Dear referee reviewer #1:

Many thanks to your insightful comments and valuable suggestions on our manuscript "Multi-model simulations of springtime dust storms in East Asia: Implications of an evaluation of four commonly used air quality models (CMAQ v5.2.1, CAMx v6.50, CHIMERE v2017r4, and WRF-Chem v3.9.1)". After careful discussions with other co-authors, we have carefully revised our manuscript, and written this point-to-point response letter here.

The revised manuscript with revision mode has been supplemented to this response letter and presented in the same compressed zip file. The revised or added contents are listed as follows (words in red are the responses):

1. The methodology, discussion, and results sections of this manuscript primarily focus on differences between the dust emission treatments used in each model simulation; however, the individual model descriptions provided in section 2.4 provide little to no information about the algorithms comprising these schemes. There really needs to be a succinct summary of the dust emission scheme equations discussed in this paper, either directly in the text or in the appendix section. Suggest using a model flow chart similar to the approach used in Darmenova et al. (2009) or LeGrand et al. (2019) for each dust emission scheme discussed and a symbology table.

Response: We want to thank the reviewer for the constructive and insightful advice. We realize that a succinct summary of the dust emission scheme is needed in our manuscript. According to your suggestion, we further introduced the dust emission schemes (such as the algorithms of dust flux and relevant parameters) used in each air quality model in Section 2.4. The differences between the dust schemes are also described. The flow charts for all dust emission schemes including equations, relevant literature and required input parameters, as well as variable lists are provided in the Supplementary file (Fig. S1~S6 and Table S1~S6).

2. The authors state that WRF v3.9.1 was used to generate the meteorological fields used to force all of the dust models discussed in the manuscript. This is confusing. WRF-Chem is an inline model. The dust emissions and airborne concentrations evolve simultaneously with the atmospheric conditions. In other words, the dust modules in the WRF-Chem assessments were likely subject to different environmental forcing conditions than those in the CMAQ, CHIMERE, and CAMx dust modules. Did the authors use the coupled WRF-CMAQ implementation as well? What was the output frequency of the WRF v.3.9.1 output (wrfout) files? This could potentially have significant influence on the results. Furthermore, are the CHIMERE and CMAQ dust modules configured to ingest windspeed (U) or friction velocity (u_*)? The dust emission calculations described in this paper, with the exception of the WRF-Chem GOCART dust emission scheme, are calculated in terms of u_*. Are the u_* fields being ingested by the dust emission flux equations in WRF-Chem, CMAQ, and CHIMERE identical? If so, please add a figure showing the surface U and u_* fields for a few time periods in the case study sequence. If not, please add a figure showing how they vary (especially if each model is doing its own U to u_* conversion) as this could be important for deciphering causative factors in model output discrepancies.

Response: Thanks your hard works and these valuable comments. We found that this sentence was not correctly expressed in this part of the manuscript. WRF v3.9.1 was used to generate the meteorological fields. The output was used to drive the air quality model of CHIMERE, CMAQ and CAMx. As to WRF-Chem, it is an inline model and dust emissions are calculated simultaneously with the atmospheric fields. Therefore, we cannot say that it is driven by the WRF meteorological fields. In the manuscript, it is revised as "The Weather Research and Forecasting (WRF) model version 3.9.1 was used to conduct the meteorological simulations, then to provide the hourly meteorological output fields to drive the air quality models of CHIMERE, CMAQ and CAMx while the chemistry module of WRF (WRF-Chem) was conducted simultaneously with the meteorological fields."

We don't use the coupled WRF-CMAQ implementation in this study. The output frequency of the WRF v.3.9.1 file for these four models is one hour.

Among these 4 air quality models, the u_* fields calculated by WRF is only used in WRF-Chem model for dust flux calculation while other 3 models implement u_* calculation independently. In CHIMERE v2017r4 model, u_* is calculated according to $u_* = \frac{\kappa u_{10}}{\ln(\frac{10}{z_0})}$, where u_{10} is wind speed at 10m, z_0 is roughness length and κ is the Karman constant. In addition, the equation of Weibull distribution ($p(|u_{10}|) = \frac{k}{A}\left(\frac{|u_{10}|}{A}\right)^{k-1} e^{[-(\frac{|u_{10}|}{A})^k]}$, where k is a dimensionless shape parameter and A is modeled wind speed, in meters per second.) is introduced for wind speed adjustment (Cakmur et al., 2004; Pryor et al., 2005). The friction velocity in CMAQ v5.2.1 was calculated based on an updated dynamic relation ($\frac{z_0}{h} = \begin{cases} 0.96\lambda^{1.07} & \lambda < 0.2 \\ 0.083\lambda^{-0.46} & \lambda \geq 0.2 \end{cases}$, where h is height of solid element, λ is total roughness density) to calculate the surface roughness length relevant to small-scale dust generation processes (Foroutan et al., 2017). The u_* in CAMx v6.50 is calculated according to the equation described in Louis (1979) expressed as $u_*^2 = \frac{\kappa^2 u^2}{\ln(z/z_0)^2} F_m(z/z_0, Ri_B)$, where F_m is a term as the function of Richardson number Ri_B. In addition, it also limits the minimum maximum value of the friction velocity to 0.4 m s^{-1}.

All the equations for u_* are presented in the flow charts of supplement file and the U_{10}, as well as u_* variations in each model during the dust episode are showed in Fig. 1 below (Fig. S11 in Supplementary Information). It reports that the u_* variations of WRF-Chem and CHIMERE are quite similar with averaged value of 0.60 m s^{-1}. In comparison, the u_* from CMAQ presents much lower with mean value of 0.41 m s^{-1}. It means that the introduction of a dynamic roughness length term in CMAQ results in lower friction velocity. This could be one of the reasons of the underestimation in CMAQ. As to the u_* in CAMx, it is the lowest among the values with mean value of 0.34 m s^{-1}because of the maximum limitation of 0.4 m s^{-1}, so that CAMx performed no dust emissions as the u_* was lower than u_{*t} during the episode.

The discussion about influence of different u_* is also presented in the Section of 3.6 in the revised manuscript.

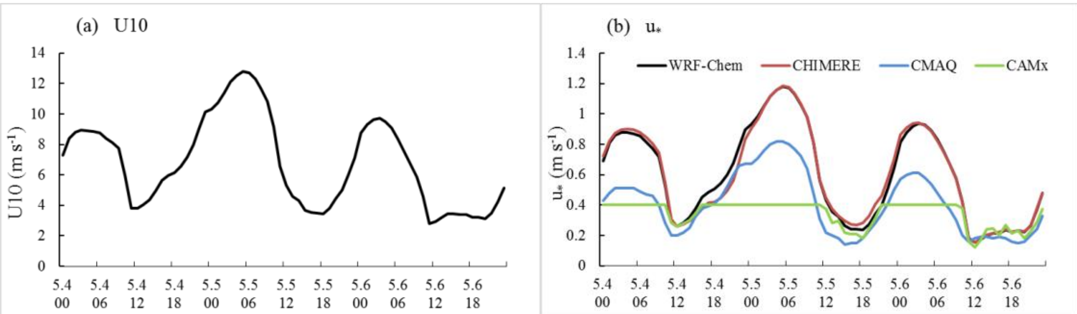

Figure 1. Time series of U10 (a), and u_* (b) from WRF-Chem, CHIMERE, CMAQ and CAMx in Changchun City during the dust episode.

References

Cakmur, R. V., R. L. Miller, and O. Torres: Incorporating the effect of small-scale circulations upon dust emission in an atmospheric general circulation model, J. Geophys. Res., 109, D07201, doi:10.1029/2003JD004067, 2004.

Foroutan, H., Young, J., Napelenok, S., Ran, L., Appel, K. W., Gilliam, R. C. and Pleim, J. E.: Development and evaluation of a physics‐based windblown dust emission scheme implemented in the CMAQ modeling system, J. Adv. Model. Earth Syst., 9(1), 585‐608, doi:10.1002/2016MS000823, 2017.

Louis, J. F.: A parametric model of vertical eddy fluxes in the atmosphere. Boundary-Layer Meteorology, 17(2), 187-202, doi: 10.1007/bf00117978, 1979.

Pryor, S., J. Schoof, and R. Barthelmie: Empirical downscaling of wind speed probability distributions, J. Geophys. Res., 110, D19109, doi: 10.1029/2005JD005899, 2005.

3. P13L4-7: The authors did not include the CAMx dust simulation in their in-depth analyses because the dust mask field required by the CAMx dust emission scheme did not include an erodible area in their region of interest. I'm confused by this reasoning. The dust mask and the dust source maps discussed for the other schemes in WRF-Chem (Figure 3) essentially serve the same purpose. Why test out different dust source fields in WRF-Chem but not the CAMx model? Claiming the paper includes an assessment of the CAMx model seems misleading to me. Recommend the authors either test the CAMx dust emission scheme with alternate dust source treatments similar to the exercise done for WRF-Chem, or remove the CAMx model and its discussion from the manuscript entirely.

Response: Thank you very much for your helpful suggestion. This question was also mentioned by reviewer #2. After discussion with all authors, we had implemented the CAMx model for further simulations. For the implementation of the dust emission scheme in CAMx, we select the seasonal dust source map (G12_0.1_seasonal in the manuscript) to replace the original dust mask file as this dust source map had the best performance among those source maps in the WRF-Chem model. The values in source map file were changed to 1 when the erodible fraction > 0 to fit the format of the dust mask file. Similarly, the performance of the CAMx dust simulation during the dust episode is analyzed and evaluated, and the result is showed in Section 3.5 of the manuscript. The daily averaged PM_{10} distribution on May 5th, 2015 is presented in Fig. 2b. It shows that the daily PM_{10} concentration simulated by CAMx ranged from 0 to 30 μg m^{-3} with high value area in the southwest part of the simulated domain, and there was no dust emitting from any erodible area in NEC. A control simulation without dust emission was also conducted and the PM_{10} pattern is same with Fig. 2b. It means that no dust emission at all and CAMx model failed to reproduce this dust episode occurred in NEC.

Considering the dust mask had been changed and the erodible areas were included in model, the failed simulation of CAMx might result from the lower value of friction velocity. In the dust model of CAMx, the friction velocity is limited to a maximum value of 0.4 m s^{-1}, making it keep a low level comparing to the values of other models (Fig. 1). It was difficult to exceed u_{*t} which was generally larger than 0.4 m s^{-1} (Fig. 5), so no dust emission occurred. Therefore, this limitation value was subsequently removed and the simulation was conducted again. The distribution of simulated PM_{10} without the u_* limitation was presented in Fig. 9c. It shows that the dust was mainly from western Jilin Province near the Songnen sandy land and transported westward. This pattern could be also observed from ground observations (Fig. 2e in manuscript). However, there was no dust emitting from Horqin sandy land. Simulated PM_{10} concentrations were generally lower than the observations with about 120 μg m^{-3} in source areas and 10~50 μg m^{-3} in the transported areas. Compared with the simulation with u_* limitation, this result was obviously improved which indicated that the limitation value of u_* in CAMx needs further adjustment to improve its performance over the areas other than barren and sparsely vegetated area.

The analysis and discussion on the CAMx result were in Section 3.5 and 3.6 of the manuscript.

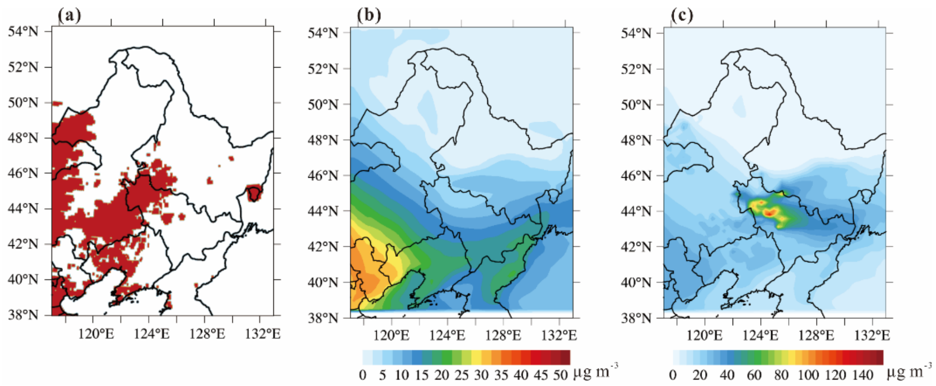

Figure 2. The substituted dusk mask (a) and daily mean PM_{10} distributions in NEC on May 5th, 2015 (b) using CAMx model.

4. P15L20-28: The strong dust emission magnitude from UoC and AFWA compared to GOCART in this study is somewhat unexpected given the findings discussed in the LeGrand et al. (2019) paper cited here. I don't think there's enough evidence to associate the excessive flux from the AFWA scheme with the saltation bin settings. I suspect these results may actually be related to the authors' use of the Pleim-Xiu (PX) land surface model (LSM) and Pleim (ACM2) planetary boundary layer (PBL) scheme. The U/u_* conversion in the PX/ACM2 setting typically produces stronger u_* values than NOAH LSM/PBL combos for equivalent U values. Operational agencies that use the AFWA dust emission scheme with the PX LSM frequently make use of the ustune tuning factor in the WRF-Chem configuration file to tone down u_* values ingested by the scheme for this very reason. It would be interesting to see a time series plot of model estimated u_* added to the time series plot in the appendix. If there is a strong sensitivity of dust emission scheme performance to LSM choice, it would be worth highlighting. Most other dust emission scheme assessment papers use the RUC or NOAH LSM.

Response: Thank you very much for your helpful comment. When we were preparing the meteorological input files, we firstly evaluated the simulation performances of WRF surface windspeed between the LSM of PX and Noah. The result showed they had same correlation coefficient (0.8) and close RMSEs (1.52 m/s for Noah-MP scheme and 1.61 m/s for Pleim-Xiu scheme). These comparisons showed close results between two schemes, however, the errors of Noah scheme had larger standard deviation showing higher dispersion than PX scheme. Considering this and the previous study by Zhang et al. (2015) which showed a good performance of PX scheme in the same research area, we finally choose PX scheme.

Later, according to this comment, we further conducted the dust emission simulation with the LSM of Noah and find that the simulated dust concentrations are much lower than those derived from PX scheme (Fig. 3). We further compared u_*, U10 and surface soil moisture calculated via PX and Noah scheme and the temporal variations of them are provided in Fig. 3. It shows that the variation curves of u_* calculated by PX and Noah scheme are quite similar. It does not present a stronger U-u_* conversion in the PX/ACM2 setting than in Noah scheme over the research area at this time. By contrast, the Noah surface soil moisture shows larger difference, with values 93.6% higher in Changchun City and 29.6% higher in the NEC area (Fig. 4 and Table 1). Moreover, the soil moisture curve with two LSM schemes are quite different. These discrepancies may result in the differences of estimated dust emissions. The lower soil moisture (which makes smaller threshold friction velocity) simulated by using PX scheme could be the reason of the stronger dust emission magnitude from PX compared to Noah LSM scheme. However, the dust emission simulated by AFWA scheme is considerably higher than that by GOCART no matter which LSM is used. Therefore, we think the differences of wind velocity and soil moisture between PX and Noah scheme could lead to the dust emission discrepancies, but it is hard to explain the differences of dust emission magnitude between

GOCART and AFWA scheme.

Many researches indicate higher dust emission simulated by GOCART than that of AFWA scheme over the areas like Mediterranean, West Asia and southwest Asia (Flaounas et al., 2017; Nabavi et al., 2017; LeGrand et al., 2019) which are quite different from our study. So we try to find out the reasons for large discrepancies between different dust schemes under the same meteorological condition. The saltation bin configuration which influences the dust mass distributions described in LeGrand et al. (2019) maybe have impact on the dust emission and concentration. One another explanation of the over-prediction of the AFWA dust concentration is that AFWA scheme considers vertical dust flux only related to the clay content which results in a higher vertical-to-horizontal dust flux ratio (Kang et al., 2010; Rizza et al., 2016; Rizza et al., 2017). We have added this into Section 3.2 of our new revised manuscript. Meanwhile, we also add the discussion about the LSM influence on the dust simulation in the same part.

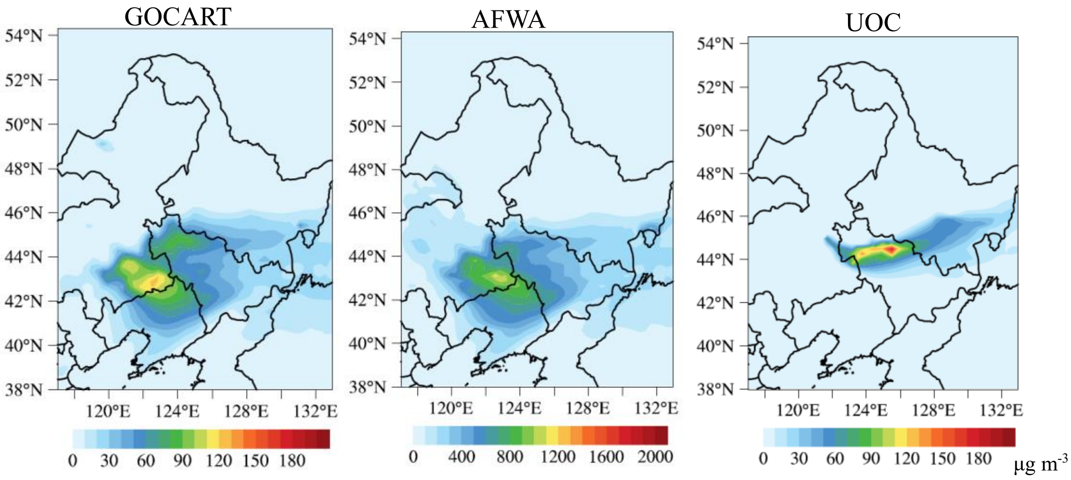

Figure 3. Daily mean PM_{10} distributions in NEC on May 5th, 2015 using GOCART, AFWA and UOC_Shao2004 with LSM of Noah.

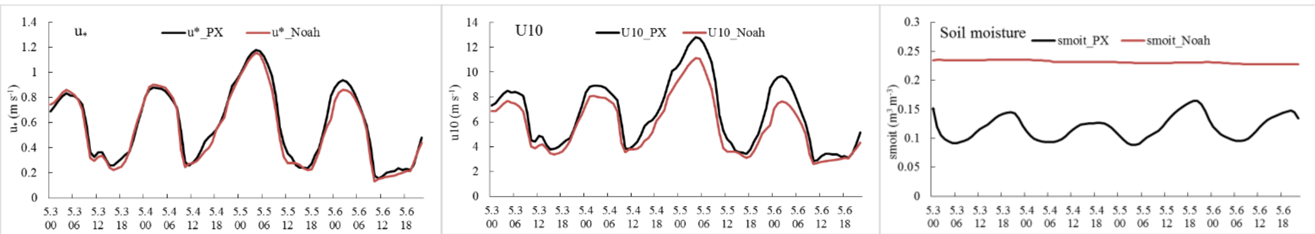

Figure 4. Time series of u_*, U10 and surface soil moisture simulated via LSM of PX and Noah in Changchun City during the dust episode.

Table 1. Mean u_*, U10 and surface soil moisture simulated via LSM of PX and Noah in the research area of NEC

	PX	NOAH
U10 (m s^{-1})	5.10	4.66
u_* (m s^{-1})	0.51	0.46
soil moisture (m^3 m^{-3})	0.27	0.35

Reference

Flaounas, E., Kotroni, V., Lagouvardos, K., Klose, M., Flamant, C., and Giannaros, T. M.: Sensitivity of the WRF-Chem (V3.6.1) model to different dust emission parametrisation: assessment in the broader Mediterranean region, Geosci. Model

Dev., 10, 2925-2945, https://doi.org/10.5194/gmd-10-2925-2017, 2017.
Kang, J., Yoon, S., Shao, Y. and Kim, S.: Comparison of vertical dust flux by implementing three dust emission schemes in WRF/Chem, J. Geophys. Res. Atmos., 116(D9), doi:10.1029/2010JD014649, 2011.
LeGrand, S. L., Polashenski, C., Letcher, T. W., Creighton, G. A., Peckham, S. E., and Cetola, J. D.: The AFWA dust emission scheme for the GOCART aerosol model in WRF-Chem v3.8.1, Geosci. Model Dev., 12, 131-166, https://doi.org/10.5194/gmd-12-131-2019, 2019.
Nabavi, S. O., Haimberger, L. and Samimi, C.: Sensitivity of WRF-chem predictions to dust source function specification in West Asia, Aeolian Res., 24, 115–131, doi:10.1016/j.aeolia.2016.12.005, 2017.
Rizza, U., Barnaba, F., Miglietta, M. M., Mangia, C., Di Liberto, L., Dionisi, D., Costabile, F., Grasso, F. and Gobbi, G. P.: WRF-Chem model simulations of a dust outbreak over the central Mediterranean and comparison with multi-sensor desert dust observations, Atmos. Chem. Phys., 17(1), 93, doi:10.5194/acp-17-93-2017, 2017.
Rizza, U., Anabor, V., Mangia, C., Miglietta, M. M., Degrazia, G. A., and Passerini, G.: WRF-Chem Simulation of a saharan dust outbreak over the mediterranean regions, Vol. 38, Special Edition, 330–336, Ciência e Natura, doi:10.5902/2179460X20249, 2016.
Zhang, X., Zhou, Q., Chen, W., Wang, Y. and Tong, D. Q.: Observation and modeling of black soil wind-blown erosion from cropland in Northeastern China, Aeolian Res., 19, 153–162, doi:10.1016/j.aeolia.2015.07.009, 2015.

5. The authors attribute over/under prediction of simulated dust conditions to dust emission scheme setting, but these conclusions are primarily based on comparison to daily average PM_{10} distributions. Simulated PM_{10} errors could also be due to issues with the atmospheric conditions (e.g., vertical mixing) and/or deposition/removal treatment. The validation methodology used for this study shows daily PM_{10} estimates are sensitive to the dust emission scheme configuration but does not provide enough evidence to confirm causality. This is especially important to note here given that multiple model frameworks are being used for this analysis.

Response: Thanks for your valuable comment. In this study, we focus on evaluating the performances of different dust models. Compared their spatial-temporal distributions and calculated the correlations, biases and errors with observational data. For this purpose, hourly ground-based monitoring PM data were used. In Section 3.2~3.5, the distributions of daily mean PM_{10} concentration were presented, in order to provide overall simulation situations (such as dust coverage and concentration level) of four air quality models. Preliminarily evaluated the dust performance of each model. Then the inter-model comparisons were conducted basing on the hourly data. We conclude the relative advantage and accuracy of dust performances of the air quality models. The error analysis and simulation assessment are both focus on the dust emission models. And indeed, the influence of atmospheric conditions and wet/dry deposition on the dust simulation are not conducted in this study, and we have mentioned this in the last paragraph of Section 3.6 in the manuscript. Thank you again for the helpful suggestion and the study on the impacts of atmospheric condition and deposition will be conducted in the future.

6. Section 3.5: I don't understand the rational for scaling PM_{10} concentrations in the inter-model comparisons. Why scale the simulation output rather than the scaling the emission fluxes?

Response: Thanks again. The PM_{10} concentrations simulated by different models varies from 10^0 to 10^4 ug/m^3 which means considerable discrepancies. The bias and error between simulated and observed PM_{10} concentration differ widely as well. It is very hard to plot all of the simulated concentrations or validation results in one figure and inconvenience to conduct the evaluation on all of the simulations at the same time. Therefore, we use scaling factor to adjust the outputs of WRF-Chem and CMAQ (as well as CAMx) to make them have similar concentration level with others. The scaling factor could help us to better understand the advantages and disadvantages of the model performance on dust simulation, rather than improve the accuracy of dust modeling at this time.

The evaluation in this study indicates tuning factor for dust flux is needed to improve model performance and puts forward an improving direction of dust simulation in present air quality models. Brief discussion and explanation about this are added in the parts of comparison and summary. The further work on tuning and localization of the dust models will be conducted basing on this study.

7. P26L13-14: The authors claim different algorithms for threshold friction velocity (FVT) resulted in significant differences in the simulated dust concentration and spatial distribution. This finding hasn't been demonstrated in this paper. The FVT treatments associated with each model haven't been introduced (again, need for model algorithm summary to guide discussion/conclusions). Recommend adding a figure of panel plots during the peak emission period showing simulated FVT estimates for a given grain size for each dust emission scheme - or - panel plots showing u*-FVT (U-FVT in the case of GOCART).

Response: Thank you very much for your helpful suggestion. We have calculated the FVT in each dust emission model according to the algorithms described in relevant references and source codes in models. The FVT variation used in each dust emission model during the dust episode is showed in Fig. 5. We have added this into Supplementary Information and also have discussion about its influence to dust emission in Section 3.6 of the manuscript.

From the figure we find that the overestimation of WRF-Chem AFWA could in part be explained by the lower FVT comparing to those used in other dust schemes. The variation of UOC FVT fluctuated widely and presented highest FVT peaks, but during the dust episode (5 May, 2015) it kept lower values making it in the middle level of FVT. This may be part of the reason of the overestimated PM_{10} concentration. However, the FVT of GOCART has the lowest value while its simulated dust emission is in the lower level. As we know, the FVT is very important in the dust emission calculation, however, it needs further adjustment and improvement as the air quality at present have difficulty in calculating FVT properly in some areas such as northeastern China. And this is also our next step work to implement field study and measurement of FVT, adjust and localize the FVT algorithm and the relevant parameters.

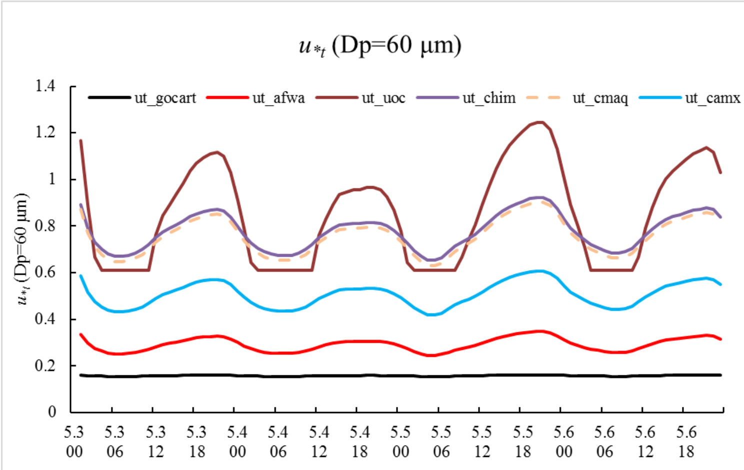

Figure 5. Time series of u_{*t} in from dust emission model of GOCART WRF-Chem, AFWA WRF-Chem, UOC WRF-Chem, CHIMERE, CMAQ and CAMx in Changchun City during the dust episode.

Minor comments:

1. P4L4-6: I would not qualify this paper as the first comprehensive evaluation of dust models for East Asia. A single dust event case study is good for examination and discussion of how the dust models function under a given forcing condition, but an extended study period with several events would be needed to truly assess model performance.

Response: Thank you for the helpful comment. This study is not the first comprehensive evaluation of dust models for East Asia, but we are sure it is the first evaluation of the dust modules in air quality models (rather than climate model (Uno et al., 2006), global model et al.). In order to keep in rigorous, this sentence is now revised as "Here we present a comprehensive evaluation of multi-model simulations of windblown dust emissions in air quality models during a dust episode in East Asia…" Now we are working on the extended evaluations of the dust emissions in air quality models, such as the evaluation on the performance of CHIMERE during an autumnal dust episode (Ma et al., 2019).

Reference

Ma, S., Zhang, X., Gao, C., Tong Q., Xiu, A., Zhao, H and Zhang, S. Simulating Performance of CHIMERE on a Late Autumnal Dust Storm over Northern China[J]. Sustainability, 11(4). doi: 10.3390/su11041074, 2019.

Uno, I., Wang, Z., M. Chiba, Y. S. Chun, Sun Ling Gong, Yukari Hara, E. Jung et al. "Dust model intercomparison (DMIP) study over Asia: Overview." Journal of Geophysical Research: Atmospheres 111, no. D12 (2006).

2. P8L10: The AFWA scheme is adapted from the dust emission scheme originally described by Marticorena and Bergametti (1995), not GOCART. It would be appropriate to cite the LeGrand et al. (2019) paper here.

Response: It had been revised as "The AFWA dust scheme is a modified version of Marticorena and Bergametti (1995) dust scheme developed by the Air Force Weather Agency (LeGrand et al., 2019)."

3. P8L19: The UoC coding error was not corrected in the public code distribution until the release of WRF-Chem v4.0. It is unclear here whether or not the authors manually corrected the coding error in their compilation of WRF-Chem v3.9.1. This was also mentioned by another community member on the forum.

Response: The error in UOC source code has been corrected before we conduct the simulations in this study. But it seems that we didn't make it clearly expressed. After revising, it is now expressed as "Note that the last term in the saltation flux formula in UOC source code is expressed as $(1+(\frac{u_{*t}}{u_*})^2)$ by error in WRF-Chem before the version of 4.0. In this study, it has been changed into $(1+\frac{u_{*t}}{u_*})^2$ in WRF-Chem version 3.9.1 according to the description in Shao et al. (2011)."

4. P12L1: The G01 acronym hasn't been defined yet.

Response: The explanation of the way we name the dust source map is provided into manuscript and it is described as "In this study, we named this source map as G01_0.25 according to first author and published year of relevant literature and its spatial resolution.".

5. P12: The dust source map discussion is difficult to follow. Table 2 provides a good summary, but the labels (e.g., G01, K08, etc.) need to be introduced in the text.

Response: The basic information and name of each source map are now introduced in the manuscript and showed as "The source map with resolution of 1°×1° developed by Ginoux et al. (2001) (namely G01_1.0) basing on the same method with G01_0.25 was obtained from the homepage of Dr. Paul Ginoux (https://www.gfdl.noaa.gov/pag-homepage). Except G01_0.25 and G01_1.0, the other source maps were obtained using satellite observations. The map with resolution of 0.25°×0.25° provided by Koven and Fung (2008) (K08_0.25 hereafter)", "The source map with only natural origins was named G12_0.1_natural while that with both natural and anthropogenic origins was named G12_0.1_ant+nat." and "The dust source map obtained from the NASA-Unified WRF (NU-WRF) version 7 (Kim et al., 2014) was similar to the former source map but divided into four seasons, therefore it was named as G12_0.1_seasonal in this study".

6. P12L20-21: The authors mention they conducted a ground survey to assess representativeness of the erodibility field. What method was used for the ground survey? Is this something that was done subjectively or with in situ measurements (e.g., wind tunnel or PI-SWERL device)?

Response: Actually, it is in-situ measurements but not the ground survey. In spring of 2013, we conducted a field in-situ observation for wind erosions over the cropland experimental station (44° 12′ N, 125° 33′ E) with black soil in Northeastern China. The SENSIT sensor (Model H11-LIN, Sensit, Co.) and wind velocity sensors (5 heights, 0.2m, 0.5m, 1.0m, 2.0m and 5.0m, model 010C, Met One Instruments Inc.)were adapted for observing the threshold friction velocity and the comment value is 0.37m/s at the surface ground.

Reference

Zhang, X., Zhou, Q., Chen, W., Wang, Y. and Tong, D. Q.: Observation and modeling of black soil wind-blown erosion from cropland in Northeastern China, Aeolian Research, 19, 153-162, doi:10.1016/j.aeolia.2015.07.009, 2015.

7. P12L20: Please list the NU-WRF version number used to acquire these fields.

Response: Now it is described as "The dust source map obtained from the NASA-Unified WRF (NU-WRF) version 7".

8. Figure 3: Was the G01 1-deg field generated manually for this analysis or acquired from somewhere? The 1-deg field was

replaced by the .25-deg field in the WRF-Chem repository back in 2012 and is no longer part of the standard WRF-Chem static dataset download. If it was generated manually, what process was used set the vegetation mask?

Response: The file of source map G01_1.0 with format of NetCDF was obtained from the homepage of Dr. Paul Ginoux (https://www.gfdl.noaa.gov/pag-homepage). And converted to ASCII format (.bin) by using write_geogrid.c in WPS. We find that the erodible fraction distribution of it shows more reasonable than the default one in WRF-Chem v3.9.1, therefore, it is introduced and the evaluation of its performance is conducted in this study. The data source is now given in the manuscript.

9. P14L1: Please describe the BELD3 dataset.

Response: The BELD3 dataset is explained in the manuscript as "when the Biogenic Emissions Landcover Database version 3 (BELD3) dataset was used during regional simulation in the USA."

10. P14L3: USGS and MODIS land use datasets are brought up here with no context. Given how often land use datasets are brought up in the discussion section, the authors may want to consider listing which land use dataset was used to configure each of the model frameworks. Please also list the number of classes associated a particular dataset, since there are multiple versions of both the USGS and MODIS IGBP land use datasets.

Response: The USGS and MODIS land use datasets in this study are the default land use and soil category datasets which provided as part of the WPS static data file. They were obtained from http://www2.mmm.ucar.edu/wrf/src/wps_files/. These default data were remapped via the geogrid program of WPS. We find that the content in P14L3 could not be clearly expressed, so it is revised to "However, the grassland fraction was not taken into account in when using the USGS or MODIS land use dataset according to the source code of CMAQ, which may lead to an underestimation of dust emission." The selections of land use data are described in Section 2.5 and detailed information provided in the Supplementary Information.

11. Sections 2.4 and 2.5 in general: The overall model descriptions are somewhat vague. Additional model configuration information would be needed if others wished to replicate this study or the authors' methodology. Given the number of model frameworks used and the current paper length, it may be better to include specific model configuration setting information in a supplement. This was also noted in the review forum by GMD editor David Ham.

Response: Thanks for your kindly suggestion. According to the advice of Editor David Ham. We had upload all the model source codes that we used, the namelist files, the sources of the input data (such as the meteorological reanalysis data and erodible fraction data), as well as pre- and post- processing script we used to the Zenodo website, and we had added the link of https://doi.org/10.5281/zenodo.3376774 in our new revised manuscript.

12. P15L9: "Control" is too strong of a statement here. Suggest changing to distribution and intensity of modeled dust are sensitive to...

Response: This has been revised to "As mentioned above, we found that the distribution and intensity of modeled dust aerosols were sensitive to the dust source maps in use."

13. P16L8-10: This seems like an odd choice to me. Figures 5k&5L and 5Q&5R suggest these two treatments produce markedly different results.

Response: Although the dust source map of G12_0.1_ant+nat and G12_0.1_seasonal (Fig. 5k&5L and 5Q&5R in manuscript) were obtained via the same methodology, G12_0.1_seasonal is divided into four seasons which makes it able to present the temporal variation of the erodible fraction. It is more reasonable than the constant field (G12_0.1_ant+nat) and makes obvious different between the distribution of G12_0.1_ant+nat and G12_0.1_seasonal (Fig. 3e&3f). Therefore, G12_0.1_seasonal is chosen for the next step in the evaluation and the reason is explained in the manuscript as "Considering that the source maps of G12_0.1_ant+nat and G12_0.1_seasonal were obtained via the same methodology but the latter one provides seasonal divisions making it more reasonable and closer to the actual environment, G12_0.1_seasonal was chosen for the next step in the evaluation."

14. P20L16-19: Dust concentration could be due to other factors outside of dust emissions (e.g., dispersion, mixing, deposition treatment, etc.). The causality statement here is too strong.

Response: This sentence is revised to "When considering that the dominant soil textures were loam and clay loam in NEC, this explained the reason for the underestimation occurred in CMAQ compared with WRF-Chem and CHIMERE."

15. P20L16-21: This discussion needs equations. See previous comment about emission scheme flow charts.

Response: The equations for calculating α (sandblasting efficiency) are provided in the flow chart of Supplementary Information and we have explained about this in the manuscript.

16. P24L1-11: Were these values tuned as well?

Response: Yes, the values in this part (discussion of Fig. 12 in manuscript) are also tuned. The comparison and analysis in Section 3.6 are both based on the tuned outputs of the air quality model. The explanation about this is implemented in the beginning of Section 3.6

17. P25L7-8: These equations need to be provided. Is this error unique to this version of CMAQ? How is soil moisture integrated into the calculation of friction velocity threshold?

Response: This part focuses on the discussion of CHIMERE output. The soil moisture is used to calculate of soil moisture correction, and then multiplies the dry/smooth threshold friction velocity to obtain the actual threshold friction velocity. The equation of soil moisture correction is showed as

$$H_w(w) = \begin{cases} \sqrt{1 + 1.21(w_g - w_g')^{0.68}} & w \geq w' \\ 1 & w < w' \end{cases}$$

where w_g is gravimetric soil moisture, w_g' is moisture without effect on capillary forces. The unit conversion from volumetric soil moisture w_v (m^3 m^{-3}) to gravimetric soil moisture w_g (kg kg^{-1}) are different in different models. Such as $w_g = \frac{w_v \rho_w}{(2.65-0.15clay)(1-porosity)}$ in WRF-Chem v3.9.1, and $w_g = \frac{w_v \rho_w}{(2650\times(0.511+0.126\times(soiltxt1+soiltxt2)}$ in CMAQ v5.2.1. However, the conversion $w_g = 100 \times w_v$ in CHIMERE v2017r4 is obviously unreasonable. Later, we also find this error exist in the previous version of CHIMERE (such as CHIMERE v2016a1). The equations are provided in the flow charts in Supplementary Information.

18. P26L6: Authors evaluated dust models, not dust emission schemes. There are too many free variables to isolate result outcomes to the dust emission schemes. Use of dust models here instead of dust emission schemes would make the language

here consistent with the intro section.

Response: It is revised to "In this study, we quantitatively evaluated the performance of several physically-based dust emission models in…."

19. P26L7: "four newly-introduced dust source maps in WRF-Chem" is a bit of an overstatement. The NU-WRF model some of these maps were obtained from is the NASA implementation of WRF-Chem.

Response: Thanks for your valuable comment. We find that these statements are inaccurate here, therefour we revise it as "Four dust schemes and four additionally-introduced dust source maps in WRF-Chem v3.9.1…"

20. P26L14-15: I agree with the authors that uncertainty associated with the U to u_* conversion is likely an important source of error, potentially more so than minor differences in dust emission physics, but there are no discussions, figures or values presented in this paper supporting this statement.

Response: Thank you very much for your helpful comment. We found that we omitted the analyses about u_* in the discussion part. On the basis of the results in the second major comment, we compared the u_* in different model in P28L22 and P29L14 when analyzing the results of CHIMERE and CMAQ. It shows as "Moreover, different algorithms for u_{*t} resulted in significant differences in the simulated dust emission. For instance, the variations of u_* in CHIMERE and WRF-Chem are similar (Fig. S11a, b) during the dust episode, however, u_{*t} presented large discrepancies from one model to another (Fig. S12). The value of u_{*t} in CHIMERE is in the upper level of the six kinds of u_{*t} used in the dust emission models while that of AFWA is much lower. This could also be one of the reasons of the overestimation in WRF-Chem AFWA." and "In addition, using the dynamic roughness length term when calculating u_* in CMAQ led to lower u_* (with mean value of 0.39 m s^{-1}, Fig. S11c) comparing to those in WRF-Chem and CHIMERE (0.58 m s^{-1}). This would be another reason for its underestimated results.", respectively. The u_* algorithm in CAMx was discussed in Section 3.5 of the manuscript (and in major comment 4). The figure of u_* variations in each air quality model is provided in Supplementary Information (Fig. S11).

21. P26L24-25: Suggest changing to "All simulations performed best near the dust source areas and degraded in accuracy with downstream advection." This may indicate issues with transport, deposition, or forcing conditions if this feature is consistent across all model configurations. Were the WRF 3.9.1 meteorological fields assessed to ensure they captured the general storm evolution prior to being applied to the various dust models?

Response: Thanks for your helpful suggestion. It is revised to "All simulations performed best near the dust source areas and degraded in accuracy with downstream advection." The meteorological fields of WRF v3.9.1 has not been test in this research area before, but we have evaluation the surface wind simulated by it in this study and it presented acceptable result (Section 2 in Supplementary Information).

22. Typos: P17L15: Typo, might?

Response: It has revised as "This difference might have arisen because the KOK scheme was mainly built on fragmentation theory"

In addition, besides the major and minor questions as mentioned above, we conducted reproducibility tests of our simulations in this study, and found an error in the WRF-Chem section. We have run the WRF-Chem model with UOC_Shao2011 scheme and it showed that PM_{10} simulated by UOC_Shao2011 presented little similarity with the observations (Fig. 6). Considering its unreasonable results, then we selected UOC_Shao2004 for the subsequently simulations. It means that the used dust scheme in original manuscript is UOC_Shao2004. However, the texts expressed the used scheme as UOC_Shao2011 by wrong. Therefore, we corrected this into UOC_Shao2004 in the manuscript.

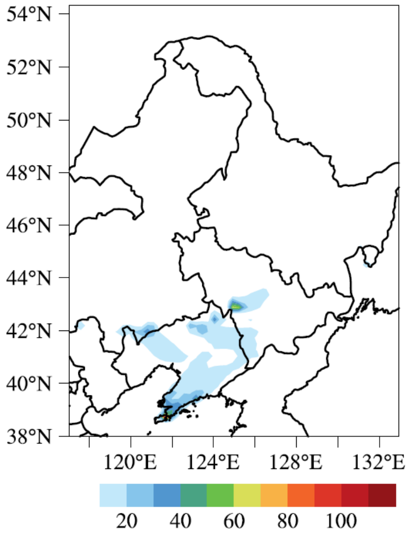

Figure 6. Daily mean PM_{10} distributions in NEC on May 5th, 2015 using WRF-Chem UOC_Shao2011.

We have tried our best to improve the manuscript and made changes in the manuscript. These changes will not influence the content and framework of the paper. And here we do not list the changes but marked in the revised manuscript. We thank the reviewer again for the constructive advice that have helped us to improve our manuscript.

All in a word, via these evaluation works, we hope to do some contributions to the community for enhance the dust forecast ability on regional scale in air quality models.