Review of "Simulation of trace gases and aerosols over the Indian domain: evaluation of the WRF-Chem Model" by Michael et al.

This paper describes and evaluates the Weather Research and Forecasting Model coupled with Chemistry (WRF-Chem) for the Indian region. The paper compares predicted ozone, carbon monoxide, sulfur dioxide, aerosol optical depth, aerosol extinction, and black carbon concentrations with primarily ground-based monitoring observations. It is great to see more regional-scale simulations being performed for the south Asia region, a region whose air quality is very important to millions of people and a region that has an impact on global climate via the Asian monsoon.

The authors reveal several discrepancies between the model and observations and suggest reasons for those discrepancies. There is essentially no quantitative evaluation, only qualitative discussion. When suggested reasons are given for the discrepancies, there is no follow-up to test the idea with sensitivity simulations. Further, I think the authors do not perform a good interpretation of their results, especially for the comparison with aircraft observations. Lastly, the article did not present new science except for the comment in the last paragraph of the paper. The authors should either focus on that science by reducing (or putting in supplementary material) the evaluation section, or the paper should be submitted to another journal, e.g. the sister journal to ACP, *Geoscientific Model Development*, which publishes numerical model descriptions, developments and evaluations of the atmosphere, and other geosciences.

Major comments:

1. The potential science question for the paper is whether there is a link between aerosols and summer monsoon precipitation based on interannual differences. Why does a model need to be used to answer this question? I suggest either focusing the paper on the model evaluation and submit it to a journal with less science focus, or expand on the science question by using the model to answer questions on what processes contribute to the interannual differences. (It could be that large-scale weather patterns change the monsoon precipitation amounts that then affect aerosol concentrations.)
2. In evaluating the WRF-Chem results, the authors need to be more quantitative and use the statistical measures that are presented in section 5.1.
3. Sensitivity simulations must be done in order to show that wet scavenging or dust emissions are affecting the model results for the comparison locations.

Specific comments:

1. Page 12290-12291. In the discussion about previous work, especially that by Kumar et al. (2012a,b), it would be good to stress that their WRF-Chem simulations use a different WRF-Chem configuration. By running a different chemical mechanism and aerosol scheme, you need to conduct a model evaluation. In the model description, it would be good to add a few sentences

contrasting the current WRF-Chem configuration to previous studies of the same region.

2. Page 12292, lines 19-24. This paragraph touches on the potential science question of whether there is a link between aerosols and summer monsoon precipitation based on interannual differences. The authors state how years 2008, 2009, and 2010 compared to the climatological average, but reported a quantitative difference (23%) for only one year. What is the average precipitation for a summer monsoon? What region is being considered (all of India, the Indo-Gangetic Plain (IGP))? For 2008 and 2010, what is the percent difference in their monsoon precipitation from the average?
3. In the results section, the authors suggest that wet deposition and dust emissions are possible causes for certain behavior seen in the model results. However, there is no description of either process. Please describe the following in the model description:
    a. Wet deposition
    b. Dust emissions (and sea salt if it is used)
    c. Initial and boundary conditions for the chemical species
4. Page 12295, lines 8-16. Why is only one station used for trace gas evaluation? It would be better to evaluate at several locations (as is done for AOD) to learn whether spatial patterns are being represented well. Another way of evaluating spatial patterns is to compare model results with satellite observations.
5. Page 12295, lines 24-30. I suggest listing the Aeronet station locations in a table. It would also be good to see these locations on a map.
6. Page 12297, lines 1-30. Currently, the introduction to mean bias, correlation coefficient, and root mean square error is not necessary because these statistics are not included in the subsequent model evaluation. However, I recommend keeping this section and using these measures for quantitatively evaluating the model.
7. Section 5.1, evaluation of the meteorology. I wonder if the model is drifting away from reality during the 15-day simulation. Is there any data assimilation or nudging of large-scale temperature, water vapor, and winds to observations during the simulation? Does a time trace of these meteorological parameters (e.g. temperature at 2 meter height) show good results early, and worse agreement later in the simulation?
8. Section 5. A time trace of the Kanpur meteorology (T, q_v, winds) would be useful for interpreting the trace gas results.
9. For the comparisons to station data, are model results from the nearest grid point examined, or are a few model grid cells averaged and compared, in case the phenomena is near the station but not exactly at the station location?
10. As the authors point out, there clearly is a difference in the results before and after mid-June. The authors attribute this difference to the lack of wet scavenging in their model run. I have several comments and suggestions.
    a. WRF-Chem V3.4 and later includes wet scavenging in the sub-grid convective scheme. It would be better to do this evaluation with this

version of the model. Further, the resolved-scale wet scavenging should also be turned on.

b. To show that wet scavenging is causing these insoluble species (CO and O_3) to decrease, then two simulations should be compared; one with wet scavenging and one without wet scavenging.
c. A monsoon definition says that the winds shift causing more transport of moisture and more precipitation in a region. It would be good to see if the ~200 ppbv decrease in CO and ~15 ppbv decrease in O_3 is correlated with the wind shifting from northwesterly to easterly (from Bay of Bengal where the storms begin). The decrease of these trace gases could simply be that the air is coming from a cleaner region.
d. While the authors wisely looked at May versus June in their results, the June averages are still a mixture of pre-monsoon and monsoon values. I suggest analyzing May to June 20 (i.e., actual day before monsoon starts) separately from the last couple of weeks of June.

11. In many of the figures, there are plotted averages and "error bars". I suspect the authors mean standard deviations, because an error would be associated with the measurement or model method. Here, I think they are simply finding the variability associated with these averages.
12. Aerosol results. Please quantify the evaluation with statistical methods (mean bias, RMSE, correlation coefficients). Another method would be to make frequency distributions of AOD from the measurements and model results.
13. Page 12303, lines 7-9. Does the Bian et al. (2011) study calculate dust emissions in the same manner as the current study? If so, then it is a plausible argument. However, one must be cautious because the dust emission parameterizations are very uncertain and must be adjusted to represent a particular region well. For example, the Shaw et al. (2008) *Atmos. Environ.* parameterization in WRF-Chem (one scheme of 3 dust schemes in WRF-Chem) is appropriate for dust in northern Mexico, but will not represent dust well in the nearby U.S. I would expect that it needs to be adjusted as well for the Thar desert dust emissions.
14. Page 12304, lines 1-13. The authors suggest that disagreements between model and observation AOD are from the misrepresentation of dust in WRF-Chem. Besides describing the dust emissions scheme in WRF-Chem in the earlier section, I suggest conducting a simulation with a greater dust emissions (see Kumar et al., 2013, ACPD). If the error is suspected to be the boundary conditions for dust, then I suggest conducting a simulation where these boundary conditions are better represented.
15. Page 12305, lines 22-28. The authors are discussing previous studies of how PBL schemes can affect WRF-Chem results. I think it's important to point out that the regions where these PBL comparisons were done (Houston and the Central Plains of the U.S.) do not have similar topography as the IGP. It would be really nice to see an evaluation of PBL schemes on trace gas and aerosol concentrations for the IGP region.

16. Page 12306, lines 8-29. I disagree with the interpretation of the aircraft profile results. There is only a small agreement between model results and measurements and that is near the surface. I do not see any agreement above the boundary layer. It looks to me that WRF-Chem is missing a lot of the variability observed, likely because of the grid spacing of the model. One last point is that these vertical profiles are from 0 to 7 km, even for Nainital, which is located at a 2 km elevation. Shouldn't these altitude plots be on a mean sea level reference?

Technical comments:

1. Page 12292, line 8. "compliment" should be "complement"
2. Page 12292, line 10. "had been" should be "was"
3. Page 12292, line 19. Use "climatological average" instead of "climate normal"
4. Page 12293, lines 1-14. These sentences match the wording in Tuccello et al. (2012). I suggest rewriting the description.
5. Page 12293, line 16. It would be nice to have horizontal resolution in km
6. Page 12294, line 18-19. Is a polar stereographic projection correct? The model domain looks like either Mercator or Lambert projection, but this paragraph is talking about emissions.
7. P. 12295, line 6. "was" should be "were"
8. P. 12296, line 6. "collocated" should be "co-located"
9. Page 12299-12300. Watch the significant digits being reported. For example, the detection limit of O_3 measurement is 0.5 ppbv, so two digits after the decimal point is not needed. One digit is sufficient.
10. Results section. Should be "data are", not "data is"
11. Figures 7-12 would be much easier to read if there were separate panels for different years (or locations).