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 #3

This paper shows the WRF-Chem simulation of aerosols in the SJV in California for one year and compares the results with observations of AOD from one AERONET site at Fresno and from MISR for a domain covering SJV, as well as measurements of aerosol mass concentrations of PM2.5, PM10, nitrate, sulfate, EC, OC, and dust from IMPROVE measurements. It tests the effects of using two different model resolution and two dust schemes, and attributes the model problems in matching observed AOD and PM10 to mainly the poor simulation of dust. It is stated in the "Introduction" that the paper a) "serves as the first step for future investigation of the aerosol impact on regional climate and water cycle in California" and b) provides a priori input for remote sensing retrievals for air quality for the MAIA mission.

While this paper has clearly shown the WRF-Chem performance over SJV that provides useful information, it lacks the vigor and thoroughness in the analysis and interpretation, and the information presented in the paper is

insufficient in helping understand the problems of the model. Given the goal of using such a model for MAIA retrieval and for climate study, much more in-depth analysis and vigorous diagnostics is necessary in order for the model improvements to be useful for those purposes. Although the content is suitable for ACP, major revisions are necessary before the paper can be considered again for publication.

General comments:

1. Dust simulations: The authors have concluded that the dust simulation is the major problem for model to capture the observed aerosol amount and variability in the warm months. Switching from GOCART to DUSTRAN just shows different problems but does not resolve the issue. However, there is no any explanation on the differences between the two schemes in terms of emission strength, source location, parameterization of dust mobilization, and deposition in order to understand why the dust amount and seasonal cycles are so different between the two schemes and yet none can capture the observations. Without understanding the cause of the problem, future improvement is not possible.

More descriptions and analyses of the two dust schemes are provided in the revised manuscript for better understanding the cause of the problem. For details, please see the last two paragraphs of section 3 in the revised manuscript.

2. Non-dust aerosols: Figure 4 clearly shows that the model does not have much skill to simulate sulfate and OC, but the problem has not been investigated. The ammonium is completely left out, which is an important part of total aerosol mass. Also, large fraction of aerosol is classified as "other", but it is not clear what the "other" aerosols are in both model and IMPROVE data.

Biases in simulated sulfate from precursor and marine intrusion are investigated in the revised manuscript.

The bias in OC is because SOA processes are not included in our simulation. It is still challenging to correctly represent SOA processes in regional climate models. We keep our current settings and discuss the impacts of SOA processes in the revised manuscript.

The ammonium is included in Fig. 4d of the revised manuscript. The performance of simulated ammonium is similar to nitrate.

"Other" refers to the difference of PM2.5 and the summation of specified PM2.5 (NO_3, NH_4, SO_4, OM, EC, dust). It is clarified in the revised

manuscript. In the model, it includes sea salt and other inorganic matter simulated in MOSAIC. In IMPROVE, it includes all other aerosols observed.

3. Optical properties: It is also not clear how AOD and aerosol extinction are computed from the simulated aerosol mass. Is aerosol microphysics package used for calculating particle sizes and mixing state? How is mass-based aerosol converted to extinction and AOD? Is the relative humidity considered in these calculations?
Description of how AOD and aerosol extinction are computed is added in the revised manuscript and attached as follows. More details can be found in Barnard et al. (2006, ACP).

"Aerosols are considered to be spherical and internally mixed in each bin (Barnard et al., 2006; Zhao et al., 2013b). The bulk refractive index for each particle is calculated by volume averaging in each bin. Mie calculation as described by Ghan et al. (2001) is used to derive aerosol optical properties (such as extinction, single-scattering albedo, and the asymmetry parameter for scattering) as a function of wavelength."

4. Chemistry: Nitrate, sulfate, and a significant fraction of OC are secondary aerosols that are produced by chemical reactions of their gaseous precursors in the atmosphere. The authors attribute the high bias of model-simulated nitrate to "high bias in nitrate emission", which is erroneous. The diagnostics should involve investigations of nitrate precursors such as NOx and HNO3, and also the formation of nitrate via heterogeneous reactions on dust and sea salt surfaces and homogeneous reactions in the sulfate-nitrate-ammonium system. It is not clear how WRF-Chem deals with nitrate formations and which is the major reaction pathway for nitrate aerosol production.
Same as sulfate – it is formed via gas and aqueous phase reactions of SO2. Better diagnostics of the problem is needed.
Thanks for the comments. Analyses of NO_2 and SO_2 are included in Fig. 6 of the revised manuscript. We also notice that switching the PBL scheme can produce better simulation of nitrate. More diagnostics of model biases are included in section 4 of the revised manuscript.

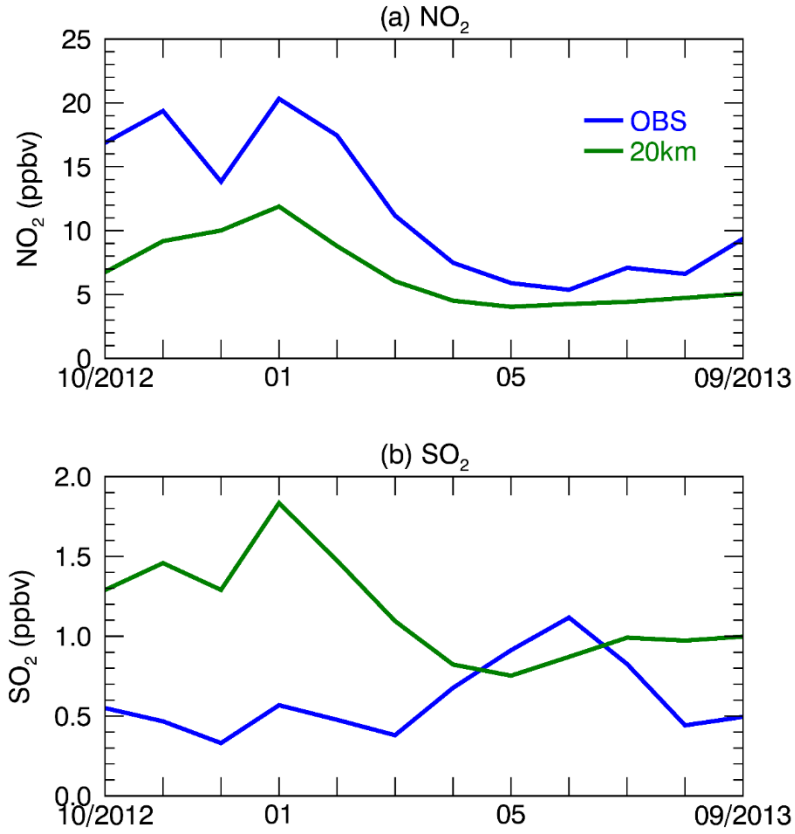

Figure 1. (a) NO_2 and (b) SO_2 from EPA (OBS) and the 20km run at Fresno, CA.

5. Other physical processes: Dry and wet depositions are the major removal processes for aerosols. The seasonal cycles of these processes also need to be investigated. For example, can the differences in seasonal variations between model and obs be partly explained by the differences in simulated and measured precipitation amount that determines the wet removal of aerosols? Or if the winds are realistically simulated in WRF-Chem that not only affect the dust emission, but also advection, both have profound effect on aerosol temporal and spatial distributions?

6. Meteorological fields: The only meteorological field compared in the paper is the equivalent potential temperature, which provides information on the atmospheric stability. Other important met fields, such as precipitation and wind speed/direction, as mentioned above, plays key roles in aerosol removal, transport, and wind-driven emissions of dust and sea salt but have not even mentioned in the paper. In addition, these fields and the physical processes driven by them are resolution-dependent, so the role of these met fields should be examined at different spatial resolutions.

The seasonal variability of precipitation is well captured in the simulations, while the magnitude of precipitation is smaller than the observations during the warm season (Supplementary Table 2). Wet removal processes are thus not likely the primary reason for the aerosol biases in the warm season.

The model simulations underestimate wind speed in the cold season (Figure 9 in the revised manuscript). 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 likely the main reason for AOD biases in the warm season.

Discussions of the impacts from precipitation, wind speed and other factors are included in section 4.3 of the revised manuscript.

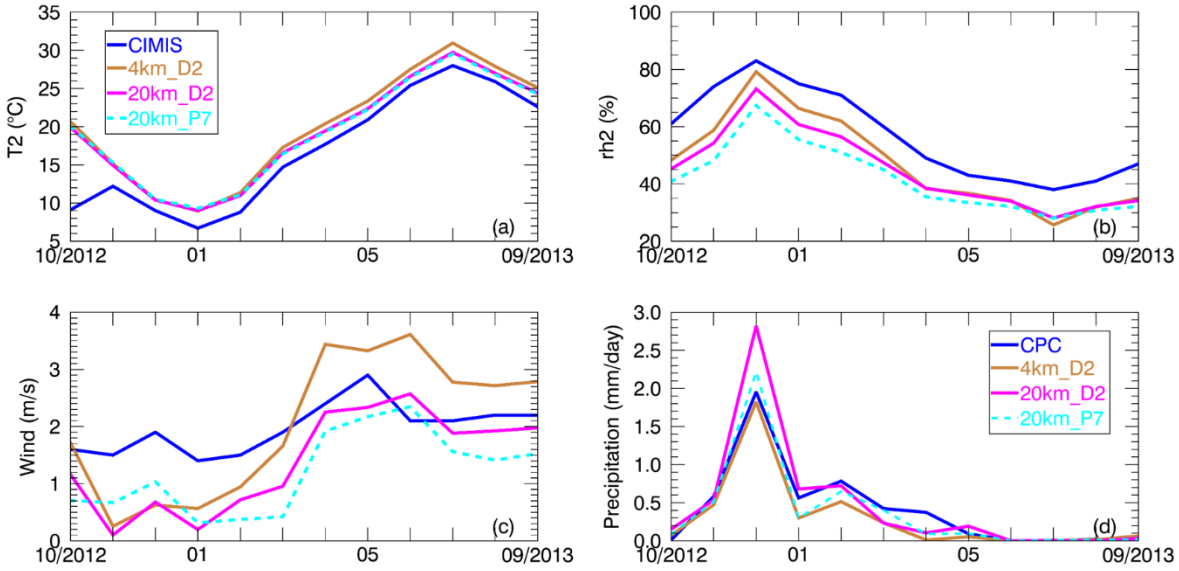

Figure 2. Monthly mean of (a) 2-m temperature (°C); (b) 2-m relative humidity (%); (c) 10-m wind speed (m/s); (d) precipitation (mm/day) at Fresno, CA. The 20km run (not shown) is similar to the 20km_D2 run while the 4km run (not shown) is similar to the 4km_D2 run.

7. Lateral boundary conditions: The effects of lateral boundary condition should be examined, or at lease discussed, particularly because of SJV's geophysical locations that is susceptible to the transpacific transport. How much of the aerosol species and their precursor gases are regionally/locally produced vs. imported from the lateral boundary, and how they affect the seasonal cycle? In other words, are the features/problems mainly produced by WRF-Chem? How important is the lateral boundary conditions to different aerosol species?

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. The impacts of the lateral boundary conditions to different PM2.5 species are small except SO4 (as shown in the following figure).

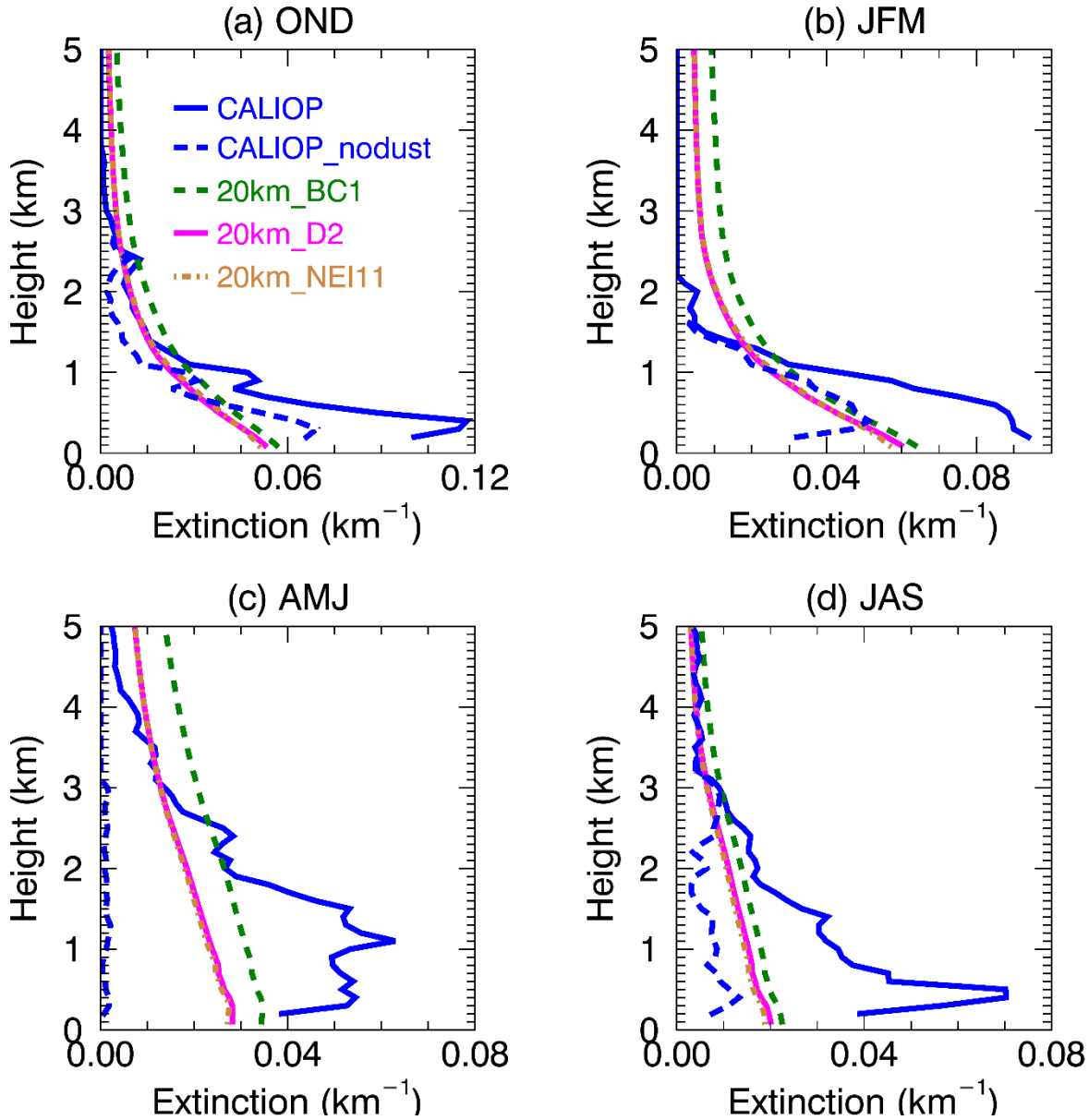

Figure 3. Vertical distribution of seasonal mean 532 nm aerosol extinction coefficient (km^{-1}) from CALIOP, CALIOP_nodust, and the WRF-Chem

(20km_D2, 20km_BC1 and 20km_NEI11) simulations over the red box region in Fig. 1a in WY2013.

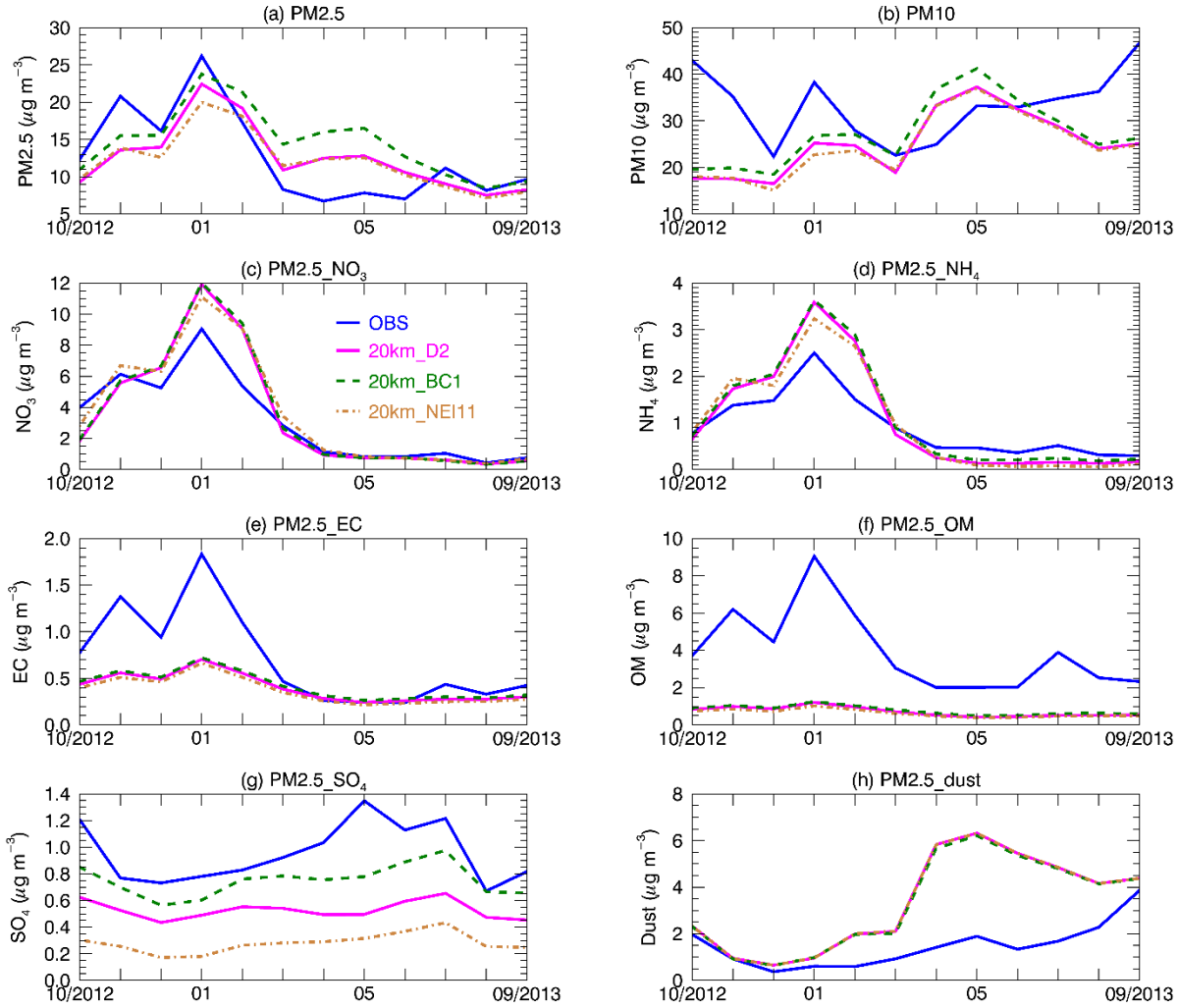

Figure 4. Aerosol mass (µg m-3) for different species from OBS, the 20km_D2, 20km_BC1 and 20km_NEI11 simulations at Fresno, CA. NH4 observations are from EPA; other observations are from IMPROVE. PM2.5_NO3 represents NO3 with diameter ≤ 2.5 µm. Similar definition for NH4, EC, OM, SO4 and dust in the figures.

8. Emissions: It seems the anthropogenic and biomass burning emissions used in this work are not up to date. For example, why the authors choose to use NEI05 emissions instead of more recent ones (e.g., NEI 2011 or NEI 2014) to better match the simulated time period (2012-2013)? Why GFEDv2 is preferred instead of GFEDv3 that was released a few years ago or GFEDv4 that has been available since 2015?

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. As shown in Fig. 4 and 5 here, the differences between NEI11 and NEI05 are small comparing to the identified model biases in this study.

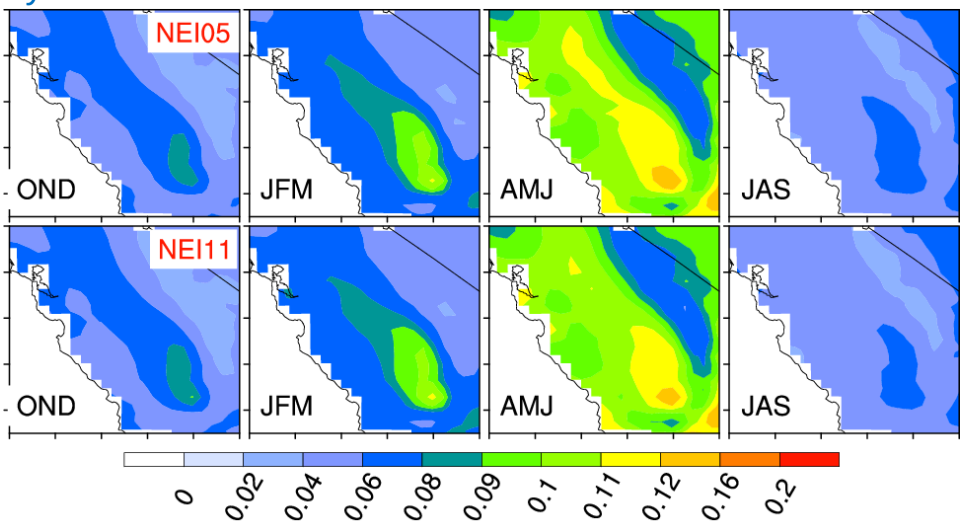

Figure 5. Spatial distribution of seasonal mean 550 nm AOD from 20km_NEI11 (NEI11) and 20km_D2 (NEI05) in WY2013.

We use the standard emission preparation program (prep_chem_sources_v1.5) for the WRF-Chem model to generate our fire emissions. Currently, only GFEDV2.1 is available in this program. Since fire emissions are not the major issues in our current simulations, we keep current settings.

9. Model-data comparison: 1) For AOD, there is only one AERONET site in the study region, and MISR's spatial coverage is limited. Why not use MODIS, which has a much better spatial coverage to have a better representation of "monthly average", in addition or even instead of using MISR?

We have compared the MISR data with the MODIS dark target and deep blue combined AOD V6 (as shown in the following figure). The MODIS data at 1°x1° cannot resolve the sharp gradient of aerosols in the SJV.

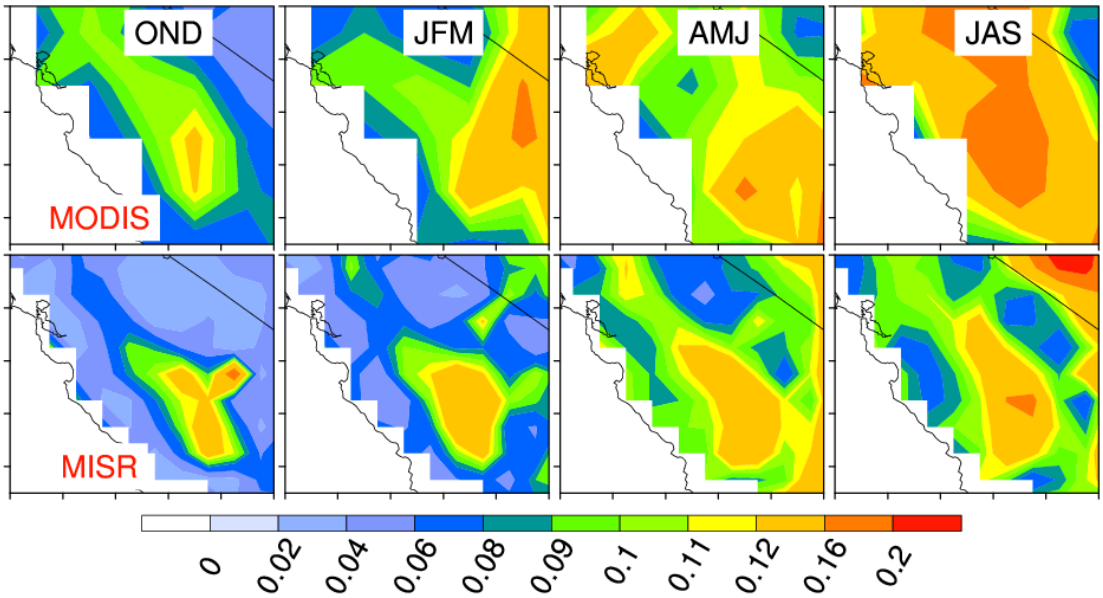

Figure 6. Seasonal mean AOD from MODIS and MISR.

2) Which months are defined as "cold" or "warm" months?
Cold months are from October to March; warm months are from April to September. The descriptions are in Line 277 and 282 in the revised manuscript.

3) More statistical quantities are needed to mark the agreement between model and observations, including correlation coefficients and seasonal/annual bias.
Correlation coefficients are included in the revised manuscript. More quantitative information are provided in the revised manuscript.

4) The authors should avoid using the subjective adjectives, such as "good agreement", "reasonably well", etc., to describe the comparisons between model and observations. More objective and quantitative methods and presentations are needed.
Following your suggestions, more objective and quantitative presentations are included in the revised manuscript.

5) Given that air quality changes quite a bit day to day and air quality forecasts are given on daily bases, why all the comparisons are done on monthly time scale instead of daily or sub-daily?
One of our goals is to evaluate model performances in simulating regional climate on the subseasonal-to-seasonable time scale. Many previous studies

have evaluated the performance of WRF-Chem in daily or sub-daily scale. It is not the focus of this study.

10. The most important step forward is to understand the causes of deficiencies in the model and suggest/incorporate improvements for better results. However, the current paper does not offer those aspects.
Following three reviewers' comments, more analyses about the causes of deficiencies in the model are included in section 4 of the revised manuscript. We summarize the model sensitivities in section 5 and indicate future directions for improvements.

Specific comments
Page 5, line 72-82: I wonder why Fast et al 2014 and Zhao et al 2013 were able to "reasonably" represent the observations with the same WRF-Chem model, either in the warm months (Fast) or on annual bases (Zhao), but this work has difficulties to do the same?
The WRF-Chem simulation is sensitive to various factors such as initial and boundary conditions, model parameterizations and emission sources. The performance of the WRF-Chem model are also different in different seasons and at different locations. Because we are focusing on different seasons and/or different locations, we can see different performances of the model simulations. Some sensitivity experiments are included in the revised manuscript to provide more in-depth analyses on model results.

Page 5, line 83: I don't think the word "extend" is appropriate – this study only focuses on SJV while Fast and Zhao showed large regions in CA.
Reworded as "we focus on simulating aerosol seasonal variability in the SJV, California using similar model configurations as that used in Fast et al. (2014) and Zhao et al. (2013b)."

Page 6, line 102-104: I don't get it – why simulation for SJV is critical to MAIA? Is MAIA only focuses on SJV?
SJV is a testbed for the MAIA retrieval algorithm development. It is clarified in the revised manuscript.

Page 7, line 116: Why are the original wavelength(s) from AEORNET that you used to interpolate to 550 nm?
AERONET AOD is interpolated to 0.55 μm from 0.50 μm and 0.675 μm. It is clarified in the revised manuscript.

Page 8, line 146: What does "speciated" mean here? There is no aerosol species information from the CALIOP data. Marine, polluted continental, etc. provided by CALIOP are aerosol types, not species.
Reworded as "Level 2 532 nm aerosol extinction data classify aerosols into 6 types" in the revised manuscript.

Page 9, line 179-180: How is convective transport (and removal) of aerosols simulated in 4-km resolution?
Convective transport (and removal) of aerosols are simulated at grid-scale in 4-km resolution. It is clarified in the revised manuscript.

Page 9-10, line 183-184: Was the overestimation by MOZART in the free troposphere a factor of 2 such that the concentrations had to be divided by 2? If the overestimation was only in the free troposphere, why the concentrations in the lower atmosphere and BL were also divided by 2?
The overestimation by MOZART is mainly in the free troposphere as shown in Fast et al. (2014) and our sensitivity experiment (20km_BC1). Lowering the boundary conditions of aerosols concentration by 50% greatly reduced the bias in simulated AOD for all regions of California. The impact of chemical boundary conditions at the surface is small in the SJV. For simplicity, all the boundary conditions by MOZART are divided by 2.

Page 10, line 198: Are the dust emissions in the GOCART and DUSTRAN also available in 20 and 4 km resolutions? What are the major differences between GOCART and DUSTRAN schemes?
Yes. More descriptions of GOCART and DUSTRAN schemes are included in last two paragraphs of section 3 in the revised manuscript.

Page 11, first paragraph in section 4.1: What PM2.5 species and precursor gases are emitted?
Nineteen gases (including SO2, NO, NH3 etc.) are emitted, while aerosol emissions include SO4, NO3, EC, organic aerosols, and total PM2.5 and PM10 masses. It is clarified in the revised manuscript.

Have you checked the domain budget between 4 and 20 km resolution to ensure the total emission for all species are identical with these different resolutions?
Yes, they are quite similar. Mean emission rates for the 4km and 20km runs are listed in Fig. 1 in the revised manuscript.

Page 11, line 215: How was AOD calculated without having information of PM2.5 composition? For example, dust and BC have very different mass to extinction conversion factor, known as mass extinction efficiency (MEE). There is no single MEE for a generic PM2.5 or PM10.
Aerosol composition is considered in AOD calculation. Different refractive index are assigned to different particles. Description of how AOD and aerosol extinction are included in the revised manuscript as the following.

"Aerosols are considered to be spherical and internally mixed in each bin (Barnard et al., 2006; Zhao et al., 2013b). The bulk refractive index for each particle is calculated by volume averaging in each bin. Mie calculation as described by Ghan et al. (2001) is used to derive aerosol optical properties (such as extinction, single-scattering albedo, and the asymmetry parameter for scattering) as a function of wavelength."

Page 12, line 237: As I said earlier, nitrate is not emitted but chemically produced. The precursor emission/concentration/transport/chemistry have to be examined to explain the nitrate.
NO_3 is included in PM2.5 emission dataset. NO_2, one precursor of NO_3, is evaluated in the revised manuscript.

Page 12, line 238: Why is simulation over Texas relevant here?
This discussion is removed.

Page 12, line 242: Be specific on what "SOA processes" is referred here.
This sentence is removed in the revised manuscript because SOA processes are not simulated in our settings.

Page 12, line 244 and 246: Be quantitative – what is the standard of "good agreement"?
Quantitative evaluations are provided in the revised manuscript.

Page 12, line 250: How large is the "large low bias"?
From 30% to 85%. It is clarified in the revised manuscript.

Page 13, line 253-254: "The 4km simulation has better agreement…", but only in the cold season.
It is clarified in the revised manuscript.

Page 13, line 254-255: "The 4km simulation captures seasonal variability of PM2.5 and its speciation": From Figure 4, the seasonal variability for the

PM2.5 species are very similar between the 4- and 20-km simulations, only the concentrations are higher from the 4km simulation. The seasonal variability of PM2.5 sulfate and OC are not capture by both 4 and 20 km simulations.
The seasonal variability of sulfate is not captured in the 4km simulation while 20km simulation has a correlation of 0.63. OM has a correlation of 0.93 for all the simulations. Reworded as "Both the 20km and 4km simulations approximately capture the seasonal variability of PM2.5 and most of its speciation" in the revised manuscript.

Page 13, line 267-268: The 4km_D2 overestimates PM2.5 by 52%, but it overestimates the PM2.5_dust by up to a factor of 4 in the warm season!
The quantitative information is provided in the revised manuscript.

Page 13, line 270-272: As I suggested earlier, please show correlation coefficients on all comparisons (in addition to the bias), which indicates how model and data agree on seasonal variations.
Correlations are provided in the revised manuscript.

Page 14, line 285-286: How much better does 4km_D2 agree with MISR than other simulations? Visually, JAS is still nowhere near MISR, and AMJ is higher than MISR. Please quantify the degree of agreement.
Quantitative information is provided in the revised manuscript.

Page 14, line 290-292: I don't understand the statement of "reasonably capture the vertical distribution", even though the model has "low biases in the boundary layer and high biases in the free troposphere". To me, this is rather "unreasonable".
Reworded as "roughly capture".

Page 15, line 298-299: "…suggesting relative good performance…": How good? Figure shows poor agreement between obs and model for sulfate and OC, so they are not "good" at all.
Reworded as "suggesting that dust is the primary factor contributing to the model biases in aerosol extinction" in the revised manuscript.

Page 15, line 303: How to explain that dust from 4km_D2 is way too high but the extinction in the boundary layer is still way too low?
The model doesn't simulate the unstable environment in the warm season. Although the dust emission at the surface is large in the 4km_D2 run, no

enough convective vertical mixing is produced in the simulations, resulting the low biases in the boundary layer. It is clarified in the revised manuscript.

Page 15, line 313 and 316: If the model has weak vertical mixing, the aerosols should be trapped within the BL and not transported to high altitudes. But the model actually overestimates the aerosol at high altitude – what is the source of high altitude aerosol?
High altitude aerosols are from horizontal transport primarily governed by chemical boundary conditions.

Page 16, line 321-322: This precisely indicates the need to quantify the role of chemical boundary conditions.
The role of chemical boundary conditions is discussed in the revised manuscript.

Page 16, line 323-324, "good performance…": But in JFM the model results are much higher (by a factor of infinity?) at above 1.5 km! How can that be evaluated as "good"?
Changed to "relatively good".

Page 16, line 330: "reasonable simulation", "good representation" – what are the measures of reasonable and good here?
Quantitative information are provided in Table 2 and 3 the revised manuscript.

Page 16, line 337: Please explain what "climatological fire emissions" mean.
Reworded as "monthly-varying fire emissions".

Page 16, line 339-340: Why can Wu et al do it right for South America fire but cannot do it for California? What are the major obstacles?
In our simulation for South America, it is a 7-day case. Daily satellite data are used to generate biomass burning emission. In this study, we are focusing on seasonal variations. Biomass burning emission is updated every month, which cannot capture the single fire event in this case.

Page 17, line 371-372: No need to spell out what GOCART and DUSTRAN stand for at the last part of the paper, since they have been introduced and used many times earlier in the text.
Most people don't read the whole paper, especially program managers. So we have all acronyms redefined to help them immediately understand what we are saying.

Page 17, line 383-385: Unfortunately, I cannot see how the evaluation in this study can be apply to other regions to ensure that aerosols are simulated correctly for the right reasons. This paper has shown the problems but has not shown how to solve the problems with what approach.

This sentence is removed in the revised manuscript.