Updated reply to the comments of the 1st Referee:

Comment1: This manuscript examined some inconsistencies with the use of the Goddard Chemistry Aerosol Radiation and Transport (GOCART) aerosol module in the fully coupled WRF-Chem model. The authors identified that 1) the diagnostic output of PM2.5 surface concentration was underestimated by 7% and PM10 surface concentration was overestimated by 5% due to the incorrect representation of the dust and sea salt coefficients; 2) the contribution of sub-micron (0.1 – 0.46 ï ̧A m) dust particles was underestimated in the calculation of optical properties with the consequence of underestimated AOD by 25-30% because the finer dust particles were not accounted for in the Mie calculations; and 3) an inconsistency in dealing with gravitational settling that led to the overestimation of the dust column loadings by 4-6%, PM10 surface concentrations by 2-4%, and the rate of gravitational settling by 5-10%. The authors further examined the impacts of boundary conditions on PM10 surface concentrations using the MERRA-2 reanalysis. These are all useful aspects of the WRF-Chem model and certainly help the improvement of the WRF-Chem simulations. However, this manuscript lacks indepth technical and scientific analyses and is rather poor scientifically. All the analyses were based on one dust case (1-12 August, 2016) over the Middle East which calls into question the applicability and effectiveness of the code rectifications in other regions and in other dust cases under different meteorological and land surface conditions. Besides, I have several major concerns as listed below:

We thank the reviewer for the valuable comments. We do not completely agree with the statement related to the technical and scientific merits of our study. The paper discusses the inconsistencies we found in the WRF-Chem v3.2 code released on April 2, 2010. We cooperated with the model developers to test and implement those corrections in the newly released WRF-Chem v4.1.3. The main objective of the presented paper is to quantify the effect of those inconsistencies on model performance. These changes could not be made without a deep understanding of model physics and the code.

Our first goal in this work was to show the effect of physical and coding errors we identified and corrected in WRF-Chem. We believe it is essential to clean the code and delineate the contributions of the introduced changes. It demonstrates the physical interconnections and evaluates the results' sensitivity to variation of specific processes and parameters. Therefore, we have chosen the idealized settings to quantify the contributions of multiple model corrections presented in the paper.

Specifically, our findings explain why WRF-Chem overestimated PM10 surface concentrations, and why realistic values of AOD were associated with overly strong dust emissions and elevated dust surface concentrations. These discrepancies have been discussed in some previous studies: Kumar et al. (2014); Eltahan et al. (2018); Flaounas et al. (2017).

Our numerical experiments are chosen to demonstrate the effectiveness of those corrections quantitatively. It is not vitally important which region of the world is selected for our experiments, just that it should be a dusty region. We use a WRF-Chem experimental setup configured over the Middle East, one of the most significant dust source regions.

However, we agree that the more realistic case-study supported by observations will undoubtedly benefit the paper. Therefore we have conducted two seven-month-long WRF-Chem simulations covering the period from June 1 to December 31, 2016 (see section 3.4 in the revised manuscript). In the first simulation, all inconsistencies in WRF-Chem are fixed (ALL_OK WRF-Chem run). In the second, none of the discrepancies are fixed (ALL_OLD WRF-Chem run). In both runs, dust emission was calibrated so that the AOD from the run fits the AERONET AODs observed at three stations (KAUST campus, Mezaira, and Sede Boker). To test the output from ALL_OK and ALL_OLD runs, we use the available PM10 observations conducted by the Saudi Authority for Industrial Cities and Technology Zones (MODON) in Riyadh, Jeddah,

and Dammam (mega-cities of Saudi Arabia) during 2016 (see Fig. 8 in the revised manuscript). More details on the MODON observations are available in (Ukhov et al., 2020).

Results presented in Fig. 8 show that PM10 surface concentrations in ALL_OK run reproduce MODON observations much better than in the ALL_OLD run, where PM10 concentrations are severely overestimated. In particular, mean biases with respect to MODON observations for ALL_OK and ALL_OLD runs are 2, 23, 77 and 72, 128, 275 (ug/m^3) for Jeddah, Riyadh, and Dammam, correspondingly. Thus, the ALL_OLD run's PM10 bias is at least three times bigger than in the ALL_OK run.

The comparison of the dust column loadings averaged over the summer (June, July, August) of 2016 (see Fig. 9) shows that dust content in the atmosphere is overestimated by up to 80% in the ALL_OLD run compared with the ALL_OK run. These results are in agreement with the statement we made in the conclusion of the original manuscript.

Comment 1: The Introduction section was poorly written. It is clear that the authors have read and cited a lot of references on the subject of dust sources, dust impacts, and dust modeling but the Introduction section was written in such a way that it was hard to gain a clear idea of why the inconsistencies occur and what the latest developments are in dealing with them and how the authors would like to address them. The Introduction section needs to be improved substantially.

We agree. We have shortened and improved the introduction section and clarified the text overall.

Comment 2: How were the "correct" dust and sea salt coefficients, d_25, s_25, d_10, in Equation 2, determined? The authors mentioned that they used the natural logarithm of particle radii but what was the rationale behind that? Was that determined from empirical relationships or lab experiments or field measurements or just trial and error? Are there any references for that?

Calculation of PM2.5 and PM10 concentrations requires the integration of aerosol volume size distribution (approximated in GOCART by five bins for dust and by three bins for sea salt) over the radius r from 0 to 2.5 um and from 0 to 10 um, respectively. Integration could be done assuming that the size distribution is a function of r or ln(r). Coefficients d_25, s_25, d_10 in eq.2 are obtained assuming that the aerosol size distribution is a function of ln(r). This method is acknowledged in CAMS reanalysis (https://confluence.ecmwf.int/pages/viewpage.action?pageId=153393481) and is justified by the fact that an aerosol size distribution is a smoother function of ln(r) than r, and therefore numerical integration is more accurate. The coefficients s_25, d_25 d_10 in the original WRF-Chem v3.2 were calculated incorrectly.

The formulas for calculating PM2.5 and PM10, as we mentioned, are presented in the CAMS knowledge base at:
https://confluence.ecmwf.int/pages/viewpage.action?pageId=153393481

We repeat these relations here for clarity:

$$PM2.5 = RHO * (1 * SS1 + 0.4 * SS2 + 1 * DD1 + 1 * DD2 + \mathbf{\underline{0.11 * DD3}} + 0.7 * OM1 + 0.7 * OM2 + O.7 * SU1 + 1 * BC1 + 1 * BC2)$$

$$PM10 = RHO * (1 * SS1 + 1 * SS2 + 1 * DD1 + 1 * DD2 + \mathbf{\underline{0.55 * DD3}} + 0.85 * OM1 + 0.85 * OM2 + 0.85 * SU1 + 1* BC1 + 1 * BC2)$$

(1)

The relations (1) show that PM2.5 and PM10 comprise the contributions from different bins that constitute the size distributions (concentration of specific aerosols within a given size range) for five aerosol types: sea salt (SS), dust (DD), organic matter (OM), sulfate (SU), and black carbon (BC). E.g., three dustbins (DD1,

DD2, DD3) cover the following particle diameter ranges (um): DD1: [0.06-1.1], DD2: [1.1-1.8], DD3: [1.8-40]. To calculate the contribution of, e.g., 3rd dustbin DD3 into PM2.5, one has to interpolate the portion of DD3 that falls into the range D<2.5 um. We checked that CAMS does this interpolation in the logarithmic space:

$$(ln(2.5) - ln(1.8))/(ln(40) - ln(1.8)) = 0.32/3.101 = 0.11$$

The contribution of 3rd dust bin DD3 into PM10 is calculated the following way:

$$(ln(10) - ln(1.8))/(ln(40) - ln(1.8)) = 1.714/3.101 = 0.55$$

These coefficients for DD3 contributions (and similarly for all other aerosol bins) are used in CAMS, see formula (1). In our paper, we adopted this approach for calculating coefficients *d_25, s_25*, and *d_10* by interpolating bin's contributions in PM2.5 and PM10 in the logarithmic space.

Comment 3&4: I commend the authors for identifying the underestimation of the AOD by the neglect of the sub-micron dust particles and their effects but I am concerned that the authors did not provide any logic behind the modifications of the corresponding numbers from MOZAIC bins (1, 2, 3, 4, 5, 6, 7, 8) to GOCART dust bins (DUST1, 2, 3, 4, 5). Please provide scientific evidence or references to support this work. Otherwise, what the authors have done is not convincing at all. 4) The authors changed the calculation of bin concentrations of dust and sea salt from using the functions of particle radius to using the functions of natural logarithm of radius. Again, what was the rationale behind this?

The mapping of GOCART five dustbins approximation of the aerosol size distribution to the MOSAIC eight bins requires the interpolation of aerosol size distribution over the radius r or ln(r). Consistent with our calculation of PM2.5 and PM10, we assume that the aerosol size distribution is a function of ln(r). The rationale is that the size distribution is a smoother function of ln(r) than r, and therefore both interpolation and integration are more accurate in the ln(r) space. In the course of interpolation, we conserve the integral of volume size distribution, i.e., the total volume of particles. Table 4 compares the bins' partitions calculated assuming aerosol size distribution is a function of ln(r) or r. Please see the response to Comment 2, where we described the procedure.

Comment 5: I did not understand how the inconsistency in the gravitational setting of dust and sea salt led to the increase of their total mass in the atmosphere. The authors mentioned that "Instead of transport the dust and sea salt mass between the layers, the default finite-difference scheme transport their mass mixing ratios not taking into account the dry air density variation with the height". Does this mean that dust and sea salt mass can't be transported across the layers? If there are vertical motions or turbulence dust and sea salt can certainly move up and down. Then where did this overestimation come from?

In the previous versions of WRF-Chem coupled with GOCART, the mass fluxes of gravitationally deposited material (dust and sea salt) were miscalculated. Let's consider two neighboring (in vertical) grid cells, one

above another. Originally, the model underestimated outgoing aerosol mass flux at the bottom of the upper grid cell and overestimated the incoming aerosol mass at the top of the lower cell. As a result, the integral mass balance was violated. In the paper, we present a new finite difference scheme that is free from the above discrepancy. It keeps the outgoing (from the upper cell) and incoming (into the lower cell) aerosol mass flux exactly equal. We conducted an idealized numerical experiment which shows that the old scheme violates the mass balance, whereas the new one does not.

6) The English of this manuscript needs to be improved.

We agree. We have now thoroughly re-edited and improved the text of the manuscript for clarity of meaning and readability.

Minor concerns:
1) On Line 21-22 of Page 8 the authors stated that "The model erroneously pushed more dust into the atmosphere to fit the observed AOD". Was this in reference to the model default values or the model runs that assimilate the observed AOD. I don't quite understand this.

Here we compare two runs with and without corrections. In both runs, emissions are corrected to fit the observed AOD. In the model without corrections, to reproduce the observed AOD, more dust has to be emitted because of erroneous AOD's calculation. Thus the run without corrections, e.g., overestimates the PM concentrations in the near-surface layer. This is in line with the cases discussed in the literature when dust emission in WRF-Chem is tuned to reproduce AOD and it overestimates PM surface concentrations. The text has corrected to clarify this issue.

2) I did not understand this statement on Line 16-18 of Page 10: "In both runs, the magnitude and temporal evolution of the AOD time-series are well correlated with the observed AERONET AOD at all sites only in the absence of dust events or when the AERONET AOD is below 1". I am wondering what the authors wanted to convey here with this statement.

We wanted to convey the fact that the WRF-Chem model with the original GOCART-WRF dust emission scheme (*dust_opt=1*) incorrectly captures strong dust events, which we defined as the cases when AOD > 1.

3) The ALL_OK run was just one of the model realizations. It is not appropriate to treat it as truth and designate the differences from it as biases.

We agree. We have corrected the text to make this point clear.

4) How useful is the Merra2BC utility in general sense? Since the MERRA-2 reanalysis is essentially a global atmospheric reanalysis generated by an atmospheric circulation model with the incorporation of trace gas constituents and aerosols. It is not a fully coupled atmosphere-chemistry model and I would think that the so created gaseous and aerosol species data may not be as useful as those from fully coupled chemistry models such as MOZART-4.

One of the advantages of the MERRA-2 reanalysis is the fact that it assimilates AOD and SO2 column loadings to constrain dust and SO2 content in the atmosphere. Additionally, MERRA-2 has a higher resolution than MOZART-4. MOZART-4 resolution is 128×64 grid boxes, while MERRA-2 has 576×360. We use MERRA-2 output specifically for that time period we consider in our simulations, while originally WRF-Chem uses MOZART output from a climatological run. MOZART-4 has 4 dust size bins with a maximum radius of 5 microns, while MERRA-2 has 5 size bins with a maximum radius of 20 microns. Besides dust, MERRA-2 calculates sea salt, ozone, sulfate, black and organic carbon, sulfur dioxide, DMS,

and MSA. MERRA-2 hourly fields are available from 1980 to present and are an invaluable asset for building initial and boundary conditions for regional aerosol and chemistry simulations. This is not to diminish the value of the MOZART-4's more sophisticated chemistry. But WRF-Chem uses chemical mechanisms and aerosol microphysics quite consistent with MERRA-2, which provides another advantage of using MERRA-2 for building IC&BC for WRF-Chem.

5) The authors mentioned that Figure 4 shows the averaged AOD time-series and scatter plots obtained from the ALL_OK and NON_LOG_046 runs. Instead, I found the spatial patterns of AOD and their differences.

Thank you for capturing the Figure 4 misrepresentation. We have corrected the sentence, which now reads: "*Figure 4 shows the averaged AOD fields obtained from the ALL_OK and NON_LOG_046 runs, as well as their relative difference (%).*"

Sincerely,

Alexander Ukhov and Georgiy Stenchikov

References

Ukhov, A., Mostamandi, S., da Silva, A., Flemming, J., Alshehri, Y., Shevchenko, I., and Stenchikov, G.: Assessment of natural and anthropogenic aerosol air pollution in the Middle East using MERRA-2, CAMS data assimilation products, and high-resolution WRF-Chem model simulations, Atmos. Chem. Phys., 20, 9281–9310, https://doi.org/10.5194/acp-20-9281-2020, 2020.