We appreciate the referee's comments, which have improved our manuscript, as detailed below:

General comments: The focus of this paper is to test the ability of estimating emissions of NO2 with data assimilation and future high resolution satellite instruments by doing an Observing System Simulation Experiment (OSSE), since the temporal and spatial variability of NO2 are high and hourly observations at a resolution of at least 5km would be required. I think the title should include "using OSSEs", otherwise it suggests that real data experiments have been carried out.

We updated the title as "Assimilation of satellite NO_2 observations at high spatial resolution using OSSEs"

The abstract and the introductory section are excellent. One of the main experimental conclusions is that the assimilation of meteorology and of NO2 should be done simultaneously, supporting Kang et al. (2012) who pointed out the need to do simultaneous assimilation of the meteorology and atmospheric CO2 and thus succeeded (in an OSSE framework) in estimating accurately the surface carbon fluxes even in the absence of any a priori information.

Sections 2-4 are not written clearly enough, in my opinion, at least not for a person who is not very familiar with the details of the subject, like this reviewer. In addition, much of the results revolve around an unrealistic experiment: having the TEMPO NO2 observations at very high spatial and temporal resolution, and at the same time assimilating the winds every 12hr or even 24hr, which, not surprisingly, ruins the results. I would suggest to replace this with a more realistic experiment: Assume that you have access to a meteorology data assimilation performed every hour, so that you have access to the mean winds but not to the uncertainties. Are the wind uncertainties essential? Experiments that we have performed suggest that the wind uncertainties are not so important for CO2, but they may be for NO2 because of the smaller scales involved.

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. First, we incorporated MADIS weather observations to increase the volume of wind observations. This enables us to preform hourly joint assimilation of meteorology and chemistry in the run named ENS.1. Following the reviewer's suggestion, we have REA assimilation run in which the meteorology is extracted from an existing reanalysis. REA reinitializes the meteorological state every hour with the best estimate of meteorological states generated by ENS.1. The result discussion and conclusion are also updated accordingly in section 4.

My recommendation is that this paper is important and could be accepted after major revision. Detailed comments:

P3, line 28: "After a spin-up time of 40 hours on the outer domain, the inner domain simulation is initialized and constrained through one-way nesting in both meteorology and chemistry". It is not clear whether this is in addition of the data assimilation, and for how long it is done.

This part only introduces the domain setup and how simulation on the outer domain serves for the inner domain. There is no assimilation on the outer domain. We describe the experiment details in section 3. The inner domain simulation is initialized on 06:00 am on 2014/07/02 and assimilation starts on 10:00 am 2014/07/02 (P8 line5). We extended the model spin-up time and modified the text as below:

“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. After a spin-up time of four days on the outer domain, the inner domain simulation is initialized and constrained through one-way nesting in both meteorology and chemistry.”

P4 line 1: "We use the RADM2 gas phase chemical mechanism for its simplicity (Stockwell et al., 1990)" I don’t know what this means.

We provide descriptions of RADM2 chemical mechanism as below.

“We use the widely-used regional acid deposition model version 2 (RADM2) as the gas phase chemical mechanism. There are 59 species and 157 reactions to represent both inorganic and organic chemical reactions under tropospheric conditions. It includes the chemical losses of NOx through reaction with OH radical to form nitric acid, and other NO_x sinks as peroxyacyl nitrates and alkyl nitrate. RADM2 predicts peak NO 2 concentrations and the timing that the peak occurs very well when the mechanism is tested against the representative environmental chamber experiments.”

Line 12: Since Kang et al. (2011, 2012) were the first to estimate surface fluxes of carbon as evolving parameters, perhaps they should be referenced.

We modified the text as follows in section 2.4:

“By including emissions in the ensemble state vector, emissions are estimated as hourly evolving parameters. Estimation of time-evolving emissions using data assimilation is first presented in carbon fluxes estimation (Kang et al., 2011, 2012). Such an approach provides emission information more than a monthly mean, or an average for a specific time period. For NOx emission estimation problem, emissions in cities show significant variation from urban to suburban. The observed columns show strong spatial variation dominated by emission hotspot as a result of the short lifetime. Combined with high-resolution observations, the goal in this work is to constrain a time-static emission error as we perturb emissions of each hour in the same way.”

Line 19-24. This is an interesting persistence approach to estimate the emission forecast model. I am not sure whether it will always work in the presence of error spin down or if the surface fluxes vary substantially. It would be good to show how well it works in this case, perhaps under varying surface fluxes.

In all model runs, the surface fluxes are varying from hour to hour. In the OSSE, the emission bias in the control run is -30% uniformly though 24 hours. We agree that in reality, the errors in emissions could vary from hour to hour. So there may not be such a constant relative bias. However, the simplicity of this assumption makes it useful for this initial trial. We hope that other experiments with alternate assumptions will emerge in future research.

P5 line 6: “showed” not “suggest”

We changed to “showed”.

Lines 13-14: Are these simulated observations obtained from NARR?

The simulated meteorological observations are sampled from the nature run (NR, see in section 3). We described the simulated MADIS observations in detail in section 2.5.

Line 24: Does you model changes in the "truth" in the simulated TEMPO observations or are the fluxes fixed?

The NO_x emissions are specified by NEI 2011 dataset. This emission provides hourly-varying emission for a typical weekday in summertime. The fluxes does not vary from day to day. NEI 2011 is used as the emission input for nature run without any perturbation. We added this information in section 2.1:

"Anthropogenic emissions for WRF-Chem are from the National Emission Inventory (NEI) 2011 for a typical July weekday 30 at native 4×4 km resolution. The NEI 2011 provides hourly-varying emission for a typical weekday in summertime. The emissions do not vary from day to day."

In section 3, we added as:

"The original NEI 2011 is used as the emission input for the NR without any emission perturbation. We begin by creating a NR and a CR simulation on the outer domain of 12 km resolution (d01) without assimilating observations using a simulation setup as described above in section 2.1. We impose a difference to the CR by using emissions in the CR that are scaled to be 70% of the NR emissions."

P6 line 10: It would be good to show a companion figure showing the diurnal cycle.

The diurnal cycle of TEMPO NO_2 column is now shown in Figure 4.

Line 14 "a mean uncertainty of 7%, which is lower than the 35%..." unclear.

This sentence is not in the revised paper.

Line 19: AMF: I couldn't find its definition. In general there is a profusion of use of abbreviations whose original definition is difficult to find in the paper. I can deal with Control run (CR) and Nature Run (NR) but then new capitals appear (AR) defined somewhere else, and I get confused. If all the acronyms were defined in the same place it would be OK, but my pdf reader, when I search for the definition of AR, gives me hundreds of words that contain "ar".

We defined box- air mass factor as BAMF previously, but not AMF. We thank the reviewer for pointing out this issue. We provide the definition of abbreviations in the text as follows.

"Finally, as TEMPO is expected to be operational no sooner than 2018, it is reasonable to expect the retrieval error dominated by air mass factor (AMF) in polluted region will be reduced as a result of future improvements in AMF simulation"

We removed AR which confuses the reader. We define nature run (NR) and control (CR) as below:

"The original NEI 2011 is used as the emission input for the NR without any emission perturbation. We consider the NR as the true atmosphere and sample meteorological and NO2 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."

P7 line 24: "the 17:00LT assimilation cycle each day" suggests that the assimilation may have been done for many days, and I really don't know yet whether this is true, or just one day was simulated (with constant true emission?).

In the updated experiments, we extended the assimilation to three days longer. We modified the text as below:

"We run assimilation experiments from 10:00 LT 2014/07/02 to 18:00 LT 2014/07/05 with an assimilation window of one hour."

Along these lines, it would be really strange for a real life researcher to use detailed TEMPO observations every hour and then do atmospheric assimilation only every 12 or even 24 hours. The experiments ENS-O and ENS-T are not realistic (and have horrible acronyms). I would drop these experiments. The results shown in Figure 3 are very obvious: when you assimilate wind observations the uncertainty in the winds becomes much smaller, and the transport errors likewise.

As indicated above, we redesigned the assimilation experiment as described in section 3. First, we incorporated MADIS weather observations to greatly increase the volume of wind observations. This enables us to preform hourly joint assimilation of meteorology and chemistry in the run named ENS.1. Also we add ENS.2 to determine whether or not concentrations should be updated when observations are assimilated to constrain unobserved emissions. In ENS.3, we use the meteorology of the next day to initialize the CR ensemble so that there is some difference between the CR ensemble mean and the NR in the meteorology. REA reinitializes the meteorological state every hour with the best estimate of meteorological states generated by ENS.1. We hoep the reviewer finds the strategy and acronyms an improvement.

P8 line 7 The acronym of "No meteorological observations assimilated" should NOT be called "BIAS-MET", the errors are much more than a bias!

We removed this run with no meteorological observations assimilated in the new experiment design.

Fig 3: Label "Local time (hours)"

We added "local time (hours)" in Figure 3~6.

Fig 4: Labels: "Longitude" "Latitude" in the figure, not in the figure caption.

The original Figure 4 is no longer in our updated paper.