Dear Editor, thank you for sending the comments to help improve the quality of the paper. We have revised the manuscript to address the reviewer's comments and provided a detailed response to each comment in this file. The comments are in regular fonts and the responses are in red.

Please note that we have performed considerable work in this major revision, but the majority of this effort has confirmed the approach and conclusions of the original manuscript. We have not included new figures in the manuscript discussing wet deposition or other features not directly related to the prediction of ground level concentrations for the reasons discussed below. Any of the figures included in this response document can be included as Supplemental Information in the manuscript if the Editor judges that this would enhance the equality of the paper.

Reviewer's report #1:

Suggestions for revision or reasons for rejection (will be published if the paper is accepted for final publication)

The reviewers brought up a number of important points that should be addressed in the manuscript changes or in future publications.

Response: We thank the reviewer for the comments to help improve the quality of the paper. We have revised the manuscript to address each comment and provided a detailed response below.

Evaluation of the model's ability to simulate deposition (wet vs. dry) is critical for gaining confidence in the model's ability to simulate long term pollutant concentrations. This need not be done for the entire period, but should be done for at least an annual cycle.

Response: Output of wet and dry deposition can be configured on or off based on the purpose of the model simulation. This study aims to provide air quality data for health effects analysis, and it generated over 40 TB data of meteorology, emissions and ground level air pollution concentrations and sources. We did not initially configure the model to save the deposition fluxes in the 9 year model simulations in order to save disk space and because these fields are not relevant to the purpose of the simulations.

In response to the reviewer's comment, we have reconfigured the model to save the deposition fluxes and repeated simulations for the entire year of 2005. We compared the calculated wet deposition fluxes to the available measurements of wet deposition fluxes from National Atmospheric Deposition Program/National Trends Network (http://nadp.sws.uiuc.edu/). No observed dry deposition data is available in California during the 9 year model period.

Figure R1 shows the comparison of predicted vs. observed seasonal wet deposition fluxes of nitrate, ammonium, sodium, chloride, and sulfate at 10 available sites in California in 2005. As the precipitation is a critical driver in the wet deposition calculations, under predicted precipitation by WRF can translate into biases in the model estimated wet deposition flux. We then followed the method proposed in Appel et .al (2011) (equation 1) and calculated bias corrected wet deposition flux (Sim_BC). The results show that nominal predictions for wet deposition fluxes are lower than observed fluxes for nitrate, ammonium, sodium and chloride, but higher for sulfate. With the precipitation bias correction, model performance for wet deposition is greatly improved for nitrate, sodium, and chloride, but becomes worse for ammonium and sulfate.

Previous studies have determined that in-cloud scavenging is an efficient for deposition of fine aerosols (Chang et al. 1987, Seinfeld and Pandis, 2006), and our analysis confirms that in-cloud scavenging dominates the wet deposition in the current study. The majority of the clouds in the California simulations had altitudes > 1km above ground level (agl). Therefore, the wet deposition fluxes were primarily affected by the airborne particle concentrations present > 1km agl, not in the breathing zone near the surface. Further analysis shows that background particles advected into the study region account for the majority of the PM2.5 present above 1km agl in the current study. MOZART global model outputs were used as boundary conditions for the wet deposition simulations, and so the reported deposition fluxes mostly reflect the deposition of concentrations predicted by MOZART due to emissions outside of California. It is possible that MOZART over predicts sulfate and ammonium concentrations and under predicts chloride concentrations during this model period.

While the analysis of wet deposition is interesting, it has little relevance for the accuracy of concentrations predicted at the surface in the breathing zone for the reasons discussed above. We provide this information to alleviate the concerns of the reviewer, but have not included the discussion in the revised manuscript since it is not central to the focus of the current paper.

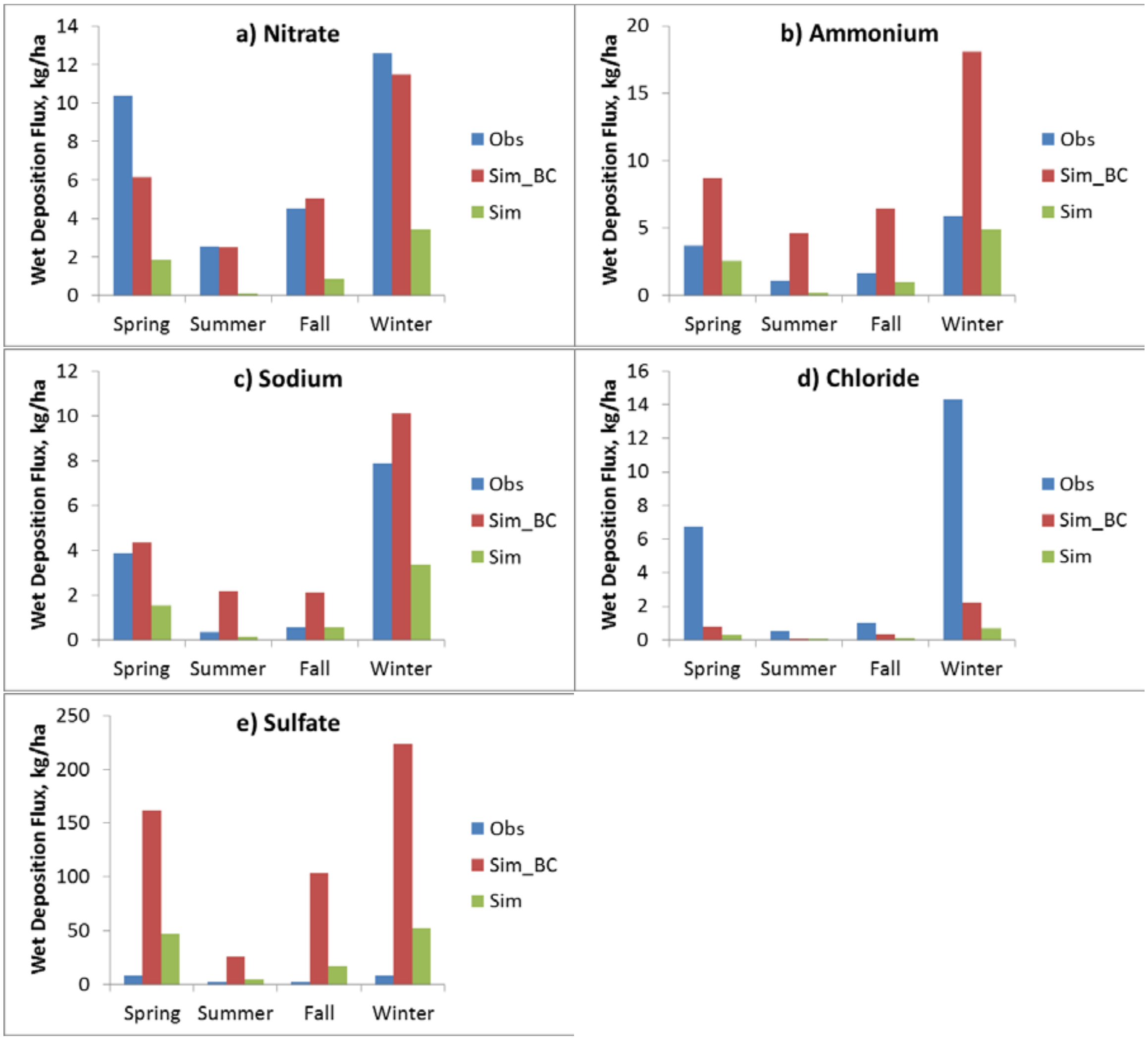

Figure R1. Seasonal observed and predicted wet deposition flux of (a) nitrate, (b) ammonium, (c) sodium, (d) chloride, and (e) sulfate in 2005. 'Obs' represents the observed wet deposition flux, 'Sim' represents the model simulated wet deposition flux with WRF predicted precipitation, and 'Sim_BC' represents the model simulated wet deposition flux with precipitation bias adjusted based on the observed and WRF predicted precipitation amount at individual sites.

More justification is needed to modify both the friction velocity and relative humidity.

Response: We have included two additional references (Mass and Ovens, 2010; Wang et al., 2014) to document the scientific basis for increasing the friction velocity by 50%. In the reference of Mass and Ovens (2010), the authors performed extensive testing and analysis of various methods to improve the surface wind biases in WRF, e.g. experimenting with all available planetary boundary layer schemes and various vertical diffusion schemes. None of these approaches solved the surface wind problem. Mass and Ovens (2010) tested WRF simulations with very fine grid resolution at 1.3 km, and found the wind bias was reduced compared to simulations conducted at 4, 12 and 36 km. This lead to a hypothesis that the wind speed over predictions were caused by WRF not resolving subgrid scale roughness elements even at 4 km resolution. Mass and Ovens then tested the effects of increasing surface drag by increasing u* by 50%, and found that this approach decreased the wind biases significantly.

The second reference of Wang et al. (2014) is part of the Air Quality Model Evaluation International Initiative (AQMEII) Phase II model intercomparison study suggested by the reviewer. In this study, the authors also implemented the same method (increasing u* by 50%) in their WRF/Chem simulations, and reported a large improvement of predicted wind fields.

In the current manuscript, we also examined the improvement of wind biases due to increasing u* by 50%. Figure R2 shows the wind speed (WS) biases as a function of wind speed (a) predicted by the original version of WRF and (b) with increasing u* by 50%, respectively. The results indicate that increasing u* by 50% reduces the positive bias for winds speeds less than ~3 m/s (illustrated by the green straight line in the figure), but increases the negative bias at higher wind speeds. However, analysis of the wind speed distribution in California (Figure R3) shows that 78% of winds are less than 3 m/s, and virtually all air pollution episodes occur when wind speed is less than 3 m/s. Therefore, increasing u* by 50% in our study improves the wind predictions for a majority of cases during the modeling period. We believe the evidence clearly shows that our approach is in line with methods employed in other state-of-the-science studies and is not an over-adjustment.

The discussion above has been briefly summarized along with the two additional references in the 1st paragraph of section 2.2 of the revised manuscript.

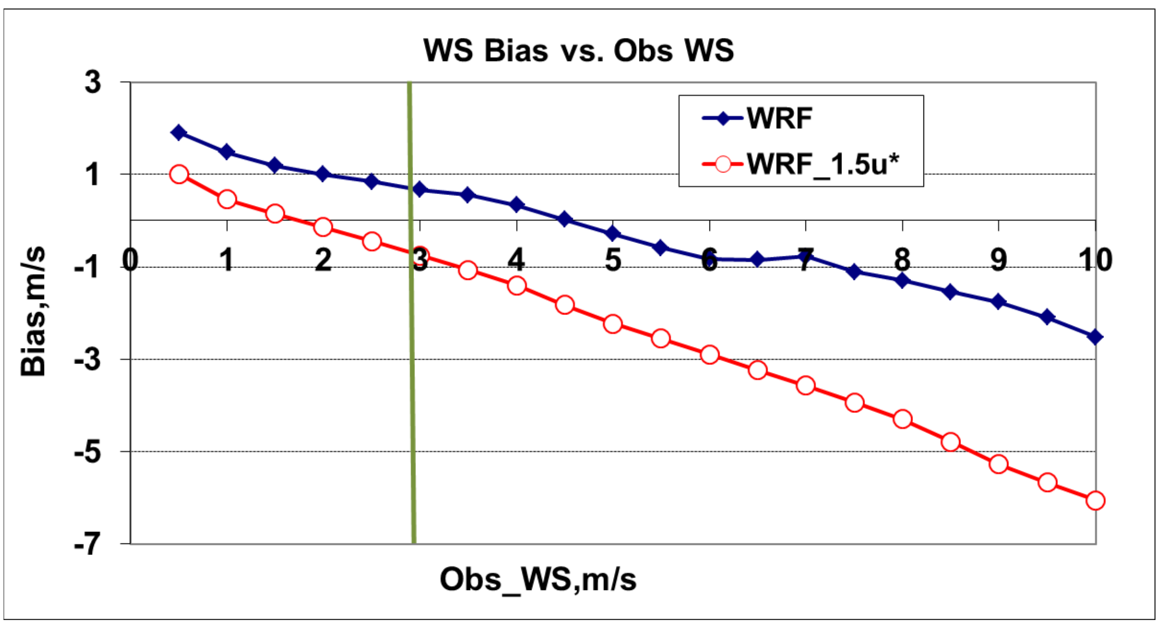

Figure R2. Wind speed (WS) bias from the original version of WRF (WRF) and increased u* by 50% (WRF_1.5u*) predictions for different wind speeds.

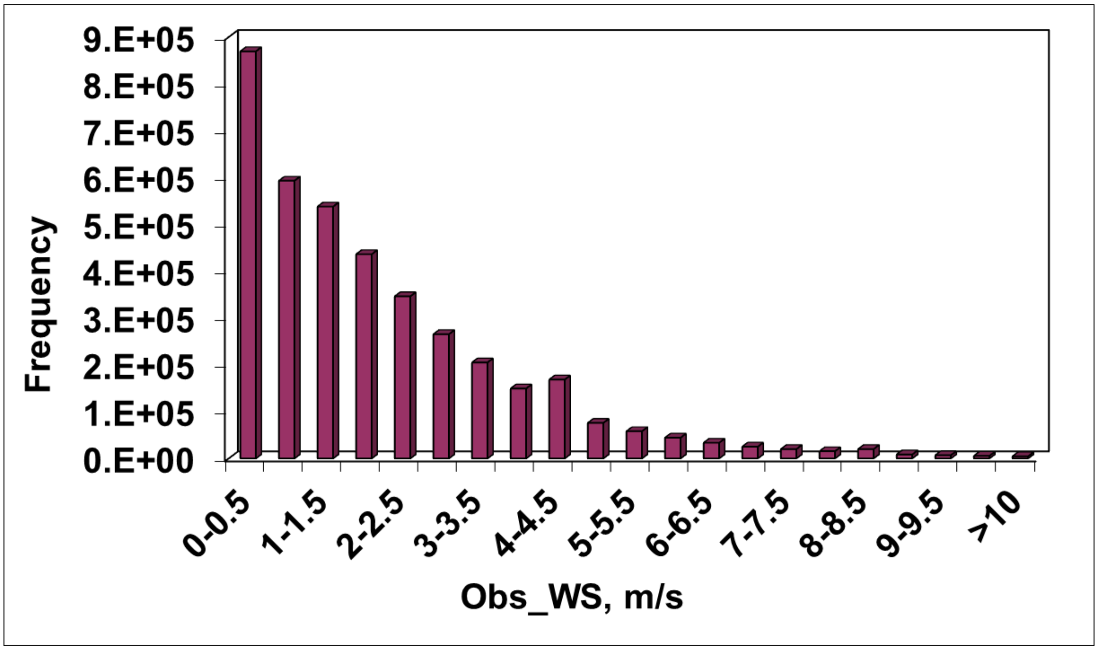

Figure R3. Wind speed (WS) frequency distributions from the surface station meteorology measurements.

The Reviewer also raised a question about modification of humidity in the model simulations. We think the reviewer misunderstood the purpose of raising the humidity during our sensitivity analysis. We did not adjust relative humidity in our 9-yr model simulations. All the model results for spatial and temporal comparison to observations are based on the relative humidity directly predicted by the WRF model. Recognizing that relative humidity was under predicted and that in turn could affect predictions of nitrate and other pollutants, we then conducted a 1-yr sensitivity simulation in which we increased relative humidity by 30% to examine how the pollutant concentration predictions responded to relative humidity. Based on the evaluation statistics for relative humidity (Hu et al., 2014, Table S1), the mean fraction biases (MFB) of relative humidity varies from -11% to -38% and root mean square errors (RMSE) varies from 21% to 36% in different air basins in California. The overall MFB and RMSE for the entire state is -27% and 31%, respectively. For this reason, we chose 30% in our sensitivity analysis. The results with 30% increased relative humidity are not used outside the sensitivity analysis in our study and these results will not be carried into other analyses or into the following health effect studies.

As a second important point, we agree with the reviewer that the under-prediction in relative humidity could also be associated with an under-prediction of precipitation in the WRF simulations. Based on the reviewer suggestion, we calculated the WRF model performance for precipitation. Table R1 shows the statistics of the precipitations during the 9 years. Not surprisingly, the precipitation was under-predicted, with a MFB of -76.1% and RMSE 2.84

mm/hr. Figure R4 shows the occurrence frequency of different precipitation events. The number of rain events with precipitation > 0.1 mm/hr predicted by WRF generally matched observations, but the number of rain events with precipitation > 0.5 mm/hr were under predicted. In California, rain events occurred about ~21% of the hours during the 9 year model period. PM2.5 concentrations decrease rapidly during all rain events with precipitation rates > 0.1 mm/hr, and so the effects of under-predicting the frequency of high-precipitation events is minor.

We acknowledge that wet deposition during rain events was somewhat under predicted (see detailed analysis in the next response) in this study. This has the potential to compensate for other possible errors associated with meteorological parameters (e.g., relative humidity and wind speed), but these factors are minor and will not change the conclusions of the study.

We have included the discussion of WRF model performance for precipitation in the 1st paragraph of section 2.2 of the revised manuscript so that readers are well informed about the potential magnitude of biases.

In summary, the friction velocity has been examined in multiple previous studies and we have adopted methods that are consistent with the results from those studies. The relative humidity perturbation carried out in this paper was only performed as a sensitivity study. The results of this perturbation were not carried into the main results.

Table R1. Statistics of WRF precipitation during 2000-2008 in California.

Mean Obs (mm/hr)	Mean Sim (mm/hr)	MFB (%)	MFE (%)	RMSE (mm/hr)
1.80	0.84	-76.1	116.1	2.84

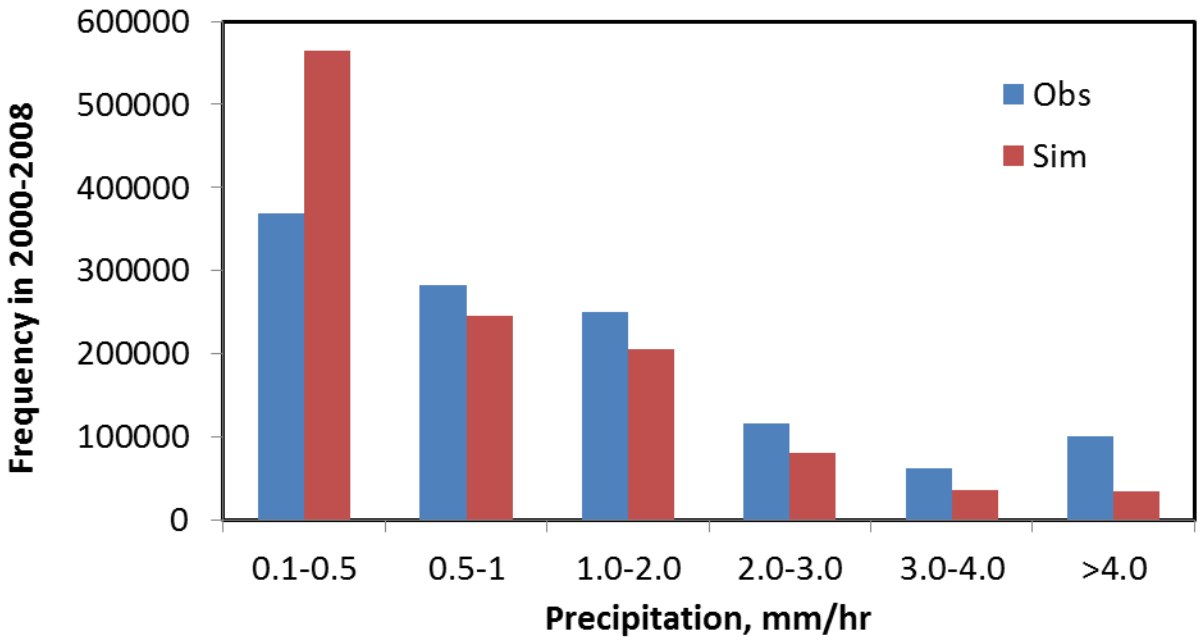

Figure R4. Occurrence frequency of rain events of different precipitation amounts.

As noted by one of the reviewers, "Could the model not also be greatly over- and under-predicting moderate actual concentrations, as long as it did so with approximately equal frequency or over-predict consistently in certain areas and under-predict in others?

Response: We do not think that is the case because nitrate is consistently under-predicted in most sites." They have the results so should be able to address the question in detail. This should be done.

Response: Figure R5 shows the mean fractional bias (MFB) in $PM_{2.5}$ nitrate and NO predictions as a function of the measured concentrations. The results clearly show that MFB is lowest for concentrations in the middle of the observed range. The greatest MFB values occur for the largest concentrations and the lowest concentrations at the extreme ends of the distribution.

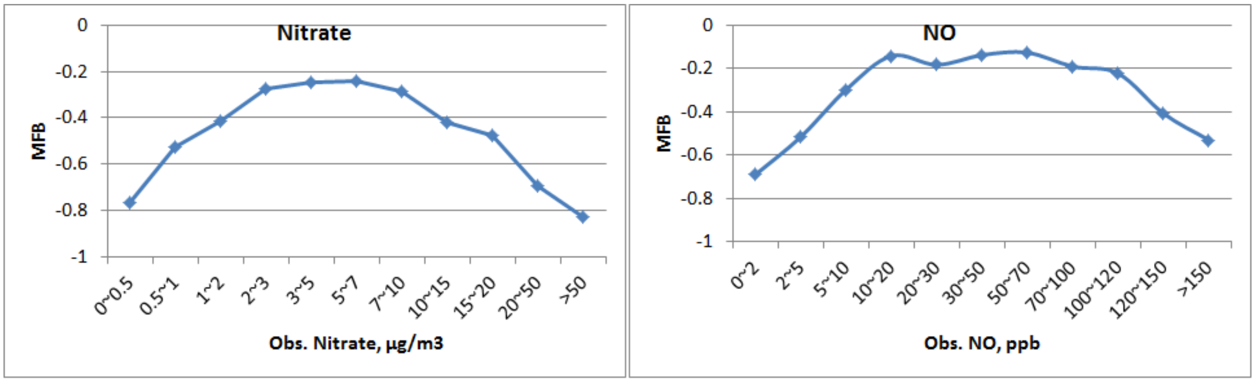

Figure R5. Mean fractional bias (MFB) for $PM_{2.5}$ nitrate and NO as a function of measured concentrations.

Figure R6 displays the frequency of underprediction/overprediction ratios of nitrate at the ambient measurement sites. Analysis shows that 11 out of 15 sites have nitrate underprediction for over 60% data points, and only 3 sites have underprediction rate close to or less than 50%. Averaging over all sites, about 2/3 of total data points are underpredicted, and 1/3 of total data points are overpredicted.

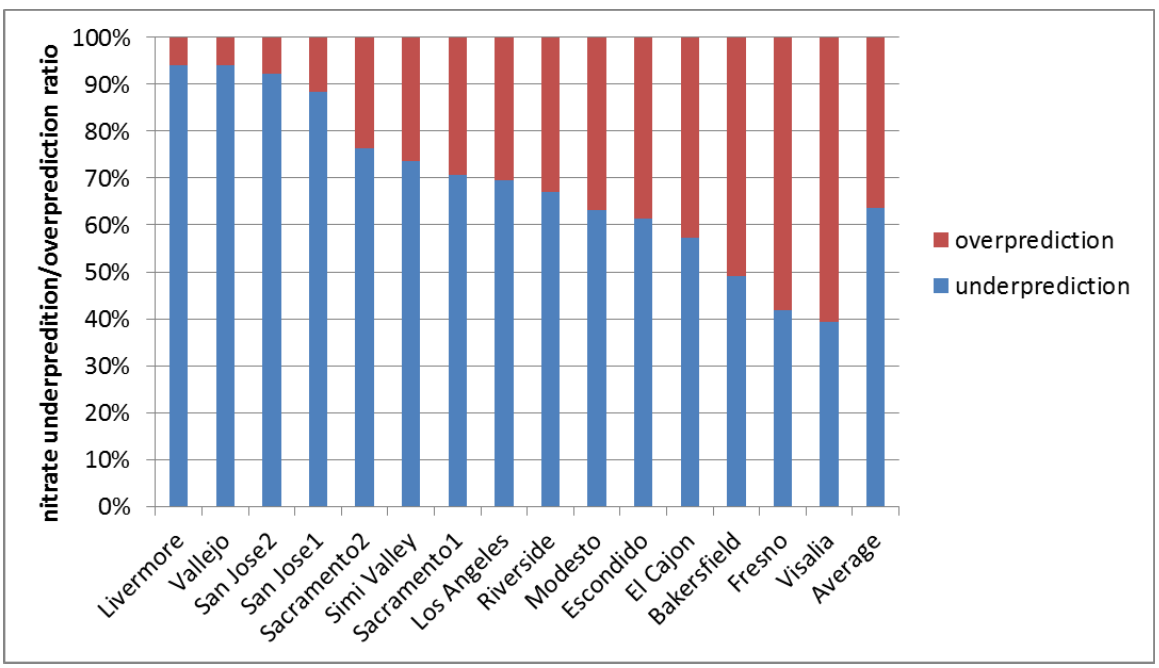

Figure R6. Underpriction/overprediction of nitrate at individual ambient measurement sites.

There are 114 NO measurement sites and so a figure comparable to R6 is impractical for this species. In summary, 73 out of 114 sites (64%) are under-predicted, and 41 sites (36%) are over-predicted. Analysis of individual hours reveals that 58% of the hours NO is under-predicted, and 42% of the hours NO is over-predicted.

The results summarize above confirm that the relatively low mean fractional bias and relatively high mean fractional error reported in the manuscript are not caused by equal frequencies of underprediction and overprediction of moderate nitrate concentrations or underpredictions at some locations counter-weighted by overpredictions in other locations. Moderate concentrations are predicted most accurately and extremely high / low concentrations are underpredicted, as stated in the manuscript.

A more extensive evaluation effort, as outlined in the AQMEII work, should be conducted.

Response: We have studied four AQMEII modeling evaluation papers to identify the methods used for model evaluation (Campell et al., 2014, Hogrefe et al., 2014, Wang et al., 2014, Wang et al., 2015). In these studies, domain-wide performance statistics, spatial distributions, temporal variations, and site-specific analyses was conducted using all available surface and satellite data. We have performed similar evaluation analyses using the available surface observation data in this manuscript. We didn't evaluate our model performance against satellite data because (1) we focused on the health-related ground level concentrations, and (2) we only saved ground level concentrations for the purpose of evaluating concentrations in the breathing zone with the minimum disk space requirement.

Reviewer's report #2:

Suggestions for revision or reasons for rejection (will be published if the paper is accepted for final publication)

The reviewer thanks the authors for so thoroughly addressing the comments and questions raised, and modifying the manuscript text to include an expanded discussion of the sulfate bias, and the effects of longer averaging times. One final comment related to the discussion of the impacts of nucleation on the observed sulfate bias in Northern California:

This reviewer agrees that the sulfate mass that can be directly attributed to nucleation will pale in comparison to that associated with condensational growth. However, in these remote areas, the background aerosol concentration is presumably low, and the increase in number concentration from nucleation would provide more 'seed' aerosols onto which sulfuric acid can condense, resulting in (possibly) a more noticeable increase in total sulfate mass.

The comment above is fairly minor, and little modifications need be made to the manuscript text to address it. No need to re-review. It is recommended that the paper be published after this minor, but important, clarification.

Response: We thank the reviewer for the comment on the impacts of nucleation on sulfate mass in Northern California where sulfate is generally low. We have added one sentence in the manuscript in the 3rd paragraph of section 3.1 to summarize the point.

"In the remote areas where the sulfate concentrations are low, the omission of nucleation processes in the current study could reduce seed aerosol surface area onto which sulfuric acid can condense. This factor could contribute to the under-prediction of sulfate mass in these regions along with the missing sulfur sources."

References:

Appel, K.W., Foley, K.M., Bash, J.O., Pinder, R.W., Dennis, R.L., Allen, D.J., Pickering, K., 2011. A multi-resolution assessment of the Community Multiscale Air Quality (CMAQ) model v4.7 wet deposition estimates for 2002-2006. Geoscientific Model Development 4, 357-371.

Campbell, P., Zhang, Y., Yahya, K., Wang, K., Hogrefe, C., Pouliot, G., Knote, C., Hodzic, A., San Jose, R., Perez, J., Guerrero, P.J., Baro, R., Makar, P., 2014. A multi-model assessment for the 2006 and 2010 simulations under the Air Quality Model Evaluation International Initiative (AQMEII) phase 2 over North America: Part I. Indicators of the sensitivity of O3 and PM2.5 formation regimes. *Atmospheric Environment* http://dx.doi.org/10.1016/j.atmosenv.2014.12.026.

Chang, J.S., Brost, R.A., Isaksen, I.S.A., Madronich, S., Middleton, P., Stockwell, W.R., Walcek, C.J., 1987. A three-dimensional Eulerian acid deposition model: physical concepts and formulation. *Journal of Geophysical Research* 92, 14681e14700.

Hogrefe, C., Pouliot, G., Wong, D., Torian, A., Roselle, S., Pleim, J., Mathur, R., 2014. Annual application and evaluation of the online coupled WRF-CMAQ system over North America under AQMEII phase 2. *Atmospheric Environment* http://dx.doi.org/10.1016/j.atmosenv.2014.12.034.

Hu, J., Zhang, H., Chen, S.-H., Wiedinmyer, C., Vandenberghe, F., Ying., Q., Kleeman, M.J., Predicting Primary PM2. 5 and PM0. 1 Trace Composition for Epidemiological Studies in California, *Environmental science & technology*(2014), 48 (9), 4971-4979.

Mass, C. and Ovens, D. (2010), WRF model physics: progress, problems, and perhaps some solutions, the 11th WRF Users' Workshop, 21–25 June, Boulder, CO, 2010.

Seinfeld, J.H., Pandis, S.N., 2006. Atmospheric Chemistry and Physics: From Air Pollution to Climate Change, second ed. John Wiley & Sons, Inc., New Jersey, USA.

Wang, K., Yahya, K., Zhang, Y., Hogrefe, C., Pouliot, G., Knote, C., Hodzic, A., San Jose, R., Perez, J., Guerrero, P.J., Baro, R., Makar, P., Bennartz, R., 2014. A multi-model assessment for the 2006 and 2010 simulations under the Air Quality Model Evaluation International Initiative (AQMEII) Phase 2 over North America: Part II. Evaluation of column variable predictions using satellite data. *Atmospheric Environment* http://dx.doi.org/10.1016/j.atmosenv.2014.07.044.

Wang, K., Zhang, Y., Yahya, K., Wu, S.-Y., Grell, G., 2015. Implementation and Initial Application of New Chemistry-Aerosol Options in WRF/Chem for Simulating Secondary Organic Aerosols and Aerosol Indirect Effects for Regional Air Quality. *Atmospheric Environment* doi:10.1016/j.atmosenv.2014.12.007.