**Response to review**

We thank the reviewers for their time spent to help us improving the paper with their useful suggestions. Below we provide a point-by-point response to the reviews and descriptions how we have addressed the suggestions. The original review text is shown in black and the response in blue.

**RC1**

The authors present a detailed comparison of NOx emissions derived from the Dutch national inventory and DECSO emissions. While the work appears scientifically sound, the manuscript would benefit from significant improvements in writing.

General comments:

1. The authors state, "we therefore present a detailed comparison of the Dutch national emission inventory and emission data generated by DECSO with the goal to verify the national emission inventory that forms the basis for emission policies. To that end, a high resolution version of DECSO-HR was developed and a new version DECSO 6.5 is introduced." While the introduction provides some context, it is unclear why DECSO-HR is specifically required to compare the Dutch inventory with DECSO. Providing additional background to clarify this point would strengthen the introduction. Furthermore, the final paragraph of the introduction contains a summary of the paper's major findings, including statements like, "The national inventory and DECSO-HR 6.5 showed good agreement of spatial emission patterns with a general agreement in total annual and monthly national emissions. On the other hand, the respective emission trends showed deviations in the years 2020 and 2021 and there were local deviations at locations with high emission intensities." Summarizing results at this stage is premature; such content is better suited for the conclusion section.

We have edited the introduction to clarify the need for a higher resolution. The following text was added to the introduction:

*"One challenge in using inverse modelling for emission verification in particular for smaller countries like the Netherlands is the relatively coarse resolution for most inverse modelling setups, which is limited by computational cost and the resolution of satellite observations and meteorological parameters. The resolution is especially crucial for analysing emissions on a local scale, e.g. for individual cities or industrial installations, and for attribution of emissions to different processes."*

We furthermore removed the summary of the results from the final paragraph of the introduction.

2. The authors assert that validating bottom-up emission inventories with satellite-derived emissions is not a widely established practice, framing this as a key motivation for the study. However, I disagree with this claim. For example, the U.S. National Emissions Inventory (NEI) has been extensively validated using emissions inversion techniques. The main novelty of this work appears to lie in the development and application of DECSO-HR compared to the original DECSO. As such, the motivation would be better framed around the introduction of DECSO-HR and using Dutch emissions as a case study. The result session should also include a subsection to discuss the

improvements/changes introduced in DECSO-HR, along with a more thorough examination of the uncertainties associated with these innovations.

We apologize for this oversight and added references to U.S. verification studies and rephrased the introduction accordingly. It should be noted that the fundamental algorithm of DECSO and DECSO-HR are identical with the main difference being the input used for the modelling. While DECSO uses TROPOMI super-observations, DECSO-HR uses the TROPOMI observations directly, with otherwise only some small adjustments. To make this more clear, we added additional details about DECSO-HR to the methods. Also the introduction and abstract were slightly rephrased to put more emphasis on DECSO-HR. In our perspective, the most relevant innovation in DECSO-HR is not the improved resolution in itself but the detailed comparison to the emission inventory it facilitated. While we do agree that other studies on emission verification of national air pollutant inventories exist, to our knowledge there are few studies that go into this level of detail, which we see as the innovative aspect of this work.

The following text was added to the method section:

*"Besides small changes concerning the harmonizing the parameters related to distance/resolution (for example a smaller time step and a shorter distance for the trajectory calculations) the main difference between the standard DECSO and DECSO-HR is their input. The TROPOMI observations used for DECSO-HR are more numerous and have higher uncertainties than the super-observations used in DECSO. On the other hand, the direct observations are better spatial matched with the emissions."*

3. Section 3.2. The discrepancies in trends observed between bottom-up emissions and DECSO inventory from 2021 to 2022 are intriguing. Although the authors list several plausible contributors to this discrepancy on the bottom-up side, I wonder whether the fixed lifetime of NOx assumed in DECSO-HR is a significant factor. It may also partly explain larger differences for high-emission-density grid cells discussed in Section 3.3. A sensitivity analysis exploring the uncertainty of the fixed NOx lifetime is strongly recommended to provide clarity.

It may have not been clear in our original text that there are actually two $NO_x$ lifetimes involved in the calculations of DECSO. First of all, there is the intrinsic lifetime of $NO_x$ in the CHIMERE model. Secondly there is a lifetime calculated for the trajectory calculations determining the sensitivity matrix that determines the Kalman gain in our Kalman Filter approach. This second lifetime can slightly affect how fast the emissions are updated to the 'real' level according to the observations. The computational cost of the lifetime increases with the improved spatial resolution of DECSO. The derived daily emissions are based on minimizing the difference between satellite observations and the model output (i.e. concentrations), which is affected only by the $NO_x$ lifetime within CHIMERE. To balance the optimal update speed and the computational cost, we choose a fixed lifetime for the Kalman Filter approach. We have also checked the effect of the fixed lifetime by comparing final monthly emissions with and without fixed lifetime and found this to be negligible.

We have added to the text:

*"This lifetime is only part of the trajectory calculation within DECSO in addition to the intrinsic $NO_x$ lifetime of the CHIMERE model. While the former lifetime determines the convergence speed of the emission updates, the magnitudes of the emissions are mainly impacted by the intrinsic $NO_x$ lifetime of the CHIMERE model, since the derived daily emissions are based on*

*minimizing the difference between the model output and the satellite observations. The difference of the final monthly emissions with and without fixed lifetime in the DECSO trajectory calculation is negligible."*

4. Section 3.3. The uncertainties associated with wind information are not accounted for in the analysis. Incorrect wind data could potentially explain the mismatched emission locations, considering the spatial resolution of wind data is coarser than TROPOMI. Performing an additional sensitivity analysis would be valuable in addressing this concern.

We agree with the reviewer that the uncertainty in wind data can potentially explain the mismatch in emission locations. Since the uncertainty in the wind fields are probably mainly manifested in random errors in its direction, this will result in small mismatches of emission locations each day. Averaged over a month this will lead to a gaussian distribution of the emission location instead of sharply located point source. To reduce this effect, we only use the observed $NO_2$ plume within 50 km from an emission source. Our study shows that the satellite-derived emissions still suffer from the spatial smoothing of emissions, which we will try to mitigate in the further development of DECSO.

Specific comments:

1. line 93. The statement, "Note that the a-priori emissions are not from an external inventory but based on the emissions of the previous day," is confusing. I assume that the emissions inventory for the "first" day relies on an external inventory. I don't quite get how the a-priori emissions are not this external inventory. Perhaps rephrasing this sentence for clarity would help avoid confusion.

To minimize the effect of the prior emissions of day one, a spin-up period of 200 days of assimilation is applied. Any effect of the inventory on the first day (which were taken from HTAP v3) is gone after this time because in our assimilation scheme we update the emissions by addition instead of multiplication of the a-priori emissions. We rephrased the sentence to:

*"Note that the a-priori emissions are not from an external inventory but based on the emissions of the previous day which are updated based on the mismatch between observations and forecast with a spin-up period of 200 days before the emission results are used. After this period any effect of the day one a-priori emissions is gone due to the additive assimilation scheme. As prior emissions for day 1, HTAP v3 was used (Crippa et al. 2023)."*

2. Figure 2 and 4. I recommend using an alternative color to highlight Schiphol airport's runways, as the current color choice makes it difficult to identify them within the plots. Improving the visual distinction would enhance clarity and accessibility for readers.

We changed the color of the symbols to improve clarity.

**RC2**

I found the paper interesting and well written.

We thank the referee for their positive assessment.

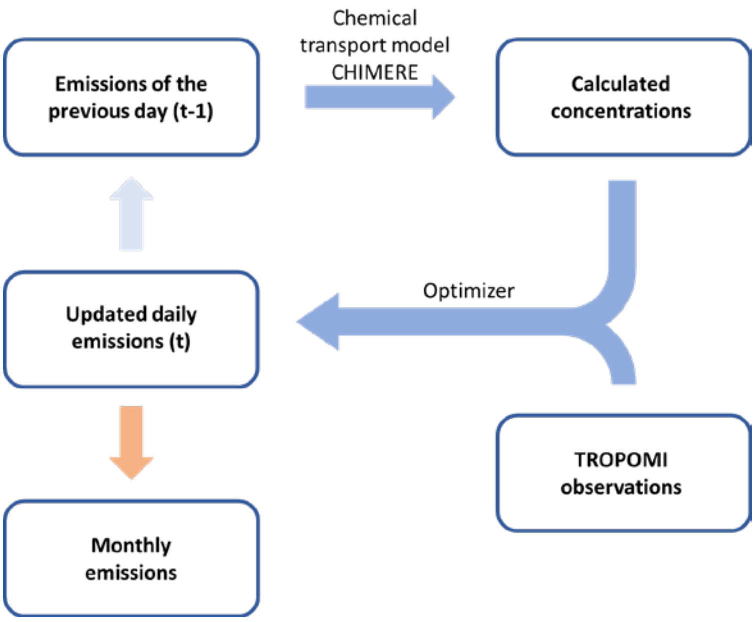
However, below I list some points that I would like to see being addressed, before publication:

- section 2.5 deals with uncertainties of the different dataset used. DECSO has uncertainty at 8%, the Dutch inventory at 19%...could you please elaborate more on how you came out with such numbers? 8% seems to be very low, and I would like to better understand from where this value comes from

We apologize for leaving out the source for the uncertainty analysis of DECSO. A detailed description of the error calculation is given in Van der A et al. (2024). The reference was added to the text and the text was edited for clarity. The stated uncertainty is the precision calculated from the error propagation of each emission estimate while taking into account the temporal correlation.

- I think a graphical scheme of how DECSO works would help the readers (and also me). DECSO is well-known, but a graphical scheme would help the reader to grasp the main features of the proposed approach

A graphical illustration of the DECSO assimilation scheme was added to the manuscript as figure 1 and is also added below.




- the authors present the Dutch inventory (section 2.2) and CAMS-REG-ANT (section 2.4). As far as I know, CAMS-REG-ANT is based on the officially reported (Dutch) data but with different gridding...but not 100% sure about this. Could you please better elaborate on the difference between these 2 inventories?

Indeed, CAMS-REG-ANT utilizes the reported national total emissions and uses a consistent method between countries for gridding of the emissions. We currently work on a detailed comparison of CAMS-REG-ANT, the official reporting and DECSO. We added a small explanation of the differences to the method section. Note, that in this work, CAMS-REG-ANT was used only for emissions outside of the Netherlands.

- in the paper the DECSO, then the Dutch inventory and also CAMS-REG-ANT are briefly described and used for further analysis. But which is the final conclusion of the authors? One should use the Dutch inventory and simply use DECSO for validation? or use DECSO directly, and for example use the Dutch inventory to split in sectors? Please elaborate on this

Each of these emission data sets was developed for different use cases and each have their individual pros and cons making them each suitable for the respective use case, for example monitoring of international policies in case of the national inventory. So we cannot give a general recommendation as depending on the use case, the recommendation for which data set to use would be different.

Since all these data sets are approximations of the same real-world emissions they should be consistent with each other and their comparison can reveal their shortcomings and inform the development of the different data sets.

We added the following text to the conclusion:

*"These two emission data sets were developed for different use cases and each has its individual pros and cons making them each suitable for the respective purpose, for example monitoring of international policies in case of the national inventory. Nevertheless, as these data sets are approximations of the same real-world emissions they should be consistent with each other and their comparison can reveal their shortcomings and inform the development of the different data sets."*

- could you please explain if the presented approach can also be used for other pollutants, also i.e. in view of future satellite sensors that will be available in the future?

The DECSO algorithm is currently also used to estimate ammonia emissions from satellite observations (Ding et al., 2024). In principle, it would also be suitable to be adopted to emissions of other short-lived air pollutants, for example sulfur dioxide. While $SO_2$ emission levels in the Netherlands are presumably too low to be estimated based on satellites, these approaches are under development to be applied elsewhere. Another big innovation will be utilization of the observations from Sentinel 4, which will give multiple observations per day and will allow to analyze the diurnal temporal pattern of emissions.