Summary

The goal of this study is to enhance the supraobservation technique, which aggregate satellite remote sensing NO_2 observations, to better align with the model resolution. The authors carefully quantify and formulate the observation error budget, which includes the error caused by the total slant column fitting, the stratospheric column, the AMF calculation, and the spatial representativeness. The author also conducted several data assimilation experiments to examine the benefits of using superobservation, concluding that the proposed superobservation method can lead to the lowest forecast error compared to other approaches.

As a result, the superobservation approach presented in this study is important to the research community and fits within the scope of GMD, and this algorithm has the potential to be extremely useful for preprocessing high spatiotemporal resolution trace gas observations from geostationary instrument. Overall, though the topic is important, there are several issues that need to be addressed before considering its publication.

General comments

Many abbreviations in the manuscript are not clearly defined (e.g., TROPOMI, TEMPO, GEMS, LER/DLER, GFED, JAMSTEC, etc.). I would recommend authors to double-check and make adjustments to the manuscript.

Line 205-208: The statements presented here are statistically correct, and the authors do not need to change them. I just want to suggest another way to think about the negative NO_2 column. As the author noted in the manuscript, the appearance of a negative NO_2 column is primarily due to limitations in the retrieval algorithm, which could be improved in the future. The biggest issue, in my opinion, is that negative NO_2 columns have no physical meaning, and they should actually correspond to small positive NO_2 columns in non-polluted areas. Taking this into account (i.e., non-negative trace gas concentration), a “negative bias” may occur when averaging negative and positive pixels.

Section 6.2: It is quite impressive to see the author put in so much effort to quantify the representative error that result from data coverage in a superobservation grid, and I really think this section is the gist of this study. However, the writing in Section 6.2 is disorganized and difficult to follow. I would encourage the authors to enhance the writing in this section, particularly the explanation/derivation of effective population size.

The level 2 product also reports NO_2 tropospheric column uncertainty (nitrogendioxide_tropospheric_column_precision). Does this term contribute to the error calculation in this study? I didn't see any discussion about it; am I missing something?

Specific comments

Line 195-197: Given the same observation uncertainty, the Kalman gain only depends on background error. As a result, low and high NO_2 observations should have the same Kalman gain, contradicting your claim that low NO_2 observations force more in the assimilation than high NO_2 observations and yield low-biased results. Could authors elaborate on this?

Figure 3: This figure is provided in the manuscript, but it is not mentioned in the content.

Equation 13: Please define Θ and Θ_0.

Line 309: Please define GFED and provide a reference for this fire emissions inventory.

Line 360-362: Is there other dataset that might be utilized to estimate spatial correlation lengths? I ask this because the version difference may not always be available.

Line 372: Could you clarify which variable you referred to as "uncertainty due to the AMF, as estimated by the retrieval" in the level 2 product?

Section 5.3: As albedo has seasonal variation, I am wondering if the AMF error spatial correlation length also has a significant seasonal change.

Section 5.3: What is eventually used for the AMF uncertainty is not clear. Is the uncertainty coming from the version difference or from the retrieval product?

Line 396-397: Is the green line just the average of the gray lines in Figure 8?

Line 421-423: It would be beneficial to further explain the calculation of fractional coverage (*f*), for example, by utilizing equations.

Line 438-439: "At 50% coverage, the increase in RE is 54% for clean areas and 263% for polluted areas." Figure 8 does not support this statement. This sentence likely describes the inflation of representative error due to partially data coverage in Figure 9. Please consider moving this sentence to the right place.

Equation 18: This equation is provided but doesn't mention in the manuscript.

Line 459: Could the author clarify what “sigma normalized RE” means here?

Line 485-490: I am confused with these sentences. Did you mean that you first estimated σ using Equation 16 and then computed the final representative error using Equation B5? Could author clarify on it?

Section 6: Many regional chemical data assimilation (DA) systems employ a finer horizontal grid (~ 10 km or less), which is comparable to the scale of a satellite pixel. In this scenario, the proposed method for estimating representative error may not be effective because the pixel population (N) becomes too small, leading to the use of the climatology value instead. I was wondering if the author could comment on the extent to which we should consider the representative error.

Line 494-495: Where did the intersection come from? I believe it is from Figure C1. Is that correct?

Line 497: “uncertainty^2” Please fix the typo.

Line 509: Please correct typo in this sentence for the figure citation.

Line 515-516: It is acceptable to retain all the descriptions here. I just want to comment on the fact that the random sample approach is not a very good method for data thinning given the large superobservation grid. A better way is to first analyze the distribution of all pixels inside a superobservation grid and then select one pixel that is closest to the mean or median. This method could pick up a more statistically representative data than a random sampling.

Section 7.1: In light of the previous comment, it is not surprising that the thinning experiment demonstrated the worst performance in DA. A data-thinning experiment is too easy to outperform as a large discrepancy between model and superobservation is expected. Keeping this experiment in the manuscript is totally fine. I would like to encourage the authors to run one more experiment in which the superobservation (error) is just a simple average of all pixels' NO_2 columns (reported error) that falls within the superobservation grid. Given that this approach is commonly used in many DA or data analysis studies, comparing the method proposed in this work to the “simple average approach” could help to highlight the benefit of the using advanced superobservation method. I recognize that it may be difficult for the author to run additional simulation, thus this is not required. Alternatively, it would be wonderful to see some discussion of the simple average method versus a more complicated superobservation method in the manuscript.

Table 1: Why is the χ^2 value for the uncorrelated error case (111.0) much larger than in other experiments? Could author verify this?

Table 1: The RMSE for fully correlated error and superobservation cases is similar. The χ^2 value for the fully correlated error case is actually closer to one, which makes it difficult to conclusively say that the superobservation case outperform the fully correlated error experiment.

Line 583-584, Please change the “o-f” to “O-F”.

Section 7.1: The error budget analysis (Figure 11) shows that representative error is not the dominant source of the observation uncertainty. As the computation of representative error seems to be very costly, did the author have a chance to examine the impact of including or excluding representative error on the DA results?

Section 7.1: I appreciate that the author conducted a thorough comparison between different DA experiments. Since the RMSE of many DA experiments are comparable, I would recommend finding an independent dataset for model evaluation to better understand the performance of individual experiments and determine which experiments better capture the real NO_2 fields. An observing system simulation experiment (OSSE) might also be useful in future work to test the performance of different superobservation techniques, since the ground truth is known.

Line 600: Please define JAMES.

Line 625: I am confusing about using correlation 0.15 for remaining observation error. This contradicts what is stated in Section 5.2, where the correlation is determined as a function of distance/correlation length, and each grid may have a different value for the AMF uncertainty. Could the author please clarify this?

Line 637: “According to Nyuist”. It would be recommended to add a reference here.

Line 640-644: The spatial correlation imposed on prior emissions and concentrations may be incorrect, and they are also updated by the DA process. As a result, using the predefined spatial correlation to justify the use of superobservations is not a good claim in my opinion.

Line 676: Please correct the typo in the equation.

Equation A1: Please define x, i, j, and n in the equation.

Line 712: The intercept in the figure is +2.5, not +3 mol m^{-2} written in the manuscript.