Review comments on "Modeling diurnal variation of surface $PM_{2.5}$ concentration over East China with WRF-Chem: Impacts from boundary layer mixing and anthropogenic emissions" by Du et al.

This manuscript provided an interesting study investigating the factors that affect the WRF-Chem performance of diurnal cycle of surface $PM_{2.5}$ concentrations in East China. This study suggested that PBL mixing coefficient is the key factor controlling the WRF-Chem model performance instead of PBL height and the diurnal cycle and injection height of emissions. This manuscript points out the importance of improving the PBL mixing process in WRF-Chem to achieve better results of diurnal $PM_{2.5}$ cycle. The topic is applicable for Atmospheric Chemistry and Physics. The text is concisely written and well documented. This study has comprehensive analysis and detailed explanation/discussion. However, this manuscript only discussed the normalized diurnal amplitude (i.e, DI) from observations and WRF-Chem simulations, but lacked the evaluation of model performance of actual $PM_{2.5}$ concentrations in East China. The comparison of monthly mean $PM_{2.5}$ concentrations between CTL1 and EXP1 (Figure 12) suggested that WRF-Chem had moderate model performance in winter (Jan) and poor performance in other 3 seasons (Apr, Jul, and Oct) even after increasing the PBL mixing coefficient. Note in winter the PBL is usually shallow and if we assume the MEIC emissions are accurate, my hypothesis is that the accuracy of PBL height simulations dominated the model performance of surface $PM_{2.5}$ in these 3 seasons. We cannot rule out that the PBL height was significantly underestimated in WRF-Chem, while increasing the PBL mixing could slightly improve the surface $PM_{2.5}$ simulations. The current manuscript claimed that the mixing coefficient is more important than the PBL height, but it did not show any comparison of PBL height from observations and simulations. I suggested the authors added 1) evaluate $PM_{2.5}$ simulations especially the spatial distribution with observations; 2) evaluate the WRF-Chem simulated PBL height with sonde or LIDAR observations to rule out the possibility that PBL height is more important. In summary, major revisions as indicated in the comments and remarks below are needed before consideration of publication in ACP.

Detailed Remarks/Suggestions for Revision

Line 48: 'significantly overestimated' the $PM_{2.5}$ concentrations compared with observations?

Line 72: The format of reference such as 'Davidson C I et al.' should be 'Davidson et al.'. Please correct all the references based on ACP bibliography requirement.

Line 94: Should 'Hu et al. (2017)' be 'Hu et al. (2016)'?

Line 96: 'CAMQ' should be 'CMAQ'

Line 294: In spring (Figure 2), I don't see two peaks of DI from observations. There is only one peak around 8 am in the morning. Similarly, the 'two peaks' in Oct is also not very clear. If the black dots are centered on 8 am, only one peak exists.

Line 305: I think the Figure S1 should be moved to the main article. It provided the spatial evaluation of DI from observations and simulations.

Line 328-329: I don't quite understand how the contribution was calculated. Looks like it is a unique function in the USTC version of WRF-Chem. Please add further explanation here.

Line 332: Similar as above, please explain how the tendency was calculated.

Line 355: Figure 5 shows the chemical composition of $PM_{2.5}$ simulated by WRF-Chem. Are there observations available to verify the accuracy of WRF-Chem?

Line 358: What are 'OM' and 'OIN'? I guess OM is equal to OC in Figure 5, while OIN is the total of NH4, NO3, and SO4?

Line 378-379: More vertical layers in the lower atmosphere could also better simulate the PBL height, which influence the surface $PM_{2.5}$ concentrations. The authors should evaluate the PBL height with observations (if available) or at least make sure the PBL heights simulated in all WRF-Chem runs are comparable.

Line 412-413: As above, this statement is only valid if the PBL heights simulated in CTL1-3 runs are comparable.

Line 434: Hard to tell from the figure, but I think the PBL height from CTL 1-3 are substantially different. For instance, at 8 am in Apr with the maximum surface PM2.5 simulated, CTL 3 has much lower PBL height as compared with CTL1 and 2. Again, if observations such as LIDAR are available, it will be great to see which run has the better PBL height. Same as in Figure 7, I suggest plotting all the PBL height from WRF-Chem runs in one figure to see the difference.

Line 474: Any reference or data to support the 5 m^2/s rate is reasonable?

Line 575-576: Does MEIC inventory treat power plants as point sources? For instance, US EPA has lat/lon information and hourly emission rate to process power plants as point sources in SMOKE. If not, how to support this argument?

Line 588: The current version of Section 4 is too long. A concise summary of the paper is needed. I suggest the authors to revise this session.

Line 1183: The unit is %?

Line 1187: Plots in Figure 1a are small and hard to see the details. Please remove the black box and symbols for cities in NO_x and SO_2 plots.

Line 1245: Surprised to see the chemistry contributed so little in the surface $PM_{2.5}$ concentrations. Does 'chemistry' stand for chemical evolution such as formation of SOA?

Line 1373: A map of data in Figure 12 is needed, maybe in the SI to show the spatial performance of WRF-Chem.