

## **Responses to Reviewer 1**

This paper shows the dust and non-dust aerosol components over the Middle East that are available from two reanalysis products, MERRA-2 and CAMS, and WRF-Chem model simulations. It first compares the 10-m wind speed among the products, then compares AOD and size distributions with remote-sensing data (such as MODIS MAIAC and AERONET) and PM<sub>2.5</sub>/PM<sub>10</sub> concentrations with ground-based measurements. With the results from WRF-Chem simulations and reanalysis products, the aerosol composition of PM<sub>2.5</sub> and PM<sub>10</sub> is presented and days of PM above the regulatory standard are estimated.

I found this paper is interesting in the sense that an evaluation of model and reanalysis products is specifically performed for the Middle East region with, albeit limited, remote sensing and ground-based measurements, and the seasonal and annual levels of PM are presented. I however have quite some comments regarding the presentation and understanding of the products used in the paper, and recommend substantial revision before accepting for publication on ACP. I believe that the revision is not difficult to deal with although it could be extensive.

We thank the reviewer for the valuable comments. Despite poor air quality, the Middle East has very sparse air quality observations. So it is essential to thoroughly test the modeling tools. This is the first attempt to reconcile observation, models, and reanalysis products in this region.

All references supporting our response are placed at the end of the text. The reviewer's questions are in black. Our answers are in blue.

### **For the presentation style in general:**

- Abstract should contain only one paragraph with acronyms spelled out (e.g., ME).
- Introduction section is too long
- should be more concise and more relevant to the point of the study. It is not a literature review.
- Conclusion section is also too long and unfocused. It has 14 paragraphs! It should be consolidated with key points summarized and highlighted, not list everything you have done.

We agree and have revised the abstract, introduction, and conclusions sections accordingly.

Aerosol composition: there are no data to evaluate the models. The surface measurement data are for PM, not chemical species. Besides, the models do not include nitrate and ammonium, and it seems they don't have the chemical mechanisms for producing secondary organic aerosols. Therefore, the chemical composition from the model omits some important components. The problem should be acknowledged at least. Is there any reference for the aerosol composition in the region?

In-situ air quality observations in the Middle east are scarce. It is one of the known problems for air quality research in this area. The things are simplified a bit by the fact that in the ME dust dominates aerosol pollution. E.g., Calipso records dust in 95% of profiles (Osipov et al., 2015). The effect of nitrates, ammonia, and organics on AOD and PMs is insignificant in comparison with dust therefore the employed chemical scheme (*GOCART-RACM*) is adequate.

To support this conclusion, we have conducted a laboratory analysis of the chemical composition of soil and dust deposition samples that show a little presence of organics and ammonium (Prakash et al. 2016; Engelbrecht et al., 2017). According to (Engelbrecht et al.,

2017) in 2015 the annual average weight percentages of soluble ions of ammonium (NH<sub>4</sub>) and sulfate (SO<sub>4</sub>) in deposition samples taken at four sites at the KAUST campus are 0.05% and 2.513%, respectively. It means that available ammonium may neutralize at maximum 5% of sulfate mass. The actual contribution of ammonium sulfate should be lower, as some ammonium may also be bound as ammonium nitrate, ammonium phosphate, or ammonium chloride. We have added this explanation to the revised text (see Sec 4.1, last paragraph).

Reanalysis products: It should be pointed out that the reanalysis products from MERRA-2 and CAMS-OA are the reanalysis of AOD, not the mass concentrations of individual aerosol species. The mass of individual aerosol species is adjusted mostly proportionally according to the differences between the AOD from native model simulation and after the assimilation of satellite data. Also, in general, a better understanding of the reanalysis products and other products is needed.

We are aware of the reanalysis machinery and mentioned in the original text on p. 4, line 88: *“They improve the aerosol total column loadings through the assimilation of observed AOD but are not capable of assimilating the aerosol vertical structure and chemical composition.”* The representatives of both MERRA-2 and CAMS development teams are co-authors on this paper.

Comparisons with data: The comparison with AERONET AOD is not an independent evaluation of WRF-Chem and MERRA-2, because the WRF-Chem is “tuned” to match AERONET AOD and MERRA-2 assimilates AERONET AOD. This evaluation should be properly addressed.

This is partially correct. All satellite retrievals use AERONET observations for calibration. MERRA-2 assimilates AERONET AOD, but CAMS-OA does not. In WRF-Chem, we tuned to the annual average AOD to fit AERONET observations. We did not tune the temporal correlations between the model and AERONET data, just the mean bias. In this sense, the correlation coefficient, which is about (0.62-0.85), between WRF-Chem and AERONET AOD provides an independent evaluation of the WRF-Chem performance (see Table 4). We clarified this issue in the text (Sec. 5.2.2, last paragraph).

Also, for the AOD comparisons with both AERONET and satellite data, it is not clear if the comparisons were done under the same spatial and temporal conditions (e.g., models are sampled under clear-sky only condition or all-sky, if model and data are temporally matched).

The model, reanalysis, and observations are temporarily matched. It was mentioned in the original text (p. 17, line 364): *“Because AERONET conducts observations only during the daylight time, we interpolated WRF-Chem, MERRA-2, CAMS-OA AODs to the AERONET measurements times and then conducted time averaging to make simulated and observed AODs consistent.”*

To me, a major conclusion is that the PM<sub>2.5</sub> concentrations over the Middle East (at least at the places the study was examined) almost never below the WHO standard because of the dominance of dust in PM<sub>2.5</sub> which cannot or very hard to mitigate. Even if all anthropogenic emissions are shut down, the air quality in the Middle East will not improve. What is the implication for that? How to improve the air quality in the Middle East under such circumstance? This problem should be discussed.

WHO provides the guidelines, not air quality standards, that are subject to the national regulations. Yes, one of the important implications of this study is that anthropogenic pollutants in the cities are coming on top of the high aerosol background maintained by natural dust aerosols. This puts stricter requirements on anthropogenic pollution control. The effect of

natural pollution could be alleviated by using specific architectural planning, increasing in-city vegetation cover, and providing air quality forecasts to alarm the population about hazardous air quality. Our work is well in line with these ideas. The text is extended to include this discussion in page 32, in the conclusion section.

**Specific comments:**

Page 1, line 4: Spell out “ME”.

Fixed

Page 2, line 26: “mass budget” – it should be “emission budget”.

Fixed

Page 2, line 55-56: AVHRR was not designed to measure column aerosol properties. It was designed to observe clouds, surface temperature, and vegetation but later was expanded to retrieve aerosols over the ocean.

We agree. The text is corrected (Page 2, last paragraph).

Page 3, line 57: Change “CALIOPE” to “CALIOP”.

Fixed

Page 4, line 92: “we improve the latest. . . emission. . .” does not sound appropriate. You just use the new SO<sub>2</sub> emission data set. What is the spatial resolution of the new SO<sub>2</sub> emission data set from Liu et al. 2018? Does it match the WRF-Chem spatial resolution? Do you have to do “downscaling” or interpolation?

It is not precisely correct. We added ship emissions to OMI-HTAP and implemented (and improved in comparison what was there) this dataset in WRF-Chem. We modified the text to make it sound more appropriate (Sec. 4.1, 2nd paragraph). We also added a reference to our recently published paper, where this emission dataset has been used. The dataset is built initially on a 0.10°x 0.10° grid, and we conservatively interpolated the emissions on WRF-Chem 10 km x 10 km grid.

Page 6, line 123: Which wavelength is your chosen reference wavelength? What the wavelengths pair you used to calculate the Angstrom Exponent? Or did you use the Angstrom Exponent provided by AERONET?

We now mention (Sec. 2.1, after formula 1) in the text that we use the Angstrom exponent from AERONET that is provided for the 440-675 nm waveband.

Page 7, line 167-168: “MERRA-2 assimilates AOD at 550 nm from the AVHRR over the oceans”: This was done before the MODIS observations. MERRA-2 assimilates the MODIS AOD over the oceans since 2000.

This is correct. As explained in Table 2 from (Randles et al., 2017), since 2000 MERRA-2 assimilates MODIS and MISR data (over land and ocean) on the Terra satellite which has an equatorial overpass at 10:30 am UTC, while AVHRR has mostly orbited with the afternoon equatorial crossing time. Therefore MERRA-2 continued using AVHRR data over the ocean until 2002 when the Aqua satellite was launched. Since Aqua has an orbit with the equator

overpass at 2:30 pm, AVHRR data was no longer needed for coverage. This information was added to the revised paper (see Sec. 3.1).

Page 7, line 169: “specially processed MODIS observations. . .”: What product is that? Any references for such “non-standard” product?

Randles et al. (2017) in section 3 (subsection d) gives details on the aerosol observing system used in MERRA-2 for assimilation, including bias correction (see Sec. 3.1).

Page 7, line 184: “CAM5-OA assimilates MODIS observations”: Be more specific on what MODIS product(s) it assimilates.

CAMS-OA assimilates MODIS AQUA and TERRA AODs. It uses observations from Collection 5 since 20090901, and Deep Blue since 20150902. The clarification is added to the revised text (Sec. 3.2, last sentence).

Page 8, line 200: “wavelengths larger than 450 nm”???

We nudge only long waves. The text is corrected to read: “We nudge waves with wavelengths longer than 450 nm.” (Sec 4, 2nd paragraph).

Page 9, section 4.1: SO<sub>2</sub> is oxidized to form sulfate aerosol. It is described that gas phase SO<sub>2</sub>+OH reaction is done with the RACM, but it is not clear how the heterogeneous reactions are treated. Such description should be added.

Oxidation of SO<sub>2</sub> into sulfate is calculated within the GOCART aerosol module. Calculation of OH and other chemical reactions is done within RACM. There are no heterogeneous reactions in the RACM chemical mechanism, only gas-phase chemistry. The reference to (Stockwell et al., 1997) is in the original text. We clarified this point (Sec 4.1, 1st paragraph).

Page 9, line 234: “the first bin appears to be very poorly populated”: Why? The small particles should be transported by the winds more easily than the larger particles. Explain.

This discussion is related to sea salt. The sea salt droplets are relatively large and there is little mass accumulated in the first bin, therefore it is relatively unimportant. The text was corrected in a few places to clarify this issue.

Page 10, line 260-261: What was the error that you are correcting? Simply saying it was corrected because it was incorrectly calculated does not help the readers/users.

We have corrected three essential drawbacks in the code. These corrections have been tested and implemented in the official WRF-Chem release v4.1.3 (released on Nov 25, 2019). In the text, we provided a brief description of each of them (see Sec. 4.2). We also submitted a paper to GMD, where the effect of those errors has been quantified. *Firstly, we show that the diagnostic output of PM<sub>2.5</sub> surface concentration was underestimated by 7% and PM<sub>10</sub> was overestimated by 5%. Secondly, we demonstrate that the contribution of sub-micron dust particles was underestimated in the calculation of optical properties and thus, AOD was consequently underestimated by 25-30%. Thirdly, we show that an inconsistency in the process of gravitational settling led to the overestimation of the dust column loadings by 4-6%, PM<sub>10</sub> surface concentrations by 2-4%, and the rate of dust gravitational settling by 5-10%.*

Page 10, section 4.2: I don't think the bug-fix needs to be described in a devoted section. It can be summarized in a few sentences in the model description.

Sorry, we prefer to present this information in a separate section.

Page 12, Figure 3 caption: change "for" to "from".

Fixed

Page 11-13, section 5.1: Why not compare soil moisture and precipitation, since you mentioned on page 11 that dust emission and deposition are sensitive to the soil moisture and precipitation.

The ME, and especially the Arabian Peninsula, where primary dust sources are concentrated, are arid regions. Winds are the primary driver of dust generation there. The precipitation is sporadic, and soil moisture is always low. A comparison of soil moisture and precipitation could be essential in other regions of the world.

Page 14, line 321: What is "the lower atmospheric layer"? i.e., what is the altitude range the dust is emitted into? Or is it emitted into the lowest atmospheric layer? Please clarify

In the model, dust is emitted in the lowest model layer, but here we discuss the physical process in the real atmosphere. Saltation injects dust particles at about 0.1 m height (Martin and Kok, 2017). Dust is mixed up by turbulence in the near-surface atmospheric layer. It is a well-known process, and we do not mean giving here extra details.

Page 14, line 324-325: "But because. . ." this sentence has been said in the WRF-Chem description section. It does not belong here anyway.

We agree. This sentence is removed.

Page 14, line 328-329: "WRF-Chem underestimated. . ." What is the evidence for that? Is there any reference or from your own simulation describing that problem? This contradicts the findings by Kok et al. that global models overestimate the fine mode aerosols but underestimate the coarse mode aerosols.

Please see Figure 4 and explanations therein. Kok (2011) found that the models overestimate the emission of a fine dust mode, and Adebisi and Kok. (2020) suggested that the models underestimate the mass of the coarse (with radius  $r > 2.5 \mu\text{m}$ ) dust mode in the atmosphere almost four times, because of too fast removal processes. The argument is not entirely valid for the dust source regions like the Middle East where deposition, which Adebisi and Kok. (2020) blame for too-quick removal of coarse dust from the atmosphere, does not have enough time to do this. Adebisi and Kok. (2020) also analyzed dust size distribution in the near-surface layer where in-situ measurements are available.

Here we compare the column integrated dust volume size distribution from the model with the column integrated aerosol volume size distribution from AERONET and find that WRF-Chem underestimates the volume of fine particles with  $0.1 \mu\text{m} < r < 1 \mu\text{m}$  and overestimate the volume of particles with  $1 \mu\text{m} < r < 2 \mu\text{m}$ . We have to increase emissions in the first bin and decrease emissions in the second bin to correct this deficiency. The text is expanded to clarify this issue (Sec. 5.2.1).

Strictly speaking, our new  $s_p$  settings (see the modified text in Sec. 5.2.1, last sentence of the 2nd paragraph) are in line with (Adebiyi and Kok., 2020), as in comparison with the default  $s_p$  set we decreased the dust mass influx fraction into two finest bins 1 and 2 ( $0.1 \mu\text{m} < r < 1.8 \mu\text{m}$ ) from 0.3175 to 0.25 (see explanation to the reviewer's comment to Page 14, line 331, below), slightly increase the mass flux fraction from 0.2275 to 0.25 for the intermediate bin 3 ( $1.8 \mu\text{m} < r < 3 \mu\text{m}$ ), and increased dust mass flux fraction into two coarsest bins 4 and 5 ( $3 \mu\text{m} < r < 10 \mu\text{m}$ ) from 0.455 to 0.5. We have added the new Appendix 3 to the paper to explain these points.

Page 14, line 330: How do you know that the total emitted dust mass is overestimated, since there is absolutely no measurements of dust emission?

We here do not mean to compare the simulated dust emissions with the absolute value of real dust emissions that are not measured. The measured physical quantity is AOD that in the model is controlled by emissions. If the model overestimates AOD, this is associated with overestimating dust emission (assuming we do not touch the dust removal processes). If the model excessively emits large particles, it generates higher dust mass flux than if it would generate emitting more fine particles, because finer particles produce a larger AOD per unit mass. We changed the wording to clarify this issue. (Sec. 5.2.1, 2nd paragraph)

Page 14, line 331, adjusted  $s_p$  fraction: The first size bin represents clay and the rest four bins represents silt. The 0.1, 0.25, 0.25, 0.25, 0.25 fractions is based on the assumption that 10% of clay will be emitted but 100% of silt is subject to be emitted to the atmosphere based on the early work in the 1990s from Tegan. Even though these numbers are arbitrary, but the sum of adjusted silt fractions (0.15, 0.17, 0.38, 0.1) is only 0.8. Please explain why you do not account for the rest of 0.2 fraction in the silt group.

The GOCART dust emission formula (2) calculates dust mass flux into the atmosphere within five dust bins. In this formula the factor  $C$  controls the total mass flux, and the  $s_p$  coefficients split the total mass flux into five different size fractions. Following this logic, we have to assume that the sum of  $s_p$  equals 1 as we stated in the text. In the revised paper, we reiterated the sensitivity of the results to the choice of  $s_p$  and slightly readjusted the  $s_p$  values. Now we use the set of  $s_p = (0.15; 0.1; 0.25; 0.4; 0.1)$ . It means that 15% of the total dust mass flux is coming as clay and 85% as silt.

In the original formulation the sum of  $s_p$  equals to 1.1. It is not crucially important, as the total flux is multiplied by the factor  $C$  that is tuned to fit the observed optical depth. So we can normalize the original  $s_p$  coefficients by dividing them to 1.1 and multiplying constant  $C$  to 1.1. It will not change any results in (2) but gives the  $s_p$  set of (0.09, 0.2275, 0.2275, 0.2275, 0.2275) that is normalized to 1 consistently with our approach. We have added the new Appendix 3 to the paper to discuss these points.

Page 16-18, Section 5.2.2: As I mentioned at the beginning, the comparison with AERONET AOD is not an independent evaluation of WRF-Chem and MERRA-2, because the WRF-Chem is "tuned" to match AERONET AOD and MERRA-2 assimilates AERONET AOD.

It is not exactly correct, at least for the model. The model is tuned to match the average value of AOD at AERONET sites using a spatially uniform time-independent factor  $C$ , which controls total dust emission. So we tuned the time-averaged AOD bias concerning available AERONET observations, not the correlation coefficient. Therefore, the high correlation of (0.62-0.85) between simulated and AERONET AOD is an independent proof of the model performance. To clarify this issue, we have expanded the text (Sec. 5.2.2, last paragraph).

Page 17, Figure 6: There are several very large spikes of AOD from the WRF-Chem simulations in Mesaira and Sede Boker in 2016. What causes these spikes?

Thanks for catching these spurious AOD spikes. We have analyzed the meteorological fields from our run for July 2016. We found that on the 27th of July 2016, a high-pressure system in the Eastern Mediterranean moving south-eastward formed high-pressure gradients reaching 3 hPa/100km. This system forced a strong gradient wind with speed exceeding 15 m/s and associated dust generation. MERRA-2 and CAMS, as well as synoptic charts based on *in-situ* observations, suggest that WRF-chem overestimates the sea level pressure gradient (see Figure A below). The preliminary analysis indicates that the boundary conditions calculated using MERRA-2 fields generated the spurious meteorological system. We re-calculated the entire July of 2016 with the boundary conditions from ERA-Interim reanalysis (see Figure A, top right panel). In the new run, the sea level pressure looks similar to observations, and spurious AOD spikes disappeared (see Figure B). We have incorporated the new July-2016 results in our analysis and corrected the figures and tables in the paper accordingly.

Sea Level Pressure Anomaly respect to mean domain SLP 06 UTC 27 July 2016

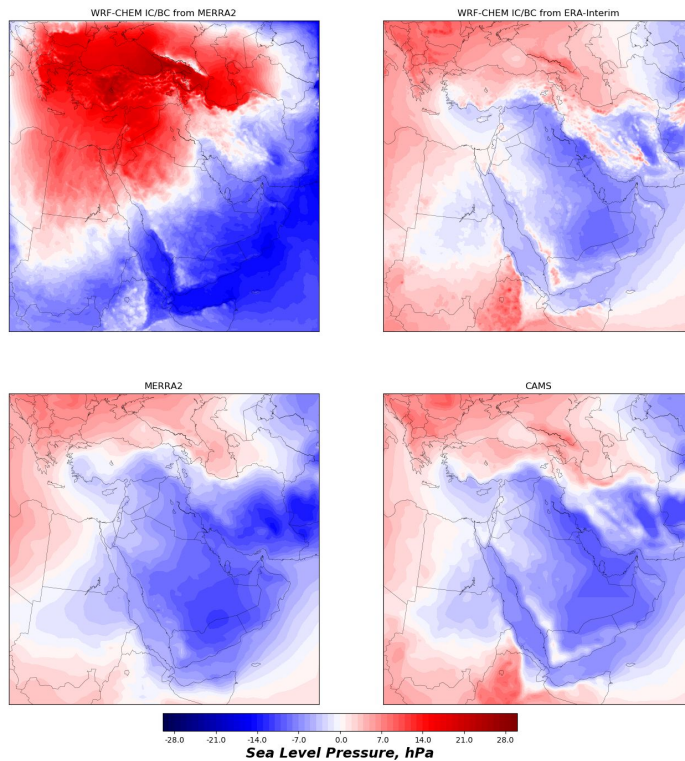


Figure A. Sea Level Pressure anomaly from MERRA-2, CAMS-OA, and two WRF-CHEM runs with MERRA-2 and ERA-Interim boundary conditions.

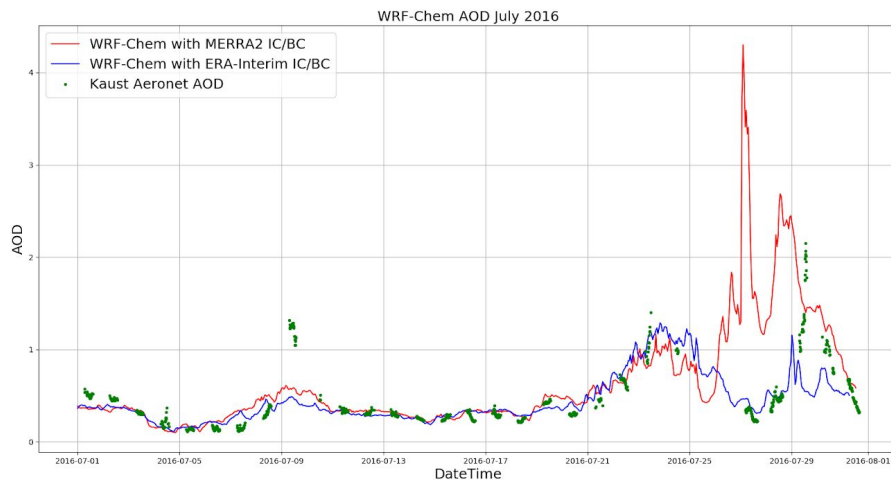


Figure B. Simulated and Observed AERONET AODs at the KAUST site for July 2016.

Page 19, Figure 7: How are the models sampled when compared to satellite data? Are they temporally matched (i.e., model results are concurrent with the satellite data, or model results are averaged for the clear sky only during the season)?

In the revised paper, we sampled WRF-Chem, MERRA-2, CAMS-OA during a day-light time (6 am-2 pm UTC or 9 am-5 pm local time) (Sec. 5.2.2, 1st paragraph). The results are not visibly different from our previous estimates when we applied 24-hour sampling.

Page 20, line 388-389: “. . . in good agreement with. . .”: What is your criteria for “good agreement”? In general, such subjective statement should be avoided. Instead, you could say something more quantitative, such as “with xx%” or “correlation coefficient within xx-yy”.

We agree and have corrected the sentence to account for the reviewer’s comment. Now it reads, as “Based on the comparison of WRF-Chem AOD with the AOD from MODIS and AERONET observations, we conclude that spatial and temporal WRF-Chem’s AOD distribution is in good agreement with the available satellite and ground-based observations, i.e. annual mean correlation coefficient  $R$  exceeds 0.6 (see Tab. 5) and correlation with AERONET is 0.43-0.85 (see Tab. 4).” (Sec. 5.2.3, last sentence).

Page 20, line 406: “. . . good agreement. . .” again! See my comments above.

This sentence was corrected. See the previous comment.

Page 21, line 420: sulfate ion: So for PM<sub>2.5</sub> you only consider the mass of sulfate ion, not neutralized sulfate that exists in the atmosphere, such as ammonium sulfate? The mass of ammonium sulfate is 37% more than just sulfate ion.

We understand that ammonium sulfate has a bigger mass than sulfate ion. But in our region of interest, there is little ammonia to neutralize a significant amount of sulfate. Therefore, we assume that most of the ME anthropogenic aerosol is sulfate.

We repeat here our response to the major concern:

To support this conclusion, we have conducted laboratory analysis of the chemical composition of soil and dust deposition samples that show a little presence of organics and ammonium (Prakash et al. 2016; Engelbrecht et al., 2017). According to (Engelbrecht et al., 2017) in 2015 the annual average weight percentages of soluble ions of ammonium (NH<sub>4</sub>) and sulfate (SO<sub>4</sub>) in deposition samples taken at four sites at the KAUST campus are 0.05% and 2.513%, respectively. It means that available ammonium may neutralize at maximum 5% of sulfate



mass. The actual contribution of ammonium sulfate should be lower, as some ammonium may also be bound as ammonium nitrate, ammonium phosphate, or ammonium chloride.

Page 21, line 422: again, what is the reason that “the first sea salt bin is poorly populated”?

This is related to sea salt size distribution. The sea salt is a relatively coarse aerosol with very poor fine fraction therefore the fine model bin is poorly populated. We have modified the analysis and accounted for the contribution of the first sea salt bin in PM for MERRA-2 in the revised paper. The results did not visibly change.

Page 21, line 426-429, PM calculations: It should be noted that all models do not include nitrate and ammonium when calculating the PM mass. Associated error/uncertainty should be estimated.

Yes, for consistency, we show the contribution of only  $\text{SO}_4$  for all models.

Page 24, 3rd line from the bottom: “As we have shown, WRF-Chem provides reliable estimates. . .”: What is the criteria for "reliable"? From Fig. 8, WRF-Chem underestimates  $\text{PM}_{2.5}$  at Jeddah and Riyadh by a factor of 2 and overestimates  $\text{PM}_{2.5}$  at Dammam. Its performance for total  $\text{PM}_{2.5}$  is inferior to CAMS. In addition, its chemical composition of PM have not been evaluated at all.

In this study, we evaluate the performance of the WRF-Chem and the best available assimilation products over the Arabian Peninsula using observed PM concentrations. This region has a poorer observation coverage in comparison with Europe or the US. Therefore model estimates are valuable to plan further analysis and mitigation measures.

The situation in the Middle East is simplified by the dominance of dust in the PM. WRF-Chem does a good job in comparison with MODIS and AERONET AOD observations, as well as predicts well the distribution of  $\text{SO}_2$ , which is the only sulfate precursor, see (Ukhov et al., 2020).

The calculations of surface aerosol concentration within a city is challenging for the  $10 \times 10\text{-km}^2$  resolution model in comparison with the point observations. E.g., we do not account for in-city dust generation, although there could be a significant amount of resuspended dust. So the larger discrepancies in PM concentrations within the city are expected.

Page 27, Figure 10: The labels and legends on this figure are way too small to be legible.

Fixed

Page 28, line 516-517: Is this a “drift of sulfate”? What is the emission patter of  $\text{SO}_2$ ?

As it is stated in the paper, Figure 10f shows  $\text{SO}_4$  concentration. The OMI-HTAP  $\text{SO}_2$  combined emissions are presented in (Ukhov et al., 2020).

Page 28, line 517-521: I don't understand what the relevancy is to refer the sulfate concentration over the US.

We added this sentence for comparison of sulfate concentrations over the US and the ME.

Page 28, line 523, MERRA-2 underestimates the SO<sub>2</sub> emission: Do you know if indeed sulfate is too low or SO<sub>2</sub> emission is too low in MERRA-2? Several issues here to challenge such statement. First, sulfate mass in MERRA-2 is not necessarily corresponding to SO<sub>2</sub> emission because the aerosol masses (including sulfate) are adjusted after the AOD simulation, which has nothing to do with SO<sub>2</sub> emission. Second, van Donkelaar's work "retrieved" PM<sub>2.5</sub> based on the satellite AOD and the GEOS-Chem model such that the sulfate (and other aerosols) concentration is adjusted based on the adjustment of model AOD to satellite total AOD. As a result, the sulfate from van Donkelaar's work is not necessarily representative of the "true" sulfate concentrations.

We clarified the text and added a reference on our recently published paper (Ukhov et al., 2020), where we compare different SO<sub>2</sub> emission dataset including EDGAR-4.2 used in MERRA-2. Ukhov et al. (2020) shown that EDGAR-4.2 underestimates SO<sub>2</sub> emissions over the Arabian Peninsula in comparison with the new OMI-HTAP SO<sub>2</sub> emission dataset.

Page 28, line 538-539: Again, I don't understand what the relevancy of the US-EPA standard being applied here. The Saudi Arabia's standard should be used. And in line 539, now you use the WHO guidelines as reference. This is confusing.

We can not avoid the comparison of air quality in the Middle East with air quality in the US and Europe. For this purpose, we specifically discussed all PM air quality limits in Table 1, and apply them when appropriate. We specifically discuss WHO guidelines, European, US, and Saudi Arabian air pollution limits to comprehensively evaluate PM pollution in the ME, and quantify its sources.

Page 30, line 554 and 556-557: I would not emphasize "for the first time" to elevate the significance of the paper. Simply state what you've done and found is more appropriate.

We agree, the wording "for the first time" is removed from the text.

Page 30, line 564: "The air pollution in the major Middle Eastern cities is evaluated" sounds overstatement. The evaluation is rather limited to only three cities and only with PM<sub>2.5</sub> and PM<sub>10</sub>, not all major cities and not all pollutants.

This sentence is replaced by "*The PM levels in the major ME cities are estimated.*"

Page 30, line 576: "improve calculation of sulfate aerosol": there is no approve that sulfate simulation is improved because there is no data to evaluate it.

The reviewer technically is correct. Strictly speaking, we did calculations with the improved SO<sub>2</sub> emissions (see Ukhov et al., 2020) that affected sulfate concentrations. The text is revised to clarify this point.

Page 31, line 582-583: CAMS-OA deficiency has been corrected: Then why don't you use the latest version that is available in 2019? What is the point to evaluate the results from an obsolete model version?

The paper evaluates the operational CAMS product. So for any given time, only the forecast and analyses of the current operational version is available. Further, rerunning the CAMS system (with data assimilation at the full resolution) is quite expensive. So it can not be easily redone. The re-analysis has a frozen model version for the whole period. So we always evaluate the best product at the time. E.g., CAMS-OA had an important upgrade of the

horizontal resolution of the operational system from T255 (80km) to T511 (40 km) on 21.6.2016. The CAMS-OA product is still in use and is distributed by ECMWF, so an independent evaluation of the existing product is useful. The evaluation period of 2015-2016 does not cover the time when the latest changes in CAMS-OA were introduced, so the evaluation of the newest version can not be done in the current study.

Page 31, line 589: “quite well” – again! Please avoid using such subjective statement.

We agree. The text is corrected.

### **References used in the response:**

Adebiyi AA, Kok JF. Climate models miss most of the coarse dust in the atmosphere. *Science Advances*. 2020 Apr 1;6(15):eaaz9507.

Engelbrecht J, Stenchikov GL, Prakash PJ, Lersch T, Anisimov A, Shevchenko I. Physical and chemical properties of deposited airborne particulates over the Arabian Red Sea coastal plain.

Kok, J. F. (2011). A scaling theory for the size distribution of emitted dust aerosols suggests climate models underestimate the size of the global dust cycle. *Proceedings of the National Academy of Sciences*, 108(3), 1016-1021.

Martin, R. L., & Kok, J. F. (2017). Wind-invariant saltation heights imply linear scaling of aeolian saltation flux with shear stress. *Science advances*, 3(6), e1602569.

Osipov S, Stenchikov GL, Brindley H, Banks J. Diurnal cycle of the dust instantaneous direct radiative forcing over the Arabian Peninsula.

Prakash PJ, Stenchikov GL, Tao W, Yapici T, Warsama BH, Engelbrecht J. Arabian Red Sea coastal soils as potential mineral dust sources.

Randles CA, Da Silva AM, Buchard V, Colarco PR, Darmenov A, Govindaraju R, Smirnov A, Holben B, Ferrare R, Hair J, Shinozuka Y. The MERRA-2 aerosol reanalysis, 1980 onward. Part I: System description and data assimilation evaluation. *Journal of Climate*. 2017 Sep;30(17):6823-50.

Stockwell WR, Kirchner F, Kuhn M, Seefeld S. A new mechanism for regional atmospheric chemistry modeling. *Journal of Geophysical Research: Atmospheres*. 1997 Nov 27;102(D22):25847-79.

Ukhov A, Mostamandi S, Krotkov N, Flemming J, da Silva A, Li C, Fioletov V, McLinden C, Anisimov A, Alshehri Y, Stenchikov G. Study of SO<sub>2</sub> pollution in the Middle East using MERRA-2, CAMS data assimilation products, and high-resolution WRF-Chem simulations. *Journal of Geophysical Research: Atmospheres*. 2020 Mar 6:e2019JD031993.