Review of Miao et al, 2020

This paper addresses the important issue of the difficulty models have in correctly simulating $PM_{2.5}$ composition. This is to my knowledge the most comprehensive evaluation of a CTM's performance in China and provides a valuable starting place for the future investigation of many issues, such as the overestimation of nitrate and the underestimation of NH_3, CO, OH and HO_2. This paper provides an analysis of the performance of the model in all four seasons, which is rare. This paper is well-written and within the scope of ACP and should be published after the minor revisions listed below.

Major Comments.

My main major comment is that the authors should be more careful in stating potential reasons for model biases, and either perform "back-of-the-envelope" calculations, or quick sensitivity tests to support their conclusions. They have already gone to huge effort to perform a large set of sensitivities, but a little more context on the environment in China could be extremely helpful in interpreting the results. It would be particularly good for example to understand when SIA is sensitivity to NH3, or HNO3, or base. Finally, the conclusions could more clearly state the main findings from this work, with the key numbers that highlight their findings, such as improvement from reduced NO2 uptake and the remaining summertime nitrate bias. The minor comments below state additional specific suggestions for this.

Minor Comments.

1. Page 4, line 108. I don't quite understand why you need to divide the observation data by 0.8 to compare to the model. Should the model represent both PM1 and PM2.5, and could you not compare to both?
2. Page 4, line 126 – provide the model doi.
3. Page 5, line 161 – Is there a citation for "the top-down estimates."?
4. Page 5, line 171 – It would be very useful to have a table of relevant emissions totals for comparison by future studies.
5. Page 5, line 189 – What season was evaluated in Fu et al., 2012 and Zhao et al., 2016? Does the conclusion still hold about the improved model performance with the simple scheme if compared by season?
6. Page 5, line 191 – It might be helpful to readers to start a new paragraph discussing the seasonality of the model bias.
7. Page 7, line 200 – Are you saying that there is too much NH4 in YRD because there is too much NOx making too much NH4NO3? I
8. Page 7, line 204 – The lack of model gradient in SO4 between urban and rural sites is striking. Do you have an explanation for this? Is there an urban/rural gradient in SO2 in the model?
9. Page 7, line 207 – Do you expect model resolution to have an effect on the ability to simulate urban aerosol?
10. Page 7, line 208 – Are you saying that winter = haze? It does not appear that you have classified winter data as 'haze'/'not haze', please clarify.

11. Page 7, line 218 – Can you be more specific about the possible causes of the seasonality in the OA bias? Could the seasonality imply an issue with biogenic vs. anthropogenic SOA?
12. Page 7, line 220 – Is it really necessary to show the model biases on a log scale?
13. Figure S3 – what are the red dashed lines?
14. Page 7, line 222 - Is the summer value for PM2.5 really within 30%?
15. Page 8, line 227 – Can you please explain the reasoning for excluding data over 150 ug m^{-3}?
16. Page 8, line 231 - Instead of "insignificant", could you state the model bias?
17. Page 8, line 233 – Could you give us more statistics on the diurnal cycle, it is hard to see in these plots that nitrate and ammonium are "flatter" than sulfate.
18. Page 8, line 243 – could the ratio of nitrate / nitrate + hno3 tell you whether there is an issue with model partitioning at this site?
19. Figure 2 – Why is there a morning peak in wintertime OA?
20. Figure S8 – What are the red numbers? Can you explain the large difference in the median vs. mean difference particularly in winter and fall?
21. Table 1 – The NMB values for wind direction don't make sense, it seems like they should be much larger.
22. Page 9, line 274 – Do you have an explanation for the seasonality in SO2 that the model is missing? Could this be for example from heating sources that the inventory doesn't capture?
23. Figure 4b – Why compare against NO2 and not NOx? The modeling partitioning could also have issues.
24. Page 9, line 279 – Could you run a quick sensitivity test to determine whether say turning off NO2 uptake brings NO2 into better agreement?
25. Page 9, line 285 – Can you calculate whether aerosol is generally sensitivity to NH3 or HNO3 in each season?
26. Page 9, line 287 – You mean in the simple scheme, right? CO wouldn't affect the semi-volatile scheme if I understand correctly?
27. Figure 5 – The uncertainties on these observations, particularly the radicals, are large. Could you put error bars, or shading etc., on the observations?
28. Page 10, line 305 – Not necessarily, depending on conditions, inclusion of clno2 can increase nitrate due to clno2 photolysis – See Sarwar et al., 2014 (GRL).
29. Page 10, line 308 – Jaegle et al., 2018 discusses the Eastern United states in winter, I am confused by this reference here. I also can't find the seasonality you reference in these citations, please clarify.
30. Page 10, line 313 – However the lack of daytime HONO is a model issue – does this provide support for photolysis of nitrate?
31. Page 10, line 318 – Just to clarify, you are aiming to address general model biases, not biases specific to haze?
32. Page 11, line 340 – Did you run all 50 sensitivity simulations at nested resolution? Did you also consider the effect of resolution itself? See Zakoura and Pandis, 2018 (Atmospheric Environment)

33. Page 11, line 341 – Why not test whether scaling up winter and fall CO improves model SOA in the simple scheme?
34. Figure 6 – It might help to have a horizontal line through the 1-1 line so we can see when the model is over or under estimating.
35. Page 11, line 350 – Would the model bias in nitrate impact aerosol water and thus result in overestimated sulfate particularly using the ALWC parameterization?
36. Page 11, line 354 – Isn't this conclusion supported by the extreme model overestimate of HONO in Figure S9?
37. Figure 6, case 8, Why is the median so much more impacted than the mean?
38. Page 12, line 360 – Why not test these things?
39. Page 12, line 362 – Why does increasing sulfate reduce nitrate in the model?
40. Figure 7. I understand the authors aim in Figure 7, but it is difficult to follow. Possibly a table would be easier for the reader to understand.
41. General comment – is there any reason to think that in-cloud oxidation of SO2 is underestimated? Could model cloud biases be part of the issue?
42. Conclusions – it would help the reader to be more specific in the conclusions about the impact of your sensitivities. For example, accurate SO2 emissions result in XX improvement in the model agreement with SO4. Generally, if the authors could put in the conclusions more numbers on their findings, for example, even our most improved model is still biased by XX % in summer, it could help improve citations by future modeling studies.
43. Page 13, line 410 – Can you provide the explanation for this here? Why is this the case?