Comments Editor (DOI:10.5194/egusphere-2024-632-EC1)

This paper thoroughly discussed and investigated constructing superobservation and the relevant uncertainty of the satellite NO2 data. However, it has serious drawbacks in writing and format. While addressing and responding to reviewers' comments and suggestions, I request that the authors also improve the manuscript to meet the standard of scientific writing. Please also see some other comments as follows.

We thank the editor for their feedback on further improving the writing and format of the paper. We have done some rewriting and restructuring to improve the readability of the paper. Below the answer to your comments. All line numbers in the answers refer to the new manuscript.

Line 73: Please briefly mention what are the disadvantages pointed out by Purser et al. (2000).

We have added the following on line 98-100:

"but Purser et al. (2000) points out two disadvantages with this method: Firstly the superobservations are not independent from the assimilation system and secondly, creating superobservations requires a statistical description of the forecast system, which is not always available"

Line 85: An unclear sentence.

We rewrite the sentence on line 111-112 to improve readability:

"Miyazaki et al. (2012a) and Boersma et al. (2016) average the observations with the overlap of the observation footprint with the superobservation grid as weights."

Line 87: (Inness et al. 2019b)

We have corrected this typo (line 115) to: "(Inness et al. 2019b)".

Line 165-172: Unclear description. Please revise these sentences with clear definitions. Is ym a simulated observation? Please clarify what will be used for obtaining x and xa.

We replace line 201-202 with the following text to improve clarity:

"Here x is the tropospheric NO_2 vertical profile from the model co-located in space and time to the footprint of the satellite and x_a is the a-priori vertical profile used in the retrieval."

Line 176: Please use a formal style for the section title.

This section has been removed due to restructuring, thus the section title is removed.

Line 196: I can't follow the argument that "given the same uncertainty, low NO2 observation force assimilation more than high observations". Did you mean low NO2 observation will have small uncertainty if the uncertainty (measured in percentage) is proportional to the column amount?

This section is removed due to the restructuring. The content can now be found in the introduction on line 85:

"If all individual observations with their individual uncertainties are assimilated in a model with a coarser resolution than the satellite, this leads to low-biased analyses, because more weight is given to low observations with a small uncertainty. With the superobservation approach described in this paper, such persistent low biases are largely avoided."

How to derive the weights shown in Fig. 2?

> The weights are the area overlap between the superobservation grid cell and the satellite footprint. We add the following to the description of the figure 2 to clarify:
>
> "The colours indicate the weight w_i, which is the area overlap (km^2) between the superobservation grid-cell and satellite observation footprint."

Figure 3 is shown, but I can't find the relevant discussion.

We add a reference to the figure on line 220:

"An example of this method is shown in figure 3."

It is not clear why Eqs. (8) and (9) are discussed. Did the authors want to explain how the superobservation affect the calculation of innovation?

> For properly comparing (super)observations to a model you need to apply the satellite kernel to the model. Thus making superobservations not only requires calculating a representative observation, but also the corresponding averaging kernel. We make this more clear by adding the following in line 232:
>
> "To compare superobservations against a model we also need a corresponding averaging kernel, which are averaged in the same way as the observations. Multiplying Eq. 1 with w_i and summing over the satellite observations we get:"

(11): Please clarify how to obtain the correlation factor, c. What is the value of c in this study?

The correlation factor c depends on the type of uncertainty. It set to 1 for the stratospheric uncertainty, set to 0 for the slant column uncertainty and defined by a correlation length for the AMF uncertainty. We make this more clear by changing lines 265-268 to:

"As mentioned in section 3.2, the superobservation uncertainty depends on the observational uncertainties and their correlation c (Eq.11). As shown in equation 2 the tropospheric column uncertainty consists of 3 separate sources of uncertainty: The stratospheric uncertainty, the slant column uncertainty and the air mass factor uncertainty. The superobservation uncertainty of these components is calculated separately because they have different correlations, which means their uncertainty propagates differently. Every component and its correlation is discussed individually in the sections below. "

Line 532-540: Section 7.1 needs significant revision. The experiment configurations should be provided with clear descriptions.

We add extra information on the experiment configurations in line 516:

"The assimilation is run with 32 ensemble member and an assimilation window of 2 hours. The localizations are based on a species-dependent localisation scale. These are derived from sensitive tests in Miyazaki et al. (2012b). Covariance inflation is achieved through the inflation of emission factor uncertainties, by inflating the spread to a minimum predefined value. Additionaly, a multiplicative covariance inflation of 7% is applied to the concentrations. The details of the assimilation approach that is used are described in Miyazaki et al. (2020b). In addition to NO_2, we also assimilate total column CO from the TIR/NIR band of the Measurement of Pollution in the Troposphere instrument. (MOPITT) (Deeter et al., 2017), the SO_2 planetary boundary layer vertical column from OMI (Li et al., 2020), and Aura Microwave Limb Sounder (MLS) O_3 and HNO_3 profiles (Livesey et al., 2022). To demonstrate the impact of different superobservation settings, the following 4 sensitivity runs were done for July 2019, only varying the NO_2 observations:"