or the North Sea).", which in the authors view is more accurate.

8. Line 281. What is 1st day forecasts?

   As described in section 3, CAMS forecasts are 4 days long, so by 1st day forecasts we mean that we select only the first day of this longer forecast. Text has been adapted in the revised manuscript to refer to the "1st day of the 4-day forecasts".

9. Line 328. What is "process modelling"?

   This is a term used to refer to modelling that is based on the simulation of physical and/or chemical processes, as opposed e.g. to statistical modelling. The text in the revised manuscript now reads "From a process-based modelling point of view…" which hopefully makes it somewhat more clear.

10. Figure 10. What is "spread"? Do you simply mean NO2 column densities here?

    The definition for spread is given in section 5.5:
    "These spreads are calculated on the basis of the difference between the minimum and maximum values for these quantities as calculated by using any of the 7 regional models and can be considered as a measure of the uncertainty of the CAMS ENSEMBLE based columns."
    In the revised manuscript, the caption of figure 10 now points to that definition.

11. Line 513. Do the authors claim a new methodology for satellite-model intercomparison here? What is the improvement compared to Eskes et al. (2003)?

    The reviewer is correct that the wording of the manuscript has been somewhat misleading. Our work does not introduce a new methodology on a technical level, but instead proposes a scheme (outlined in Figure 1 of the manuscript) to present and discuss approaches for comparing modelled and satellite/observed atmospheric gas columns (TROPOMI NO2 in our case), as well as point to the indicated methodology in order to do so.
    The text in the conclusions section of the new manuscript now reads:
    "A methodological scheme was introduced to elucidate approaches for model-satellite comparisons which builds on the fact that relative differences can be made independent of t he prior profile shape used in the satellite retrieval.",
    which should make it clear that we are not suggesting a new technical method.

Anonymous Referee #2

Received and published: 9 September 2022

John Douros and co-workers report on comparisons between TROPOMI NO2 column observations and results from the 7 air quality models which are currently operational within CAMS, providing forecasts and analyses over Europe at 0.1x0.1 degree resolution. The comparison shows a reasonable agreement during summer, but a substantial (factor of about two) model overestimation in winter. The use of high-resolution a priori profiles from the CAMS model ensemble (instead of the global 1x1 profiles) in the tropospheric NO_2 column retrieval from TROPOMI results in higher retrieved columns over emission hotspots by about 30%. The authors further performed validation of the new satellite TROPOMI NO2 dataset using remote sensing column measurements and found that despite the overall overall bias reduction compared to the operational TROPOMI product, the new dataset is not able to close the large gap between observed and modelled NO_2 columns in wintertime.

The manuscript does not include significant advances in modelling and has quite limited novelty. It uses pre-existing models and their outputs which are routinely available. However, the study proposes an alternative TROPOMI NO2 dataset over Europe based on high resolution model profiles which could be useful for the community. The method used for this derivation has been already applied in previous studies. I find the comparison of the data with the output of the 7 models interesting, in spite of the fact that the reasons for the large mismatches are not investigated in the manuscript. The scientific approach and the methods are not new but they are valid and widely used in the literature. The results are discussed in a balanced way, although in most instances the discussion is only qualitative. The writing is not always very precise. The language should therefore be improved in the revised version. Some references are incomplete or not defined, and additional references are needed. I could recommend publication after the following points are adequately addressed.

The authors thank the reviewer for reviewing the manuscript, for the insightful comments and for the useful suggestions for improving it.

As regards the scientific merit of our work, our manuscript does not introduce a new methodology on a technical level, but instead proposes a scheme (outlined in Figure 1 of the manuscript) to elucidate approaches for comparing modelled and satellite/observed atmospheric gas columns (TROPOMI NO2 in our case), as well as point to the indicated methodology in order to do so. The wording used at certain places in the manuscript may have been ambiguous on this and has now (i.e. in the revised manuscript) been phrased differently to make it clearer. What we also consider important in this work lies in the fact that a comparison is performed utilizing the recommended approach, based on a sizeable collection of European operational regional air quality models, which provides insights into the state-of-the-art atmospheric composition modelling, especially above the surface. The more novel part is indeed the introduction of the alternative TROPOMI NO2 dataset over Europe based on the CAMS ENSEMBLE analysis, which is arguably the best available near real time modelling regional atmospheric composition product available for the European continent.

The revised manuscript contains improvements, including clarification of terms and wording to address most of the specific comments of the reviewer.

Comments:

l.8: "7 up to 11 models". Not precise and misleading since the manuscript presents only results from 7 models.

We agree with the reviewer, this has been removed and more care has been put to describe what the "7 up to 11 models" meant, in the main body of the text, section 3 in particular.

l.13: remove "quantitative"

Removed.

l.13: provide information (e.g. bias, correlation) about how close this agreement is

The agreement is now quantified in terms of relative difference between the TROPOMI and CAMS derived columns. The text now reads:

"In summer, the comparison shows an close agreement between TROPOMI and the CAMS ensemble NO2 tropospheric columns with a relative difference of up to 15% for most European cities."

l.14: 'significant discrepancy', provide numbers

The discrepancy is now quantified in terms of relative difference between the TROPOMI and CAMS derived columns. The text now reads:

"In winter however we find a significant discrepancy in the column amounts over much of Europe, with relative differences up to 50%."

l.25-28: here again provide figures of the bias reduction and correlation obtained from this validation

Quantitative information has been added, i.e. bias of the new product is 5% to 12% smaller compared to the standard (S5P) product.

l.34: 'values above the surface which are otherwise very scarce', replace by 'measurements at the surface which are very scarce'
Replaced.

l.36: read 'at kilometer scale'

Corrected.

l.39-44: This information does not seem relevant for this paper.

This is only a few sentences to provide context on the current state of efforts to monitor atmospheric pollutants (including NO2) from space, with the use of geostationary satellites in particular. This will become quite relevant for this work in the future we will be given the opportunity to compare with models on a near hourly basis. The text in the manuscript has been improved to better reflect this.

l.50: What are the CAMS systems? I would replace by 'CAMS makes'

Replaced.

l.51: Inness et al. 2019b is not defined

The reviewer is correct, now fixed.

l.53: 'consistent' appears twice in the same line, replace by "to daily (re)analyses of concentrations and emissions which are consistent with…"

Corrected.

l.57: changes are not sharp for pollutants other than NO2, see https://doi.org/10.1029/2020GL091265, https://doi.org/10.3390/atmos12080946, DOI: 10.1126/sciadv.abg7670. I suggest to drop 'sharp' from the sentence and add some more references.

The general remark about the sharp decreases was based primarily on the extensive review by Gkatzelis et al (2021) which covers a wide range of pollutants and relies on various kinds of observations, not only satellite based. This is of course a complicated subject which to a large extent defies such generalizations, i.e. the observed changes are found to be inhomogeneous in time and space and have not always been visible in satellite retrievals. In any case NO2 indeed exhibited the sharpest drop. “sharp drops” has been removed, the text has been adapted to reflect the above and two extra references have been added.

l.58-60: poor wording, Replace 'dedicated studies have been launched to study' by 'dedicated studies have been performed to investigate'

Replaced.

l.62: near daily basis

Replaced.

l.72: TROPOMI appears twice, replace 'measurement series' by 'measurement period', mention that TROPOMI NO2 is derived using the global TM5-MP profiles

The text has been adapted taking into account the reviewers suggestions.

l.73: mention clearly the horizontal resolution of the CAMS and the TM5 models

The exact horizontal resolutions of CAMS and TM5 are mentioned in subsequent sections (i.e. section 4.2, table 2). We thing it might be too early to go into this level of detail already from the introduction.

l.71-75: improve the clarity

As mentioned, the text of this paragraph has been improved.

l.81-82: remove 'very small', replace 'very large' by 'high'

“Very small” removed, “very large” replace by “wide” which seems more appropriate in this context.

l.84: provide references for your statement

The ability of the TROPOMI instrument to identify power plants, highways and ships is documented in various works which have now been added in the bibliography (also below).

Goldberg, D. L., Lu, Z., Streets, G. D., de Foy, B., Griffin, D., McLinden, A. C., Lamsal, N. L., Krotkov, A. N. and Eskes, H., Enhanced Capabilities of TROPOMI NO2: Estimating NOX from North American Cities and Power Plants, *Environ. Sci. Technol.* 53, 21, 12594–12601, 2019, https://doi.org/10.1021/acs.est.9b04488

Miyazaki, K., Bowman, K., Sekiya, T., Jiang, Z., Chen, X., Eskes, H., Ru, M., Zhang, Y., and Shindell, D.: Air quality response in China linked to the 2019 novel coronavirus (COVID-19) lockdown, Geophys. Res. Lett., 47, e2020GL089252, https://doi.org/10.1029/2020GL089252, 2020.

Liu, F., Page, A., Strode, S. A., Yoshida, Y., Choi, S., Zheng, B., Lamsal, L. N., Li, C., Krotkov, N.A., Eskes, H., van der A, R., Veefkind, P., Levelt, P. F., Hauser, O. P., Joiner, J., Abrupt decline in tropospheric nitrogen dioxide over China after the outbreak of COVID-19. Sci. Adv.6, eabc2992, 2020, https://doi.org/10.1126/sciadv.abc2992

Georgoulias, K. A., Boersma, K. F., van Vliet, J., Zhang, X., van der A, R., Zanis, P. and de Laat, J., Environ. Res. Lett. 15 124037, 2020, http://dx.doi.org/10.1088/1748-9326/abc445

l.85: remove 'the paper by' here and throughout the manuscript

Removed.

l.85-93: check your references, for example Eskes et al., 2021a is missing

Fixed.

l.97-98: 'to force the stratosphere to be consistent with TROPOMI', weird statement

We agree with the reviewer, this has been changed to “… the modelled characteristcs of the stratosphere…”.

l.108-114: 'do not have a large impact', 'rather stable', 'considerable change', provide quantification

More details on the quantitative differences between the TROPOMI products produced with the successive versions of the level-2 processor can be found in the next paragraphs of the manuscript (lines 118-135) but also in (mentioned as van Geffen et al, 2021b in the original version of the manuscript, Geffen et al, 2022 in the revised version):

van Geffen, J., Eskes, H., Compernolle, S., Pinardi, G., Verhoelst, T., Lambert, J.-C., Sneep, M., ter Linden, M., Ludewig, A., Boersma, K. F., and Veefkind, J. P.: Sentinel-5P TROPOMI NO2 retrieval: impact of version v2.2 improvements and comparisons with OMI and ground-based data, Atmos. Meas. Tech., 15, 2037–2060, https://doi.org/10.5194/amt-15-2037-2022, 2022.

as well as in:

http://www.tropomi.eu/data-products/nitrogen-dioxide/

l. 121: MAXDOAS or MAX-DOAS, not both

We now use MAX-DOAS throughout the revised manuscript.

l.120: mention that the Verhoelst et al. comparisons do not account for averaging kernels

The reviewer is correct, this is now mentioned in the revised manuscript.

l.124: reference missing

Reference to Lambert et al (2021) has now been added in the revised manuscript.

l.135: could you mention the impact of the new version v2.2 described in https://doi.org/10.5194/amt-15-2037-2022 ?

Indeed, van Geffen et al (2022) argue that "on average the NO2-v2.2 data have tropospheric VCDs that are between 10 % and 40 % larger than the v1.x data". This is now explicitly mentioned in the revised version of the manuscript. Further analysis on the impact on the impact of version 2.2 (and above) of the processor is included in the supplement.

l.146-48: link not accessible (and too long)

Link is functional but is indeed long and is thus in more than one line, which makes it not easily clickable. A solution for this can hopefully be found at the final editing stage of the manuscript.

l. 153: correct typo

Corrected.

l.164: 'have', not 'has'

Corrected.

Figure 1: Acronyms are not explained in the caption.

Providing explanation in the caption for all quantities mentioned in the figure would not be very practical, thus a comment has now been added in the caption to point to the part of the manuscript where the quantities/acronyms are defined.

l.179: not necessary

Removed.

Section 4.1: I find this section describes well-known methods in a confusing way.

The reviewer is correct that the content of this section follows closely Eskes et al. (2003). The main aim of this section is however to introduce the acronyms/naming conventions used in the manuscript (and figure 1) in a natural way, so that the reader does not need identify the correspondence between these and the reference. For that reason, the authors thought it would be useful to include elements of the methodology introduced by Eskes et al. (2003). In any case, the section is improved in the revised manuscript to clarify its purpose.

l.225-26: avoid repetition of the word 'gridded' in the same line

Removed.

Sections 5.1, 5.2, and 5.3 could be merged, all of them consist in briefly presenting the figures 5-8

Sections 5.1 and 5.2 have been merged in the revised manuscript as both of them indeed discuss maps (figures 5 and 6). Section 5.3 was retained as it discusses a qualitatively different result, the time series of NO2 columns at various cities and European regions.

l.325: I could not find Huijnen et al. 2010b in the list

Fixed.

Table 3, add additional columns with the ratio S5P and S5Pcams and CAMS-RG-A. Or add another table. This would help your discussion.

The authors thank the reviewer for the suggestion. Instead of ratios an additional column with the relative difference between S5P and CAMS-RG-A has been added which is useful as a way of quantifying the agreement between TROPOMI and CAMS derived columns.

l.455: did you use 8 or 9 MAX-DOAS stations for validation? In the abstract you mention 8

We used 9 MAX-DOAS stations. The abstract has been corrected in the revised manuscript.

Section 6.1, the discussion is again only qualitative. For example, 'CAMS is higher close to he suface': higher by how much?

The reviewer is correct that there are occasions in the text were remarks are only, or mainly qualitative. While in many cases we tried to complement these remarks with quantitative information, discussion of figure 14 (mean averaging kernels and NO2 profiles by the various models), is purposefully qualitative as it is basically used as a means to explain the preceding figures/results, some of which may seem counter-intuitive at first glance, e.g. how complementing the free tropospheric part with data from the global model can lead to lower columns (S5P-RG being generally lower to S5P-R).

Fig.15 inset statistics are too difficult to read

The figures have now been redrawn at higher resolution which should make these easier to read when zooming in. The preference of the authors would be to keep all statistics in the figures as the increase in dispersion along the RMA regression line mentioned in the discussion can more easily be seen with the standard deviation, but much less with the 1/2IP68.

Fig.16: Is S5Pcams and S5P-RG the same thing?

Indeed, they are. The naming has been unified in the revised manuscript (only S5P-RG is used).

l.500: 'this is not done here', improve the wording

The sentence has been improved in the revised manuscript.

l.564: '10% column enhancement', is this on average?

The reviewer is correct, the text should have read "a column enhancement of at least 10%". Now corrected in the revised manuscript. This is with the exception of Helsinki which exhibits quite low S5P-RG/S5P ratios.