The reviewers' insightful comments are highly appreciated. Below we have listed the referees' comments in black and our response in blue.

We have made the following major revisions in the revised manuscript:

1. More descriptions of aerosol properties simulated in the model are added in the revised manuscript.
2. Two aerosol precursors (NO_2 and SO_2) observed by EPA are included to diagnose model biases in NO_3 and SO_4, respectively.
3. Analyses of meteorological variables, including temperature, relative humidity, wind speed and precipitation, are included.
4. Analysis of Ångström exponent is included to diagnose the model simulated aerosol particle size.
5. More quantitative information, including correlation and bias, is included in the discussion.
6. We have performed some sensitivity experiments to provide more in-depth analyses on model results, including changing the anthropogenic emission source (20km_NEI11), the chemical boundary conditions (20km_BC1) and the PBL scheme (20km_P7).
7. A bug in calculating equivalent potential temperature is fixed in the revised manuscript. The unit of relative humidity was wrong in previous version. The updated profiles of equivalent potential temperature do not change the conclusions of this study.
8. The OC (organic carbon) from observations are converted to OM (organic matter), which is simulated in the model, by multiplying by 1.4 to account for hydrogen, oxygen, etc.

Anonymous Referee #1

This paper examines the performance of a regional-scale chemical transport model in representing aerosol properties in the San Joaquin Valley over a one year period. The model is compared with surface measurements of composition and AOD as well as satellite measurements. The motivation for the paper is sufficient (although could be improved), but the main weakness is their approach and interpretation of the simulations. In addition, the paper is poorly written.

Major Comments:

The most important problem the manuscript has is how the model was configured to address the purpose of the study. WRF-Chem is a useful tool, but as with all models can only perform well when it is configured properly.

The following is a discussion of items the authors should consider to revise and/or address.

Domain and Dust Emissions: It is clear that the model domain is larger than the one shown in Figure 1. But it is hard for me to assess the importance of dust emissions since those are not shown. For local sources, dust is likely generated in the desert areas to the southwest of the SJV. It would be useful to show the emission regions from GOCART and DUSTRAN. My understanding is that the emission regions in DUSTRAN as implemented in WRF-Chem are rather ad hoc. They may depend on vegetation type. I suspect that dust is being generated locally in the SJV in DUSTRAN but not in GOCART.

Thanks for the suggestion. Dust emissions are included in Figure 7 in the revised manuscript (also in the following Figure 1). As the reviewer hinted, dust is being generated locally in the SJV in DUSTRAN but not in GOCART. Discussions about the differences between DUSTRAN and GOCART are included in the last two paragraphs of section 3 in the revised manuscript.

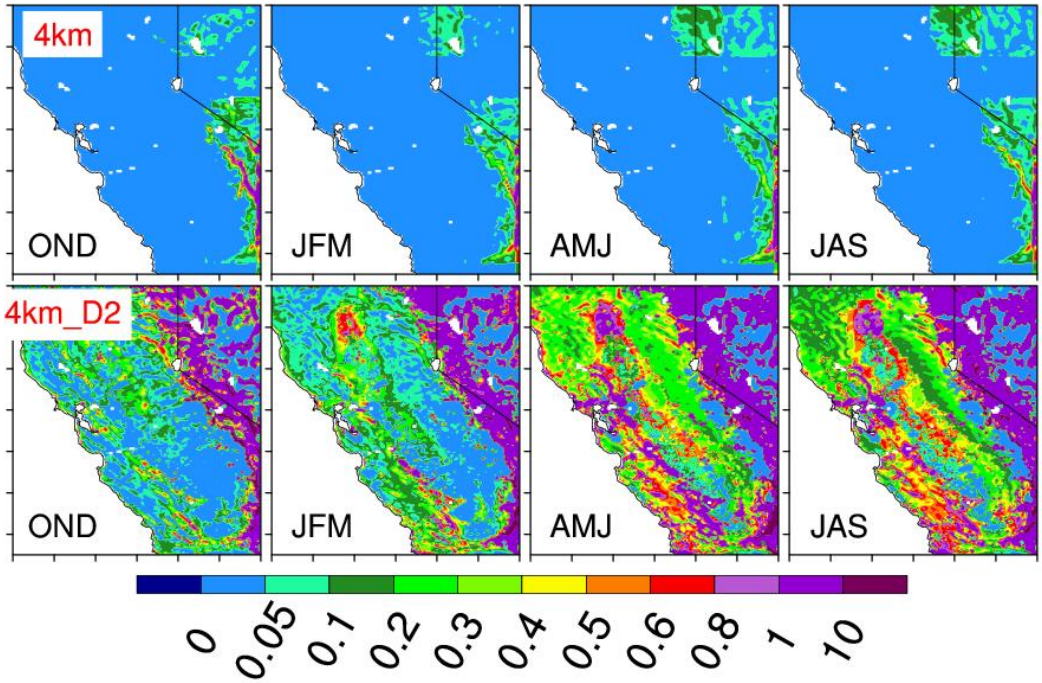

Figure 1. Seasonal mean of dust emission rate (μg m^{-2} s^{-1}) for (upper panel) GOCART; (lower panel) DUSTRAN.

The authors mention how many grid nodes are used in the vertical direction, but should give an idea of the vertical resolution near the surface that will affect dust emissions.

The vertical resolution from surface to 1 km gradually increases from 28 m to 250 m. It is clarified in Line 204 of the revised manuscript.

Dust emissions will depend in part on wind speed, and representing wind speed in California depends a lot on circulations affected by terrain. Both a fine horizontal and vertical resolution is needed to represent those winds that will affect dust emissions. It is not clear how well the model performed in winds – particularly over the dust emission regions. While some evaluation of the thermodynamic structure is given, there is nothing for the winds.
The evaluation of wind speed comparing to surface observations from CIMIS (California Irrigation Management Information System) is included in Figure 2b of the revised manuscript. The model simulations underestimate wind speed in the cold season. In the warm season, the 20km run underestimates wind speed except June while the 4km run overestimates wind speed, which indicates wind speed is not the main reason for AOD biases in the warm season. Discussions of wind speed impacts are included in the first paragraph of section 4.3 in the revised manuscript.

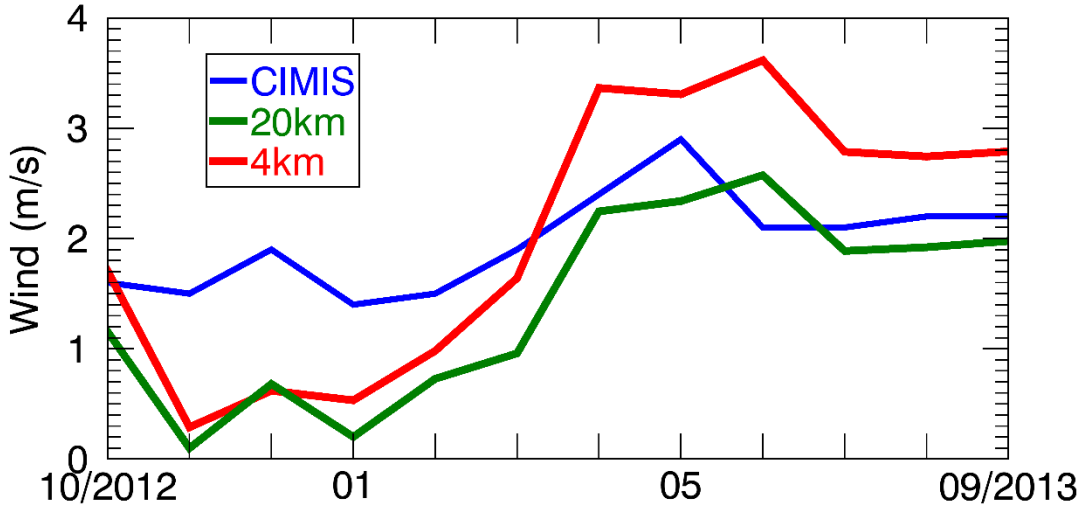

Figure 2. Simulated monthly 10-m wind speed (m/s) at Fresno, CA compared to CIMIS (California Irrigation Management Information System) observations.

Boundary Conditions: The authors half the amount of aerosols from MOZART following Fast et al. (2014). But the errors in a coarse global model, like MOZART, will likely change in time and depend on meteorological conditions. There is no sensitivity results or evidence whether such a change in boundary conditions is warranted in the present study. I believe the version of MOZART the authors use prescribes dust using climatology which would affect the simulations over California. The potential errors in MOZART that will contribute to AOD over California will likely vary over a year-long period.

We have run two sensitivity experiments with DUSTRAN at 20 km resolution, one with MOZART divided by 2 (20km_D2) and the other with original MOZART (20km_BC1). AOD maps are shown in the Supplementary Fig. 1 and the following figure. It is clear that the 20km_BC1 run overestimates AOD in the rural regions from OND to AMJ. Both the 20km_D2 (BC0.5) and 20km_BC1 (BC1) runs underestimate AOD in the rural regions in JAS, which indicates chemical boundary condition is not the main reason for the underestimation of JAS AOD in the simulations. Thus, we keep the setting of halving the amount of aerosols from MOZART in the simulations.

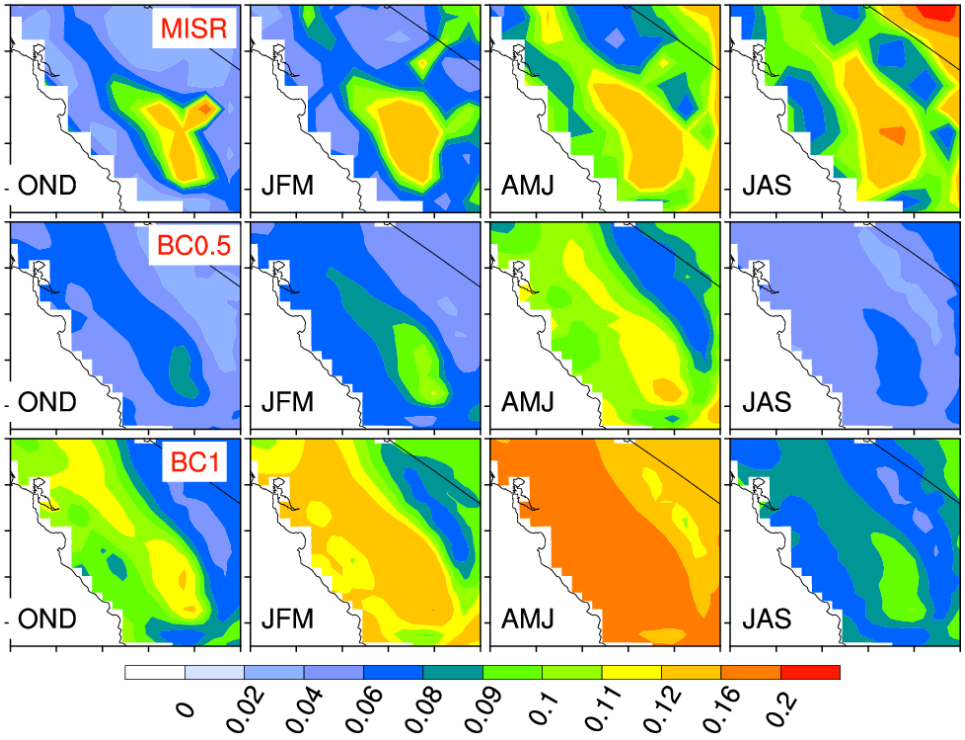

Figure 3. Spatial distribution of seasonal mean 550 nm AOD from MISR, the 20km_D2 (BC0.5) and 20km_BC1 (BC1) in WY2013.

Simulation Period: On line 167, the authors state that the simulation period is from 2012 to 2013. There is no rationale as to why this period is chosen. Perhaps it does not matter and they are only looking at seasonal variations. But this are these seasons "typical" or not?

We are only looking at seasonal variations. Similar results are also shown in our initial experiment in WY2012. For further investigation of model performance by comparing with the DISCOVER-AQ field campaign datasets in 2013 (a future study), we switched all our experiments to WY2013.

Anthropogenic Emissions: The authors use the 2005 NEI, but it would have been more appropriate to use this 2011 inventory which is closer to the time of

the simulation period. Even more ideal, would be to use emissions generated by CARB that are likely to have local emissions in California better represented. There are papers describing this inventory that at least be cited and the changes in SO2 and NH3 emissions in the SVJ valley (which are likely to be very different that the NEI 2005) will contribute to the nitrate and sulfate errors described in the paper. Since dust is an important factor over a large portion of the year, the differences in anthropogenic emissions are not likely to affect that conclusion. But it would affect the relative contribution of anthropogenic to natural sources over the year.

The 2011 NEI was not available in the WRF-Chem emission datasets when we initiated this study. We have run two sensitivity experiments with the 2011 NEI (20km_NEI11) and 2005 NEI (20km_D2) at 20 km resolution with the DUSTRAN dust scheme. Results are shown in the supplementary materials and the following figures. The differences between NEI11 and NEI05 are small comparing to the identified model biases in this study. As the reviewer pointed out, the differences in SO_4 and NH_4 are relatively large. However, SO_4 in NEI11 has larger biases than SO_4 in NEI05.

As shown in Fast et al. (2014), "reducing the default CARB emissions by 50% led to an overall improvement in many simulated trace gases and black carbon aerosol at most sites and along most aircraft flight paths; however, simulated organic aerosol was closer to observed when there were no adjustments to the primary organic aerosol emissions". We can see all the emission datasets (CARB, NEI11 and NEI05) have uncertainties in the aerosol emissions. We decide to keep our current model setup and include discussions of the uncertainty in the emission data sources in the revised manuscript.

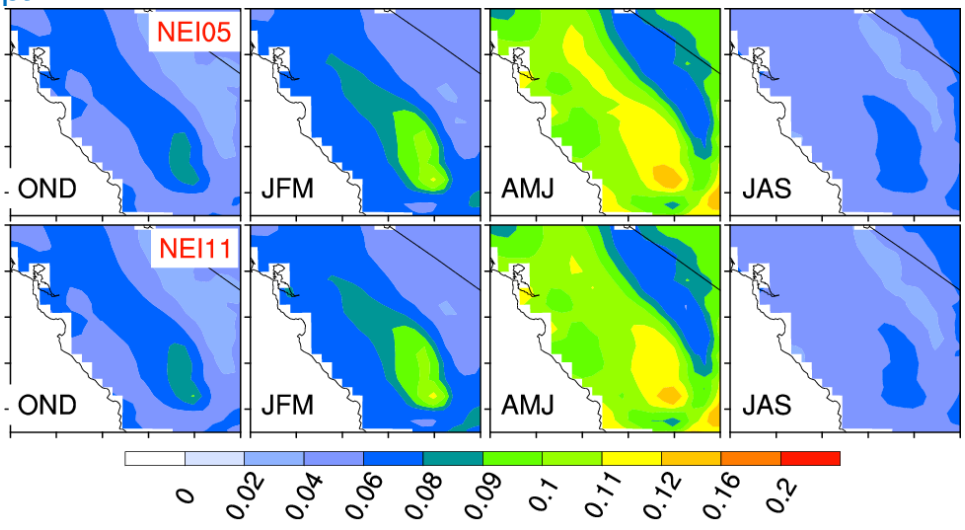

Figure 4. Spatial distribution of seasonal mean 550 nm AOD from the 20km_NEI11 (NEI11) and 20km_D2 (NEI05) runs in WY2013.

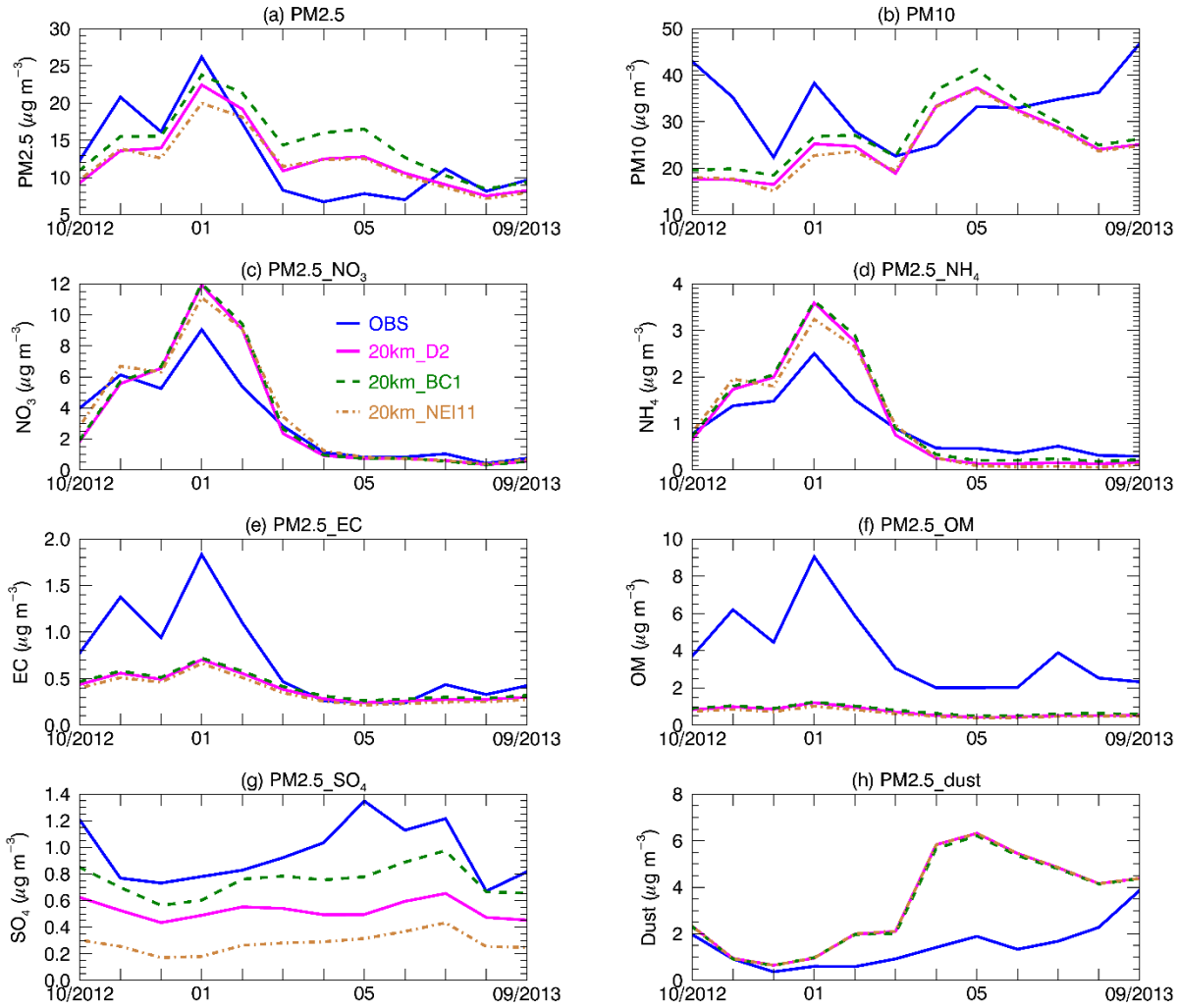

Figure 5. Aerosol mass (μg m^{-3}) for different species from EPA-CSN (OBS), the NEI05 (20km_D2) and NEI11 (20km_NEI11) runs at Fresno, CA. PM2.5_NO_3 represents NO_3 with diameter ≤ 2.5 μm. Similar definition for SO_4, EC, OM, dust and NH_4 in the figures.

Model Evaluation: The authors used satellite equivalent potential temperature to evaluate the temperature profiles in the model. As seen in Figure 9, it seems that the vertical resolution is coarse so it is not the best source to examine near-surface temperature gradients. Two of the near-surface AIRS profiles look unrealistic to me. In addition it appears to have a 1 deg uncertainty (which is large for temperature) and is from a 1 degree grid – which will average out substantial temperature variations in areas affected by terrain. Using radiosondes would be a much better way to evaluate the model. The coarse vertical resolution of AIRS also leads to misinterpretations about boundary layer mixing. They claim that boundary layer mixing is too weak and explains why the simulated extinction profiles are wrong in AMJ and JAS.

There is simply not enough aerosols around, no matter what the vertical distribution.
Unfortunately, there is no routine radiosonde observation available in the SJV. AIRS data have been extensively evaluated using radiosondes in other regions. We agree that the coarse vertical resolution of AIRS data cannot fully resolve near-surface temperature gradients. However, AIRS is the best dataset currently available to evaluate seasonal variations of the vertical temperature/moisture profiles in the model simulations over the SJV. Evaluation of surface temperature/RH is conducted by comparing with surface observations in the revised manuscript. Results are consistent with evaluations of vertical profiles comparing to AIRS. More analyses of aerosol biases in the boundary layer are included in the revised manuscript.

We have found that the unit of RH is wrong in our code to calculate equivalent potential temperature. It is fixed in the revised manuscript. The profiles look reasonable now. It doesn't change the conclusions of this study.

Missing Aspects: While the authors have evaluated simulated aerosol composition and PM25/PM10 mass, they have not examined aerosol water. During dry conditions of the summer months, this may not be a large factor contributing to extinction. Aerosol water is likely to become more important aloft, where RH is likely to be higher. But one does not know unless it is examined. Is there significant aerosol water in the simulations?
Aerosol water will be influenced by simulated RH, so an evaluation of simulated RH is in order.
Evaluation of simulated RH is included in the supplementary and discussed in the revised manuscript. As shown in following figures, there are dry biases in the model simulations. However, due to the relative dry environment (RH<50%) in the warm season, the dry bias may not be responsible for the underestimation of aerosol extinction in the boundary layer and column-integrated AOD through hygroscopic effects (Feingold and Morley, 2003).

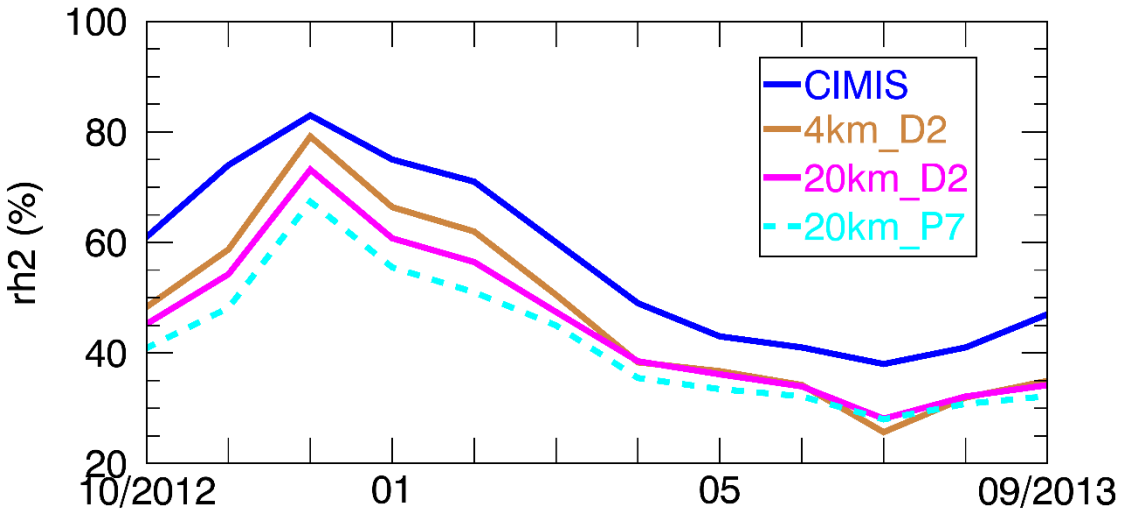

Figure 6. Monthly mean 2-m RH (%).

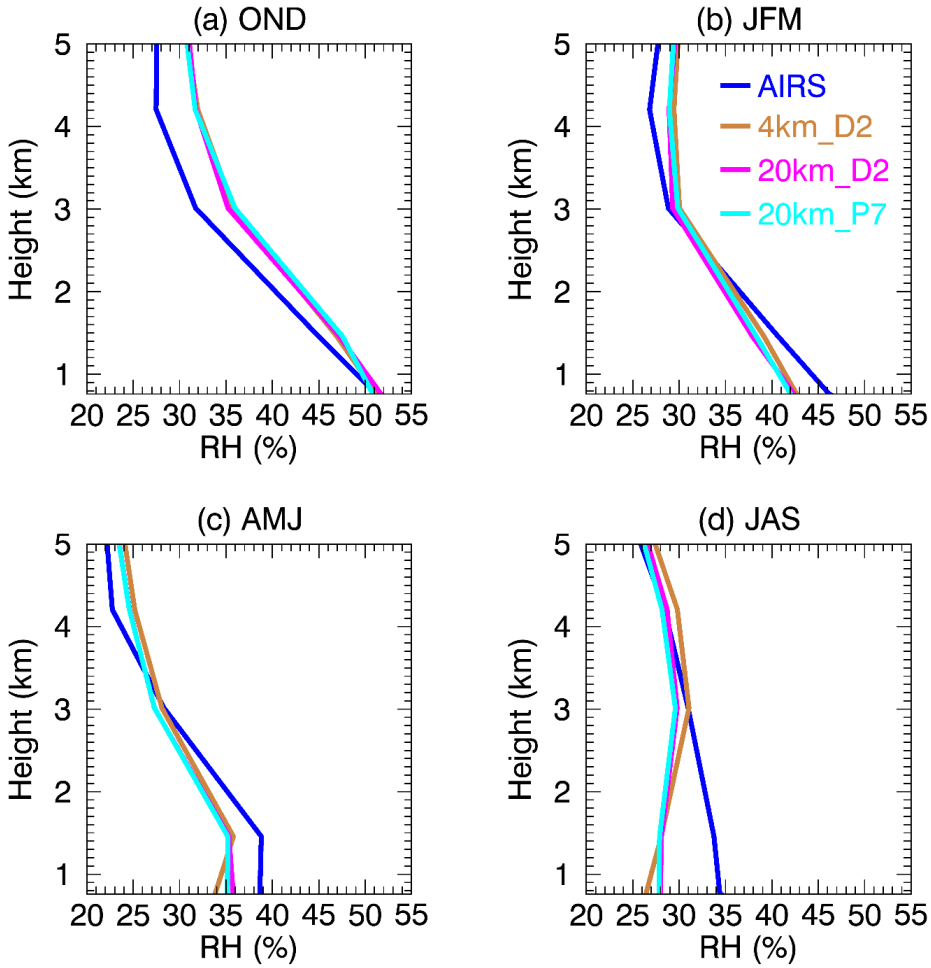

Figure 7. Vertical profile of seasonal mean relative humidity (%) in the WRF-Chem simulations comparing to AIRS. The 20km (not shown) run is similar to the 20km_D2 run while the 4km run (not shown) is similar to the 4km_D2 run.

A second missing aspect is SOA. I assume the version of MOSAIC they use does not include SOA. Yet SOA has been shown to be a major factor in PM25 for much of the year in California. While SOA concentrations will be lower than dust concentrations (when significant dust is present), it seems that omitting SOA is problem. One motivation factor in the study was related to using and air quality model (such as WRF-Chem) to guide emission control strategies. That would include OC emissions. But it seems that only primary OC is included, so that comparing simulated OC to observed OC is misleading.

SOA processes are not included in our simulation. Fast et al. (2014) used the simplified two-product volatility basis set parameterization to simulate equilibrium SOA partitioning in the WRF-Chem model. SOA is still underestimated in their simulation in May and June. We tried to run the WRF-Chem model at 20 km resolution (20km_VBS2) following the settings in Fast et al. (2014). However, our simulation can only produce comparable AOD in AMJ while AOD in other seasons are underestimated. Since it is challenging to correctly represent SOA processes in regional climate models, we keep our current settings and discuss the impact of SOA processes in the revised manuscript.

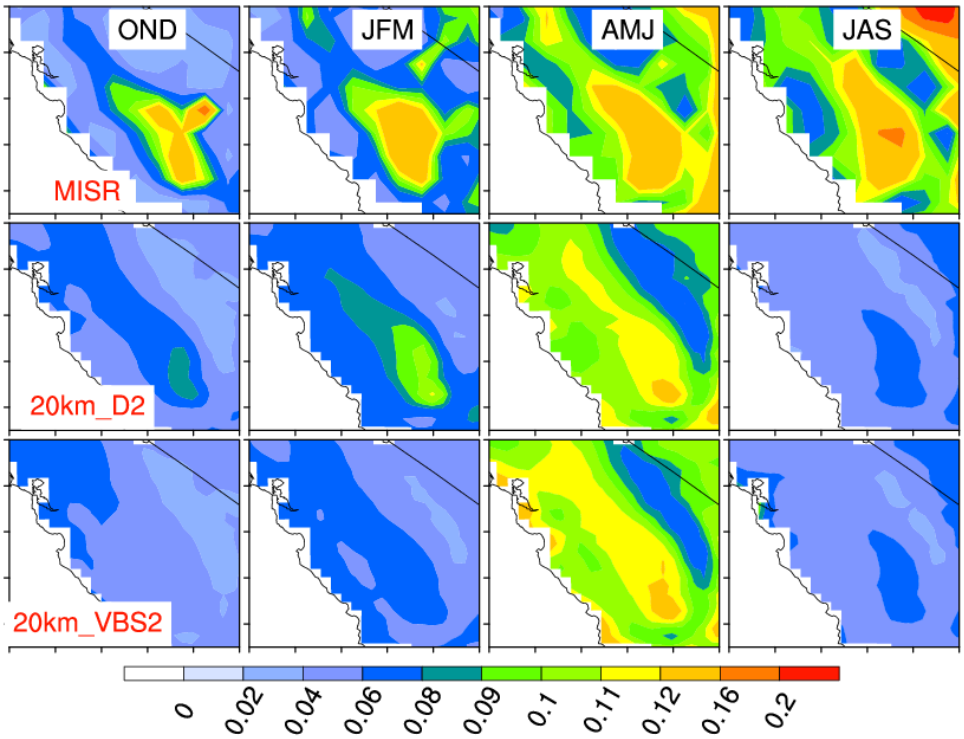

Figure 8. Spatial distribution of seasonal mean 550 nm AOD from MISR, 20km_D2 and 20km_VBS2 in WY2013.

Also, MOSAIC simulates organic matter (both carbon and oxygen), so do the authors account for the missing oxygen parts in the measurements that are labeled OC?
Thanks for your comment. The observed OC is converted to organic matter (multiply by 1.4) to compare with the simulated organic matter in the revised manuscript.

The authors also use a 4-bin version of the model which coarsely represents the aerosol size distribution. The authors should at a minimum discuss how this assumption affects their results and conclusions.
Discussion of the impacts of this assumption is provided in the revised manuscript as following:

"Zhao et al. (2013a) compared the impacts of aerosol size partition on dust simulations. It showed that the 4-bin approach reasonably produces dust mass loading and AOD comparing to the 8-bin approach. The size distribution of the 4-bin approach follows that of the 8-bin approach with coarser resolution, resulting in ±5% difference on the ratio of PM2.5-dust/PM10-dust in dusty regions. Dust number loading and absorptivity are biased high in the 4-bin approach comparing to the 8-bin approach."

It would have been useful to see some sort of evaluation of aerosol size distribution, since that also affects extinction and AOD. So the authors are really not probing all the aspects that affect uncertainties in simulated extinction and AOD.
Evaluation of Ångström exponent (AE), an indicator of aerosol particle size, is included in Fig. 4b of the revised manuscript. WRF-Chem captures the seasonal variability of the AE well, with a correlation of 0.90 in both the 20km and 4km simulations. The magnitude of AE is also approximately simulated in the cold season, with a mean of 1.15 (1.20) in the 20km (4km) runs compared to 1.33 in the observation. However, the simulated AE is underestimated by ~30% in the warm season, indicating that the simulated particle size is biased high during this period.

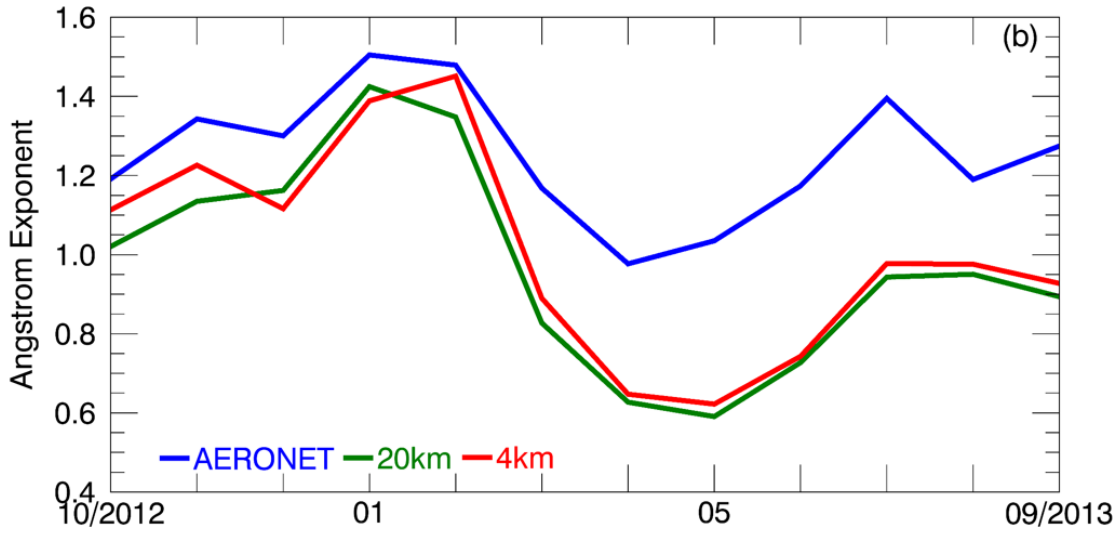

Figure 9. Monthly mean Ångström Exponent between 600 nm and 400 nm at Fresno, CA.

Model Interpretation: All of the above factors will affect the interpretation of the model results and whether local (due to WRF-Chem) or long-range transport (not WRF-Chem related) sources of dust contribute to the errors in simulated dust concentrations and the vertical distributions. As stated in the summary, the authors claim the errors are largely due to errors in the dust emissions (not clear whether they mean local emissions or those from long-range transport) and vertical mixing. Given how the model has been used, they have not provided sufficient evidence to convince me that is the case.

The simulated aerosol extinction in the free troposphere above the boundary layer is close to or larger than CALIOP, suggesting that aerosols transported from remote areas through chemical boundary conditions (e.g., the differences between the 20km_BC1 and 20km_D2 runs in Supplementary Fig. 3) may not be the major factor contributing to the underestimation of dust in the boundary layer in the SJV. It is clarified in the revised manuscript.

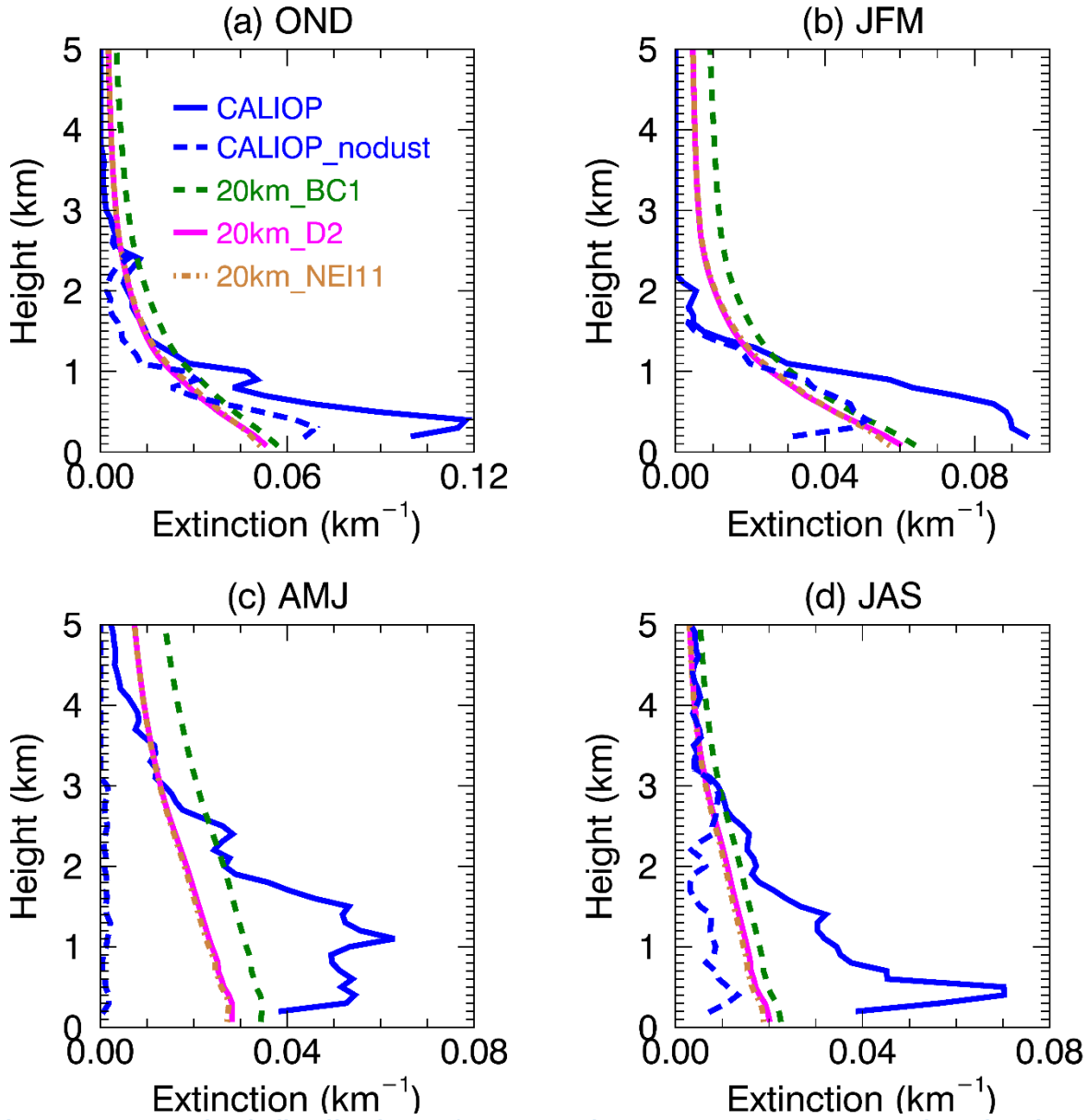

Figure 10. Vertical distribution of seasonal mean 532 nm aerosol extinction coefficient (km^{-1}) from CALIOP, CALIOP_nodust, and the WRF-Chem (20km_D2, 20km_BC1 and 20km_NEI11) simulations over the red box region in Fig. 1a in WY2013.

Specific Comments:

Lines30-31: Change “in cold season” to “in the cold season” and similarly “in warm season” to “in the warm season”. This is the first instance of poor use of English in the text. I will not comment on other problems since I seem my role as commenting on the science, rather than correcting the grammar. The

authors should use an editor if the co-authors are not willing to help out with the English.
Careful proofreading is provided by the co-authors (James Campbell and Hui Su) for the revised manuscript.

Lines 43-45: This statement is an obvious one and I am not sure it is needed. The focus of the paper seems to be on dust, so this is a secondary issue.
Removed per your suggestion.

Lines 92-104. This paragraph provides an important motivation for the study, but could be strengthened. Many readers will not know why models, such as WRF-Chem, are needed to develop/verify/modify satellite retrievals. It would be useful to add a few sentences describing how such models are used to demonstrate the purpose.
The following sentences are added in the revised manuscript to describe how the WRF-Chem model will be used in the MAIA retrieval algorithm.

"A significant challenge for aerosol remote sensing in retrieving spatial information on specific aerosol types, especially near the surface, is due to the lack of information on the vertical distribution of aerosols in the atmospheric column and limited instrument sensitivity to aerosol types over land. The WRF-Chem model will be used to provide near-real-time estimation of particle properties, aerosol layer heights, and aerosol optical depths (AOD) to constrain the instrument-based PM retrievals."

Line 214: "averaging process" is a phrase that is not clear or specific enough. It is not clear how the authors apportion the NEI 2005 emissions to the WRF domain, and the procedure should be some sort of "reapportionment" rather than interpolation. Simple interpolation cannot be used since that would not conserve mass. Did they check to make sure the total mass emitted from NEI 2005 with the WRF domain was actually the same as what was used after the emissions were reapportioned to the WRF domains?
Reworded to "reapportionment process". We use the standard emission conversion program in the WRF-Chem (convert_emiss.exe) to reapportion the anthropogenic emission. The domain-averaged emission rates for the 20km and 4km simulations are quite similar, as listed in the updated Fig. 1.

Line 257: The sensitivity experiment mentioned does not contain sufficient details for the reader to know why or how it was performed.

Reworded as: “The underestimation also exists in a sensitivity experiment (not shown) with the same model setups except initialized in April, indicating that the identified model biases in the warm season are not caused by potential model drift after a relatively long simulation period.”

Line 264: The authors start discussion Figure 5c before 5a. Why not change the order of the panels then to match the progression of the discussion in the text?
Order changed as suggested.

Line 338: There are far more studies evaluation WRF-Chem in simulating biomass burning than simply the one the first author led.
Two more references (Grell et al., 2011; Archer-Nicholls et al., 2015) are included in the revised manuscript.