The authors thank the reviewer for reviewing the manuscript, for the insightful comments and for the useful suggestions for improving it. Here are some replies on the comments provided.

Our manuscript does not introduce a new methodology on the technical level, but instead proposes a scheme (outlined in Figure 1 of the manuscript) to guide interested parties as regards recommended 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 been phrased differently to make it clearer.

What we consider important in this work lies in the fact that the comparison is performed using 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.

Some answers to specific comments:

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 remark about the sharp decreases was based on the extensive review by Gkatzelis et al (2021) which covers a wide range of pollutants and relies on various kinds of observations. It is true however that these changes are not always visible in satellite retrievals.

1.84: provide references for your statement

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

Daniel L. Goldberg, Zifeng Lu, David G. Streets, Benjamin de Foy, Debora Griffin, Chris A. McLinden, Lok N. Lamsal, Nickolay A. Krotkov, and Henk Eskes
Environmental Science & Technology 2019 53 (21), 12594-12601, DOI: 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.

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

Aristeidis K Georgoulias et al 2020 Environ. Res. Lett. 15 124037, http://dx.doi.org/10.1088/1748-9326/abc445

1.73: mention clearly the horizontal resolution of the CAMS and the TM5 models

This information is contained in table 2.

1.108-114: '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 manuscript):

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.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 NO\textsubscript{2}-v2.2 data have tropospheric VCDs that are between 10 \% and 40 \% larger than the v1.x data". This is explicitly mentioned in the revised version of the manuscript.

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

The main purpose of this section is to introduce the acronyms/naming conventions used in the manuscript. For that reason, it was necessary to resort to elements of the methodology introduced by Eskes et al. (2003). The section is restructured in the revised manuscript.

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

Thanks for the suggestion, an additional column with the S5P to CAMS-RG-A ratio is included in table 3.

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 qualitative remarks should be accompanied by include quantitative information. Wherever possible, these are addressed in the revised manuscript. Discussion of figure 14 (mean averaging kernels and NO2 profiles by the various models) however, is qualitative as it is basically used as a means to explain the preceding figures/results, some of which may seem counterintuitive 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).

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

This is not accurate, the text should have read “a column enhancement of at least 10%”. This is with the exception of Helsinki which exhibits quite low S5P-RG/S5P ratios.