This manuscript addresses an interesting feature of sensitivity of NOx emission inversions to a priori profile shape assumptions in AMF of satellite NO2 columns. Authors conclude "As the difference between the simulated profile shape and the a priori profile shape increases, so do the corresponding assimilated emission errors". In the discussion section, however, the authors indicate that the adjoint inversion is less sensitive to vertical representativeness errors in cases where emissions are poorly constrained. It is noted that choice of AMF will become increasingly important to adjoint inversions as emission inventories improve. The manuscript delivers some new and intriguing messages to satellite and air quality modeling community. I think the manuscript is well written. The introduction and revisit of AMF and averaging kernel are neat and helpful. There are some parts that need further investigation and explanation. Hope the authors revise and improve the manuscript before final publication.

Thank you for your comments. We address the specific comments below.

Major point Line 304-310: the manuscript deals with the "truth" emissions that deviate only 5% from the original anthropogenic NOx emissions. Here, it is written that the tests using random 15% or 30% perturbations to emissions were insensitive to the AMF. In real cases, NOx emission inventory errors are quite large (> 30%). Do the authors mean that choice of a priori shape factor is not important for most of real emission study cases? Please show the results from the random 15% or 30% perturbation tests (or other new cases if possible) and discuss more on applications to the real world problems (e.g., Qu et al., 2017).

Results from the random 30% perturbation tests are now included in the Results section.

We now clarify in the conclusion that our results show that recent emission study cases are likely insensitive to the AMF (Line 330):

"This indicates that while the adjoint cost function is mathematically dependent on the AMF, the inversion is less sensitive to vertical representativeness errors in cases where emissions are poorly constrained, as is the case in recent adjoint inversion studies (e.g. Qu et al., 2017). However, choice of AMF will become increasingly important to adjoint inversions as emission inventories improve."

Minor points 1. Examples of inconsistent a priori shape factor: I do not think these days retrieval groups use SF_BL, SF_Trop type a piori. In the abstract, up to 80% increased error is based on this choice. I am not sure if readers need to take this number seriously.

We clarify in Section 3.2.2 that the SFBL and SFtrop tests are extreme cases that do not represent typical retrievals (Line 264):

"The SF_{BL} and SF_{trop} tests do not represent any modern retrieval algorithms, but are used as extreme examples of an *a priori* that assumes no spatial variability."

We also changed the value cited in the abstract to 30%, representing the error of the more realistic test cases.

2. Line 274-276: I am not sure what these mean.

We have replaced these statistics with root mean square differences, which more clearly describe the difference between the *a posteriori* emissions from the SF_M and the other tests.

3. Line 288-292: Examination of the mathematical frameworks behind two common methods for comparing simulated and retrieved columns highlights how the method introduced by Palmer et al. (2001) facilitates separation of observations sensitivity (scattering weights) from the profile shape (shape factor) enabling the model-retrieval comparison to be independent of a priori profile assumptions. In the last part, model-retrieval comparison to be independent of a priori. . .It is confusing because the main conclusion of the manuscript is that the model-retrieval comparison is not independent of a priori (in certain cases).

The mathematical framework outlined in Section 2 shows that indeed any model-retrieval comparisons is dependent on the prior. One conclusion in the manuscript is that the 4D-Var assimilation process, while mathematically dependent on the prior, is not sensitive to the prior under some conditions. We have clarified this by adding Table 2 which indicates how the model-retrieval comparisons are affected by the *a priori* profile, and by adjusting the discussion to clarify this distinction (Line 326):

"Inversion tests performed using synthetic observations based on random 30% perturbations to emissions were insensitive to the AMF, despite large differences in *a priori* vertical column densities. In these tests, the adjoint cost function was more sensitive to the larger difference between the observed and simulated slant columns (i.e. $\Omega_{s,m} - \Omega_{s,o}$ in Eq. (13) and (19)) than to AMF. This indicates that while the adjoint cost function is mathematically dependent on the AMF, the inversion is less sensitive to vertical representativeness errors in cases where emissions are poorly constrained…"

4. Line 307: Add "s" in the subscript.

This has been corrected.

5. Line 315-Line 320: It is good to emphasize these again. But I believe that retrieval groups are already doing this. It might be good to mention various data supported by the retrieval groups.

Some retrieval data products do include scattering weights, but not all of them. However, we feel it is best to emphasize this as a general recommendation without singling out any particular retrieval groups.

6. Can the posteriori NOx emission difference (%) for highly polluted cases be shown in Table 2? This type of results will also be useful for higher perturbation cases.

Percent differences were not included in Table 2 for clarity. However, we now include mean "true" emission values in the captions to allow the reader to make such comparisons if desired.

7. Is the number of iterations of 4D VAR assimilations for all the test cases the same? How many iterations are requited for the tests?

We now mention the number of 4D Var iterations on Line 198: "Tests performed here required 20-30 iterations to minimize the cost function."

8. Is there a possibility that model spatial resolutions affect the results?

The mathematical framework outlined in Section 2 presents the main point of the paper, which is that consistency between simulated profiles and shape factor profiles used to calculate AMFs is essential. This is true regardless of model resolution. We now include a test where the *a priori* profile is generated by a higher resolution model (GEOS-Chem run at 2x2.5 resolution). This test further supports the main conclusion of the paper. From Line 316:

"The SF_{finer} test indicates that using a higher resolution model to generate *a priori* profiles does not provide an advantage in simulation-observation comparisons, as consistency between the simulation profile and the AMF shape factor is of greater importance."