Answers to Referee #2 comments: ***Review of Impact of physical parametrizations and initial conditions on simulated atmospheric transport and CO_2 mole fractions in the US Midwest***

We thank the referee for the helpful comments that will improve the manuscript. In the text below, we have tried our best to respond to all the general and specific comments provided by the reviewer.

Comments to Author:

This paper presents the evaluation of multiple WRF model configurations over the US Upper Midwest. The model configurations are constructed by selecting different PBL, cumulus and microphysics schemes. These configurations are also used to simulate atmospheric CO_2 mole fractions by using the CO_2 fluxes from the Carbon Tracker global inversion system. It is important to use well constrained meteorological drivers in CO_2 inversion studies. As the previous studies have shown, uncertainties in the meteorological drivers can lead to large discrepancies in simulations of atmospheric CO_2 mole fractions.

This study is trying to identify how much different physics parameterizations contribute to error in CO2 transport within the WRF-CO_2 model. I have some reservations on the setup and conduct of the modeling exercises and evaluations in the paper:

REF-C1: The WRF-Chem CO_2 model needs more explanation. How different is this model from the regular WRF-Chem model, which can simulate passive tracers including CO_2?
Author-C1: The WRF-Chem CO_2 module has been developed following the passive tracer scheme of WRF-Chem, modified to include multiple tracers (from 10 to 25 tracers currently). The physics is identical to the passive tracer original module, using the turbulent mixing in the Planetary Boundary Layer and the mean wind fields for advection and diffusion elsewhere.

REF-C2: How vertical mixing of CO_2 is parameterized in this model configuration? Does this version of the model include convective tracer transport?
Author-C2: In the WRF-Chem version used in this paper, the only convective parameterization available is based on precipitation rates to diagnose the vertical mass flux (1D) with a pre-defined cloud base and cloud top. We are currently developing the coupling between 3D mass fluxes from the Kain-Fritsch scheme with the chemistry CO_2 module in order to properly account for convective transport. This work is still ongoing. In this paper, we used a resolution of 10km that may under-estimate convective transport for small-scale features but should be sufficient to resolve some of the large-scale convective events in summer. We added the following sentences to clarify in the text in ***bold/italic***:
P17/L31-L34: ***The impact of CP on CO_2 mole fractions requires more evaluation, because our convective scheme is not coupled with the tracers (i.e., CO_2 mole fractions), however, we can still use the convective schemes to evaluate its impact on wind fields and PBLH.*** Consequently, we cannot yet quantify the impact of the lack of parameterized cumulus transport of CO_2 transport on our findings.

REF-C3: The spatial resolution of the inner domain (10km) isn't very high. Given the importance of the representation of heterogeneity in simulation of both anthropogenic and biospheric CO_2 fluxes, it'd greatly help to conduct higher resolution simulations (2-4 km).

Author-C3: Certainly, having a higher horizontal resolution model (i.e., 2-4 km) will improve at least the atmospheric transport over the US Upper Midwest region and provide more variability or heterogeneity to simulated CO_2 mole fractions linked with weather systems. However, in terms of CO_2 mixing ratios during non-frontal conditions, the spatial scale of structures over the US Upper Midwest is on the order of tens of kilometers. Therefore, increasing the spatial resolution is not expected to improve the representativity of the simulation results. High-resolution simulations would, however, increase considerably the required computational time to execute multiple simulations over a large domain (i.e. 1,600x1,600km). We also note that WRF-Chem is coupled to the only operational inversion product available for CO_2 fluxes at 1°×1°, CarbonTracker. If higher-resolution fluxes become available, the WRF grid spacing could be increased to explore fine-scale structures.

We add a note about the decision of the 10-km resolution in the text (***bold/italic***):

P4/L24-26: The coarse domain (d01) uses a horizontal grid spacing of 30-km and the nested or inner domain (d02) uses 10-km grid spacing (Figure 1). ***Because of limited computational time and the resolution of the CO_2 surface fluxes described on section 2.6, we decided to keep our highest resolution of the model up to 10-km.***

REF-C4: The vertical resolution of the domain is quite high (40 levels <2km) near surface. However, there are only 19 levels above 2km. How the vertical levels aloft are distributed? The vertical grid spacing will impact how the model captures capping inversion layer, clouds and so on.

Author-C4: This was a mistake in the document. We have a total of 59 vertical levels where 40 levels are within the first 4km of the atmosphere instead of the first 2km (see Figure1). Figure 1, shows a vertical profile of temperature at OAX site for configuration of our ensemble. The first black line shows the first 40 levels, above level 40 the figure includes dashed lines every 5 levels up to level 59. Because the statement in the article was incorrect, we end up fixing line 22-23 on page 4 as follow:

P4/L21-L22: "*The atmospheric column in each simulation is described with 59 vertical levels, with 40 of them within the first 4-km of the atmosphere. Two nested domains were used.*"

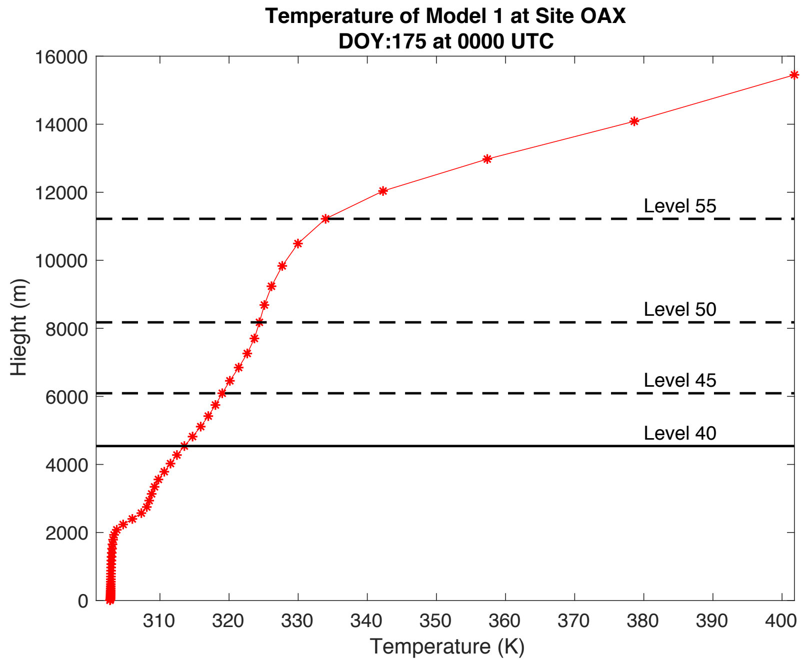

Figure 1. Vertical profile of potential temperature at OAX rawinsonde site from Model 1 on DOY 175 at 0000 UTC. Black solid line represents the height at level 40 of the model and the dashed lines shows the height of every 5 level.

REF-C5: 14 radiosounding sites are used to evaluate the meteorological simulations. The accurate simulation of vertical mixing is important for CO_2 simulations. Unfortunately, the regular radiosoundings in the US don't capture (00 and 12Z only) the deep boundary layers during the afternoon hours in the Midwest due to timing. I recommend using additional data (e.g. ceilometers) to evaluate the daytime evolution of the boundary layers in Midwest.

Author-C5: We agree with the reviewer that radionsonde data are collected at the same time of day, usually after the PBL height peak, and cover only certain parts of the domain. We considered adding more observation to our evaluation but decided not to for several reasons. The radiosonde data provides the best spatial and temporal resolution that we can have for our project. Other data such as CO_2 mole fractions from the NOAA aircraft program (or other aircraft campaigns) will bring the limitation of the time and/or spatial coverage. Only five sites are within our WRF simulation domain and profiles are usually collected every two weeks (i.e. about 2 per site over our simulation period). Most aircraft campaigns do not sample the mixing depth but rather collect long transects over the continent (e.g. COBRA, ATom). In addition, we have no available intensive campaign in that part of the country and for the simulation period we selected. In the near future, the Atmospheric Carbon & Transport (ACT) project funded by NASA, that is currently performing multiple fields campaigns over the East and Midwest of United States for different seasons, will provide a significant spatial coverage to address transport model errors. The campaign has just finished its fourth deployment in the Spring of 2018. Regarding surface stations, including them in our analysis would bias our results towards near-surface model errors, which we know are not representative of the whole Planetary Boundary Layer (PBL) (e.g. Hu et al., 2010; Rogers et al., 2013; Deng et al., 2017). CO_2 molecules are mixed over the entire PBL during daytime. Therefore, adding tens to hundreds of surface stations will not represent the actual model errors in the PBL. For these different reasons, we decided to focus on operational radiosondes launched at 00z, using mid-PBL wind measurements. Because we are aware of this limitation in our study and this is a concerned from both reviewers we decided to add text to the last paragraph of our Discussion section (P18/L1-L3).

Citation: Rogers, RE, Deng, A, Stauffer, DR, Gaudet, BJ, Jia, Y, Soong, S, Tanrikulu, S, 2013, Application of the Weather Research and Forecasting Model for Air Quality Modeling in the San Francisco Bay Area. *J. Appl. Meteor*., 52: 1953–1973. DOI: https://doi. org/10.1175/JAMC-D-12-0280.1

REF-C6: The choice of the meteorological data for the model verifications is limited to the radiosounding data, which are quite limited in time and space. I suggest adding surface wind, temperature and other measurements to the model evaluations.
Author-C6: We are aware of the limited data used to explore the transport errors. However, our main interest as mentioned on section 2.8 of the paper is to evaluate meteorological observations at height similar to the observation that means that we want to use observations that are within the well-mixed layer. Surface station will fall out of the height of our interest and can represent more the surface layer than the convective layer. With respect to the variables explored, we also clarified on Section 2.9, that our main interest was to explore all the variables that we assumed contribute the most to the representations of the CO_2 mole fraction distributions within the PBL. Additional to that, current atmospheric inversions do not use surface wind speed and direction as input, however, it uses these two variables within the PBL.

It will be ideal a study over a region and a period of time where multiple observations can be used to have broader view of the transport uncertainty. This is one of the main goals of the Atmospheric Carbon & Transport (ACT) project by NASA, that is currently performing multiple fields campaigns over the East and Midwest of United States for different seasons. ACT main objective is to understand how different weather systems transport different greenhouse gases (GHGs), evaluate this transport and improve the estimates of sources and sinks of these GHGs.
Because we are aware of this limitation in our study and this is a concerned from both reviewers we decided to add this to the last paragraph of our Discussion section, where we introduce the different limitation of this study (see ***bold/italic*** lines in the next paragraph).

P17/L1-L3: "***We also note that models were compared only to rawinsonde data, the only type of observation that had both the temporal and vertical resolution needed to evaluate the models within the PBL. More observations with higher temporal, spatial and vertical resolution will be an asset for future evaluation of transport models, focusing on intensive campaigns over multiple seasons.*** Our meteorological results, however, are broadly consistent with past literature."

REF-C7: It's recommended to use the GF cumulus parameterization instead of the G3 scheme in WRF. I suggest testing the model with the GF scheme.
Author-C7: We have selected the Grell-3D and Kain-Fritsch schemes to represent two families of convective parameterization in order to capture model sensitivities. Recent papers suggested the Grell-Freitas parameterization to produce more reliable results in WRF simulations. We selected the older version which has been used extensively in the literature when we designed this study. While modelers often aim at simulating the atmospheric dynamics the most accurately, we focus here on the differences when using various combinations of schemes, and found that surface

schemes and PBL schemes are the most critical parameterizations. The convective parameterizations have also some impact on our results to a lower extent, but no combination of surface and PBL was able to capture the variability in mixing ratios without systematic errors. We concluded that no single configuration can guarantee a reliable representation of near-surface CO2 transport, and recommended to use ensemble-based approaches. Despite the use of the GF scheme, the biases from PBL and surface schemes will remain. We added the following sentence to encourage future ensemble approaches to include the GF scheme:

P17L28-L31: ***Also, as noticed in recent studies, the Grell-Freitas convection scheme produced more reliable simulations of the atmospheric dynamics (Gao et al., 2017; Gbode et al., 2018). Therefore, we recommend the use of newly developed schemes for future studies as model schemes are made available in new model versions.***

REF-C8: There is one key uncertainty associated with using the CO_2 fluxes from CT in WRF-CO_2 modeling here. The spatial resolution of the EDGAR and CASA CO_2 emissions are much coarser than the inner WRF grid spacing here. Consequently, the WRF- CO_2 model can't capture the regional CO_2 variability even with "perfect" meteorology. This needs to be discussed in this paper. At least, high resolution anthropogenic CO_2 emissions (e.g. VULCAN) could be used in such model setup.
Author-C8: In the Upper Midwest, the spatial gradients in CO_2 fluxes are on the order of tens of kilometers, mostly driven by land cover types. Corn and soybean fields extend over wide areas (entire Iowa, Illinois, and beyond), as well as forested land in Missouri, and wheat fields in the western part of the region. For that reason, major spatial gradients are correctly represented at 1x1 degree resolution. Currently, a high-resolution CASA simulation at 500-m resolution has been developed and will be coupled to WRF-Chem. In future studies we plan to investigate differences in CO_2 fluxes. The main concern comes from the location and magnitude of the sink, as CarbonTracker is a global inversion model not aimed at representing regional fluxes. The summer drawdown and seasonal cycle show differences that could impact the CO_2 mixing ratios. However, no significant differences are expected in terms of spatial gradient representativity.

We added the following sentence on page 14 (***bold/italic***):
P14/L30-L31: This result suggests that transport model errors from our ensemble only represent a fraction of the total uncertainty in our modelling system. ***In this study we use CarbonTrakcer fluxes which is a global inversion system and does not aim to represent regional fluxes.*** Therefore, Additional errors can be due to incorrect CO_2 surface fluxes and boundary conditions.

REF-C9: Minor comments: Abstract: ". . .is this variability is. . ."?
Author-C9: Thanks for the comment, this part of the sentence was fixed as follow:

"PBL height varies across ensemble members by 300 to 400 m, and this variability is controlled by the same physics parameterizations."

REF-C10: For the WRF model, please cite the recent WRF/WRF-Chem description paper: Powers et al., Weather Research and Forecasting Model: Overview, System Efforts, and Future Directions, AMS. https://doi.org/10.1175/BAMS-D-15-00308.1

Author-C10: Thanks for the reference, however, the article that you are suggesting is more an overview of WRF-ARW and WRF-Chem. For this type of paper, we would prefer to cite a more technical paper such as Skamarock et al. (2005).