We thank the reviewer for the time and review. Your feedback has improved our manuscript.

In their paper, X. Liu and co-authors discuss the impact of future geostationary NO2 observations on air quality models and the potential of these observations to improve estimates of emissions.

General assessment: The highlight of this paper is a documentation of the sensitivity of NOx emission estimates on the quality of the meteorological fields and the winds in particular. As shown, these issues are especially critical when the resolution of the observations and the models advances to the km scale. This sensitivity is demonstrated in a clear way, and to my opinion is an important aspect to keep in mind for the development of future regional analysis systems. As such, I am in favor of publishing these results in ACP.

The paper is well written, has a good introduction, but the overview of the method C1 and tools is very short and condensed. The implementation of the OSSE setup is not provided with enough detail for other groups to repeat the experiment, see my detailed comments below. I therefore would urge the authors to expand these parts of the paper.

We have made major revisions in section 2~4. In section 2, we provide detailed description on the assimilation system including the forecast model, DART facility (with localization settings), initial and boundary condition ensemble, emission update scheme and synthetic meteorological and chemical observations. In section 3, we redesign the assimilation experiment.

In order to judge the importance of the meteorological uncertainties for emission estimates, it is important to provide estimates of real-life uncertainties in state-of-the-art regional weather analysis systems, and to compare these uncertainties with the OSSE setup. This would be a valuable addition to the conclusions section.

We agree that it's important to compare the identified wind uncertainties in this work with state-of-the-art regional weather analysis. The available atmospheric reanalysis products (https://climatedataguide.ucar.edu/climate-data/atmospheric-reanalysis-overview-comparison-tables) provide a dynamically consistent estimate of the climate state at each time step. These assimilation scheme and models ingest all available observations every 6-12 hours and produce global and regional reanalysis at mesoscales (above 32 km). The products with such assimilation frequency and spatial resolution do not provide analysis uncertainties at the fine scale as in this work, which is hourly assimilation at the resolution of 3 km. In this work, we downscale NARR (32 km) to initialize the WRF forecast and use WRFDA to generate the ensemble. WRFDA system uses a state of the art method for generating the ensembles/quantifying the meteorological uncertainty. It uses the NMC method to generate the static error covariance and then uses the static error covariance to perturb the meteorological fields. Since the WRF meteorological forecast errors are comparable to other regional system, we would expect the uncertainties of the associated ensembles to be representative and reasonable. So we revised the text as follows:

"For meteorology ensemble, random perturbations were added to each member by sampling the NCEP background error covariance using WRF Data Assimilation System (http://www2.mmm.ucar.edu/wrf/users/wrfda)."

Since we redesign our experiment to assimilate meteorological observations hourly, the meteorological uncertainties becomes small enough to allow chemical observations to constrain emissions successfully in our results. Then we analyze whether the wind uncertainties are important for emission estimation in REA. Based on our result in REA run, we conclude as following:

"Last but not least our results show that the information of wind *uncertainties* is not important for NO_x emission estimation. Assimilations of NO_2 only with the meteorology constrained from an hourly weather assimilation product would perform as well as the joint assimilation of meteorology and chemistry."

It seems the authors make very optimistic assumptions on the uncertainty of the TEMPO retrievals. Is this justified and are the conclusions sensitive to the choice made for the observation error?

In ensemble assimilation method, the adjustment on the emissions depends on the relative magnitude of TEMPO observation uncertainties and the prior emission uncertainties. We agree that the uncertainty of TEMPO retrieval in this study is optimistic compared to current OMI retrieval. However, previously we have conducted experiments with TEMPO observation of 30%. We find the emission estimation results not promising because the observation uncertainties are too large to let the observations influence the emissions. In addition, the prior emission uncertainty is set to be 30% in our OSSE. This can also be very optimistic given the dispute on the accuracy of bottom-up inventory. In real application, with a reasonable estimation of prior emission uncertainty which will be higher than 30%, it is highly possible that more weights will be given to TEMPO observations to adjust the emissions. The choice of TEMPO retrieval uncertainty in this study is limited by the estimation of the prior emission uncertainties. In real application, even if the reported TEMPO uncertainty is too large to adjust the emissions, the conclusions of our OSSEs on the optimal approach of NO_x emission estimation will not change.

Last line p2: 3.3 to 5 molecules/cm^2 ?

We change the text as, "biases of 34% (3.3 to 5.0 $\times 10^{15}$ molecules/cm^2) are found in the modeled average NO_2 column over Los Angeles at resolutions of 96 km compared with 12 km."

p3, l27: "idealized profile setting provided by WRF-Chem." This is unclear and needs to be explained.

In our updated experiments we initialize WRF-Chem using the global chemical transport model outputs that will provide a realistic initial condition for chemical species. We modified the text as, "We use the global chemical model output from MOZART to initialize the chemical simulation on the outer domain and to provide the chemical boundary condition."

p4, top: Please discuss the state augmentation approach for NOx in more detail: how are the emissions perturbed as compared to the inventory, and how is the ensemble constructed.

Compared with the NEI 2011 which is referred as the truth, the CR emissions are perturbed by adding a 30% low bias uniformly in temporal and spatial scale. This is used as the ensemble mean of emissions in the assimilation process to construct the emission ensemble. Each ensemble member is then generated by adding a perturbation to the ensemble mean. The resulting emission ensemble in this work shows ~35% uncertainty presented by the ensemble standard devition. The correlation between emission perturbations of two grid point is modeled by a simple isotropic exponential decay function with a characteristic correlation length of 50 km.

We discuss the state augmentation approach for emissions in section 2.4 in detail. We described the ensemble construction in section 2.3. To better explain the emission perturbation, we added the text as below:
"We consider the NR as the true atmosphere and sample meteorological and NO_2 observations from the NR. The control run (CR) is a parallel model calculation to the NR and suffers from imperfect model input and parametrization. The differences between the NR and the CR in this study are the emission inputs and the initial conditions for the meteorology. … . In all experimental runs, we bias the CR initial emissions to be 30% below the reference model and examine the ability of the assimilation to recover the reference emissions."

p5, l1: "We chose the 10 km distance based on sensitivity experiments" How are these experiments done?

We changed the text to the following: We perform univariate experiment by tuning the localization distance from 5km, 10km, 20km to 50km and keeping other settings in the system unchanged. By comparing the posterior emission RMSE of REF (described in section 3) of all experiments, we find the smallest RMSE with the localization distance of 10km.

p5, l26: "We calculate a layer dependent Box-Air Mass Factor (BAMF) : : : follow the latest version of the NASA : : :" It is unclear from the text how this is done. If the authors compute this themselves, then which RTM is used? How is the geometry computed?
What about the terrain parameters. Please provide these details.

We didn't calculate BAMF by running a RTM directly as it is not required when we follow the method in (Bucsela et al., 2013). "The scattering weights are computed and stored a priori in six-dimensional look-up tables (LUT) generated from a RTM. The six LUT parameters are solar zenith angle (SZA), viewing zenith angle (VZA), relative azimuth angle (RAA), terrain reflectivity (Rt), terrain pressure (Pt), and atmospheric pressure level, (p). ". We modified the text as below:

"The geometry related parameters (SZA, VZA and RAA) are computed hourly for each TEMPO observation using Matlab functions sun_position.m and geodetic2aer.m with inputs of the location and time of each TEMPO observation, and the location (36.5°N, 100°W) and altitude (35,786 km) of the TEMPO sensor. The terrain reflectivity and terrain pressure are sampled from the WRF-Chem nature run (NR, see section 3) for each TEMPO pixel."

p6, l3: "The other parameters are sampled from the model run". Which parameters?

Other parameters indicate terrain reflectivity and terrain pressure. We clarified this in the updated text.

p6, l14: The mean uncertainty of 7.5 % seems very optimistic. Is this justified somehow?
Would the results be very sensitive to this choice?

We agree that it is very optimistic to set the TEMPO observation uncertainty as 7.5%, which is the lower bound of current observation error for OMI NO_2 tropospheric column product. This paper focuses on the optimal approach for NO_x emission estimations from TEMPO. It is difficult to separate systematic and random contributions to the uncertainty in TEMPO measurements. Future experiments should explore variations in systematic and random uncertainties in the measurements.

p6, l29: "scaled to be 70%" A uniform scaling is very idealized, and may be more easily recovered in a DA system than spatially-varying emission perturbations. It seems logical to add also a more random

perturbation in emissions to reflect the uncertainty in the emission spatial distribution. Would this impact the conclusions?

The conclusions in this work will not change by using a spatial-varying emission perturbation. We agree that a uniform emission perturbation is idealized. Success in recovering emissions with a uniform scaling can't not guarantee the success to recover emissions with spatially-varying perturbations. However, the approach that didn't work in the ideal case will fail in the realistic scenario with spatially-inhomogeneous emission perturbations. In our new experiments, we show the shortcomings of some emission estimation strategy in the scenario of ideal emission perturbation. We added the following text in section 5.

"In our OSSE, we start with an ideal case of a priori emissions, which is spatially-uniform 70% of true emissions. Success in recovering emissions with a uniform scaling can't not guarantee the success to recover emissions with spatially-varying perturbations. However, the approach that didn't work in the ideal case will fail in the realistic scenario with spatially-inhomogeneous emission perturbations. The degradation of emission estimation in the ENS.2 and ENS.3 will still hold for correcting emission errors under more complex circumstances, such as spatially- or temporally- varying emission errors. For future work, further analysis is required to examine whether spatially-varying emission errors can be reduced under the condition of low wind RMSE."

p7, l29: Please discuss explicitly the formula to compute the uncertainty. In line 31 the RMS is defined, but "uncertainty" is unclear.

We added the text as follows:
"We also analyze the uncertainty of the prior and posterior estimates. The uncertainty is expressed by the 1-σ standard deviation of the ensemble."

p8, l11: "initial uncertainty of 41.70 mol/(km^{2}hr)". What is this absolute number? Is it an average over the red domain in Fig. 1?

The sentence is removed in our updated discussion section 4.

Fig.3: It would be helpful to indicate with symbols at what times an analysis is produced (for the green and blue curves).

Since we updated the experiments in this paper, the original Figure 3 is removed.

p8, l18: Please explain how this correlation is computed.

We removed the discussion of correlation in our updated paper.
P10, l3: "future TEMPO NO2 observations will enable us to constrain surface emissions on a city scale". I would claim that this is not fully demonstrated by the authors, because the OSSE setup is still idealized. In particular, the same model is used to construct reality (the nature run), and emissions are idealized, which will lead to too optimistic results. Please formulate the conclusions in a more careful way.
Please also report on typical wind uncertainties which are within reach in present-day high-resolution regional meteorological analysis systems. This may be compared with
RMS and uncertainty values as reported in table 2. Is a performance like reported for e.g. ENS-H feasible in reality?

We clarified our point as follows:

"We test the emission constraints from TEMPO NO_2 observations in an ideal case assuming no errors associated with the modeled meteorology. In the experiment of joint assimilation of meteorology and chemical NO_2, we find that the emission estimation is successful in the morning but degrades in the afternoon when the prior wind RMSE grows above 1 m/s. Considering the dependence of errors in estimated emissions on the wind forecast errors, we recommend to guarantee the accuracy in modeled wind with RMSE below 1 m/s for the success of chemical assimilation to infer emissions at the scale of our model grid."

Reference:

Dee, Dick, Fasullo, John, Shea, Dennis, Walsh, John & National Center for Atmospheric Research Staff (Eds). Last modified 21 Jan 2016. "The Climate Data Guide: Atmospheric Reanalysis: Overview & Comparison Tables." Retrieved from https://climatedataguide.ucar.edu/climate-data/atmospheric-reanalysis-overview-comparison-tables.