Updated reply to the comments of the 2nd Referee:

General comments

Comment 1: There are a very large number of appendices, some of them very short, which results in a disjointed manuscript that doesn't flow very well. I would consider reformulating these so that the overall paper flows better (possibly merging those which are fundamental to the paper, e.g. the non-technical description of Merra2BC, into the body). Appendix F doesn't even appear to be referenced anywhere in the manuscript.

We thank the reviewer for the valuable comments. We have merged Appendix H, which presents the improved numerical scheme for gravitational settling calculations, with the main body of the paper. Appendix F is combined with appendix G and is now referenced in Sect. 3.3. The appendices are specifically designed to consolidate the technical information about diagnostic calculations and pre-processing (Merra2BC). We feel it is convenient to have this information on hand at the end of the main text.

Comment 2:

There are several places (particularly in the introduction, but also in section 2.1.1) where an excessively long list of citations is given to exemplify a point – please consider whether all of these are necessary and if not cite the most pertinent examples.

The reference list in section 2.1.1 has been shortened.

Comment 3:

In a number of places, configuration parameters of WRF-Chem are referred to without explaining their meaning. While the manuscript is obviously of most interest to those familiar with this model, it should be understandable more widely.

We have added Table 1 into the Introduction section, where a short description of the *chem_opt* namelist options affected by our modifications is provided.

Comment 4:

Finally, the experiments carried out should be described prior to the results section.

We have added a description of each found inconsistency into the Introduction section (before Table 1). Although this gives an impression of how the test experiments are organized, we prefer to keep the detailed description of the experiments in the specific sections, as it is combined with the explanation of a specific problem and how it was rectified.

Specific comments:

p.3, lines 29–31: please include a table explaining what these options are.

We have added Table 1 at the end of the Introduction section, where all *chem_opt* options are explained.

p.4, line 14: please state what "chem_opt=300" means.

Table 1 now includes a description of the *chem_opt=300.* The text has also been modified to read: "*We use chem_opt=300 namelist option, which corresponds to simulation using GOCART aerosol module without chemical reactions.*"

p.4, line 24: is AFWA an acronym? If so, please expand on first use.

Acronym AFWA has been expanded in the text into "Air Force Weather Agency"

p.6, lines 28–29: should this be ERA-Interim (not ECMWF Interim)?

This sentence has been changed to read: "*IC&BC for meteorological fields are derived from the ERA-Interim (Dee et al., 2011) global atmospheric reanalysis produced by the European Centre for Medium-Range Weather Forecasts (ECMWF).*"

p.7, lines 9–12: a little more explanation of Merra2BC is warranted in the main body of the paper, especially given that its introduction is later highlighted in the conclusions suggesting it's more than a minor element.

This has been edited so that we now discuss the rationale of using MERRA-2 output for constructing the IC&BC. MERRA-2 aerosol and chemical species fields are superior (in terms of spatial resolution, time coverage, and because they are constrained by observations) in comparison to those used in WRF-Chem so far for calculation IC&BC (MOZART-4, for example). Merra2BC is a preprocessor that conveniently transforms MERRA-2 output into WRF-Chem IC&BC.

p.7, line 14: this is confusing, because it refers to "all inconsistencies" when these haven't yet been enumerated in the text. Please reformulate so that the inconsistencies, changes made, and experiments carried out are introduced prior to the results section.

Thank you for your comment. To clarify this issue, a new paragraph has been added into the Introduction section, where we briefly introduce the found inconsistencies and their effects.

p.7, line 30: is it documented that a log-based distribution is the one on which the parameterisation is based, and that therefore this is an inconsistency? Or is it the author's assumption/assertion that such a distribution is universally the correct one to assume, whatever the parameterisation?

Calculating PM2.5 and PM10 concentrations requires the integration of aerosol volume size distribution over the radius r from 0 to 2.5 um, and from 0 to 10 um, respectively. Integration can be performed assuming that the size distribution is a function of r or ln(r). We calculate coefficients d_25, s_25, d_10 in eq.2 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.

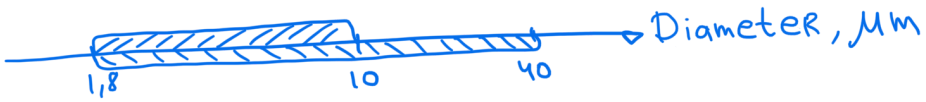
The formulas for calculating PM2.5 and PM10 presented in the CAMS knowledge base:

$$PM2.5 = RHO * (1 * SS1 + 0.4 * SS2 + 1 * DD1 + 1 * DD2 + \mathbf{0.11 * DD3} + 0.7 * OM1 + 0.7 * OM2 + O.7 * SU1 + 1 * BC1 + 1 * BC2)$$
$$PM10 = RHO * (1 * SS1 + 1 * SS2 + 1 * DD1 + 1 * DD2 + \mathbf{0.55 * DD3} + 0.85 * OM1 + 0.85 * OM2 + 0.85 * SU1 + 1* BC1 + 1 * BC2) \quad (1)$$

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.

p.8, line 6: "overestimated" sounds like it is a comparison to observation, but I think this means only in comparison to the modified model? Please make this clear, and if possible illustrate if this is an improvement against relevant observations. (It is not given that a model which appears to be more theoretically correct actually improves results.)

We agree and have clarified the text accordingly. We also have added a new section 3.4 where we have conducted two seven-month-long WRF-Chem simulations covering the period from June 1 to December 31, 2016. 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).

Fig. 8 shows 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.

p.8, line 19: it's not quite true that these are not accounted for - all of DUST1 is included if you add up the coefficients; it's merely that some of it is treated as larger than it should be (and thus less optically effective). Please clarify this in the text.

Thank you for your comment. The dust mass from the DUST1 bin was not omitted but was erroneously mapped onto coarser MOSAIC bin-sizes than required. This resulted in the underestimation of AOD, as discussed in the text. We have now clarified this issue throughout the text.

p.10, lines 3–4: please clarify what you mean here. A "function of natural logarithm of radius" is also a "function of radius". Do you mean specifically a linear function of each?

Not exactly; here we are talking about linear interpolation of a nonlinear function. It could be performed in radius space, or in ln(r) space. We argue that it is more accurate to interpolate in the ln(r) space, as discussed above. The text has been corrected to clarify this issue.

p.14, lines 20, 23, 27: again, “overestimation” suggests this is by reference to some actual measurement rather than to the modefied model - please clarify if this is just “one model version is higher than the other” or if the change is an improvement or degradation compared to measurements.

Thank you, the text has been clarified. Please see our response to comment p.8, line 6. (see above)

p.14, line 29–p.15, line 3: Section 3.4 is very brief – to be meaningful, this needs to show to what extent the contribution from both initial and boundary conditions is significant relative to one another and to sources within the domain.

We show the capability of Merra2BC in constructing IC&BC. ICs are important for making an accurate forecast. BCs impact is seen near the boundaries, and we show how far their signal propagates. The effect of BCs scales with the magnitude of the fluxes through the boundaries of a domain. We can make their effect stronger but that is not the point. The advantage of accurately calculated IC&BS is that they allow for improvements in such things as air-quality forecasts, and in many cases, for reliable use of a smaller domain, when the signal from boundaries is well defined.

p.16, line 9: as above, please clarify that submicron particles aren’t omitted as such, but treated incorrectly.

This issue has been clarified. Please see our response to comment p8, line19 (see above).

p.18, line 9: is it really correct that MERRA-2 has a globally-uniform constant surface pressure of 1000hPa? That seems highly unlikely in a meteorological reanalysis – please check and clarify, as this is what the current text suggests.

Thank you for pointing this out, the sentence in question has been deleted.

p.19, Figure A2: both singular and plural should be “species” (not “specie”).

Thank you, this has been corrected.

p.21, lines 7–11: please state the AERONET wavelength(s) from which these calculations are performed to generate the 550nm value.

We use a 440-675 nm wavelength range. The text has been modified to state this.

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.