Responses to Reviewer 2

General comments

The manuscript by Ukhov et. al., presents a detailed comparison of WRF-Chem, MERRA-2 and CAMS aerosol data with respect to the two MODIS data products, AERONET and ground-based network of PM measurements. The paper is well written and extensive sets of data are considered for comparison which represents aerosol optical depth, PM10, PM2.5 and their spatial and temporal patterns. In addition to the comparison, the composition of aerosol among dust, sulfate, sea salt and other constituents have been discussed. The impact on aerosol air pollution has also been investigated. This study could have been completed by also including some comparison for vertical profiles of aerosol extinction or various components of aerosol (e.g. dust, sulfate), with the measurements (if available) or at least among the model and assimilated products. I have a few major and several minor concerns with the manuscript, which upon being addressed, I recommend publication in ACP.

We thank the reviewer for the valuable comments. We agree that comparing vertical aerosol distribution in the models and in observation is very useful. But it is not a small side issue. To address it, we have established a micropulse lidar (MPL) site at the KAUST campus and observed the aerosol vertical profile since 2014. There is no way currently to separate in observations the vertical profile of dust from that of sulfate and other aerosols. We are not sure we have enough space in this paper to address the important issue of the aerosol vertical profile properly; therefore, we refrain from doing this here. This work is mostly the subject of another paper submitted to ACP recently.

All references used to support our responses to the reviewer's comments are presented at the end of the text. The reviewer's questions are in black. Our answers are in blue.

Major issues:

1. The manuscript primarily focuses on the various aerosol product and WRF- Chem code has been modified to calculate these parameters. In section 4.2, authors mention that the code modification will be published in a forthcoming publication. Since the data produced for use in this publication is simulated with modified model code, yet not peer-reviewed, I can only recommend publication after the technical publication.

The model modifications have been reviewed, tested, and implemented in the v4.1.3 official version of the WRF-Chem code (released on Nov 25, 2019). The forthcoming paper in GMD is about the quantitative evaluation of the effects of those changes on simulations. The GMD paper is independent of the current study, and, we believe, should not delay the publication of the present manuscript.

2. For sections 2.2, 5.2.2 and 5.2.3 authors use the MODIS combined deep blue (DB) and dark target (DT) product. It a level 3 gridded product at a much coarse spatial resolution of 1◦ × 1◦. DB has poorer performance over water, while DT has limitation over land. In my opinion, authors should use separate DB and DT, level 2 gridded products, which are available at much finer (10km × 10km) resolution (comparable to WRF-Chem and MAIAC). Moreover, level 2 product also allows the possibility of applying a quality assurance criterion, which has shown improvement in the comparison previously (for e.g. Liu, N., et al. (2019)).

Following the reviewer's recommendation, we calculated the 10x10 km^2 MODIS DB&DT level 2 AOD product. The AOD fields with 10x10 km^2 and 100x100 km^2 resolution are shown below. We see that qualitatively the AOD structure is similar at both resolutions, but 10x10 km^2 fields have much more fine details. In the revised manuscript, we now use the 10x10 km^2 resolution MODIS DB&DT product, and all tables in the revised paper are corrected accordingly.

Figure A. 10 x 10 km² MODIS AOD

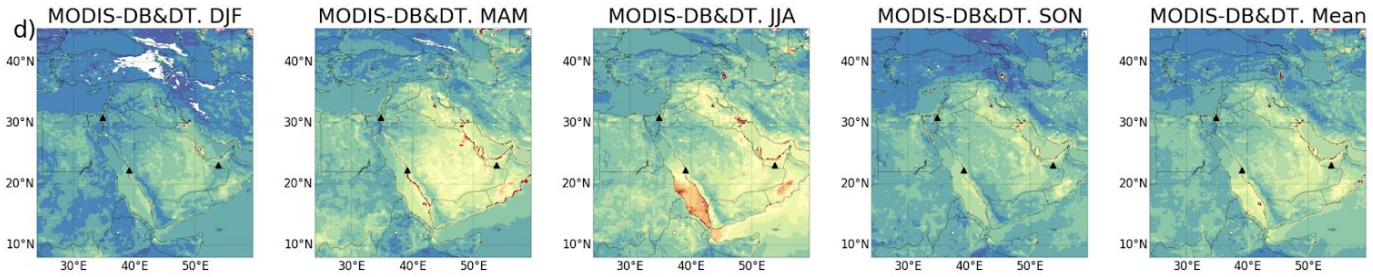
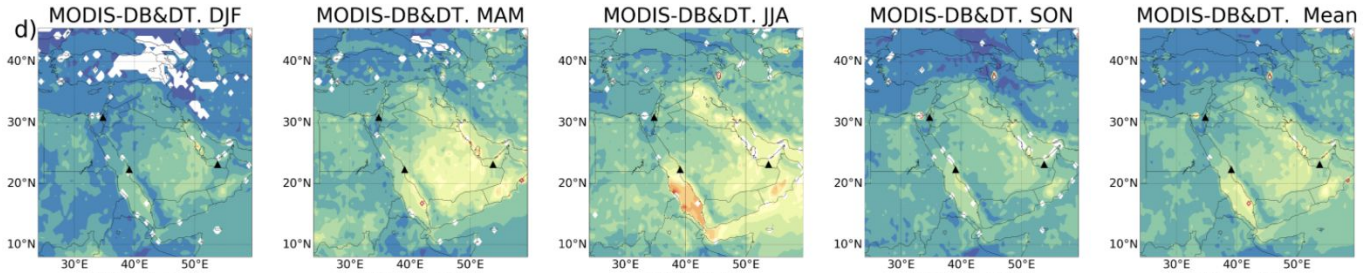


Figure B. 100 x 100 km² MODIS AOD



2. For comparison with MODIS data products, the model data should be sampled around satellite overpass or at most averaged ± 1h around satellite overpass. Further, in order to avoid sampling bias, only those days should be considered for calculating seasonal means when both measurements (AERONET/Satellite) and model (or assimilation) data are available.

Because we analyze the seasonal mean AOD fields over the entire ME, we have to use multiple overpasses to compile a map for the whole domain. Therefore AOD's from WRF-Chem, MERRA-2, CAMS-OA are sampled at the day-light time (6 am-2 pm UTC or 9 am-5 pm local time). We added this explanation in the text (Sec. 5.2.3, 1st paragraph).

There is no missing data in the model and the reanalysis outputs, and only a little in the MODIS products, therefore all available observations and model outputs were used for calculating seasonal means. The undefined pixels detected in observations are synchronously excluded from the statistical analysis in all datasets.

4. The introduction is very long in general and can be curtailed by only keeping the content most important to the study. Some (not all) suggestions: - Lines 55-60- MISR, AVHRR and CALIOPE are not relevant to this study. Some restructuring is also needed. For example, the description of the work to be presented in the text fits better towards the end of the introduction. In line 75, authors mention about evaluation to be presented in the subsequent section but this is followed by further literature review. Line 93 again starts with the highlight of work to be presented in this work.

We revised the text to remove the redundancies

5. The conclusion needs to be curtailed. Redundancy in the conclusion can be reduced. Some examples (not all): Lines 558-559 and lines 571-543; Lines 560- 561 and 574-575. Numbers should be provided in conclusion rather than only qualitatively stating “overestimate” / “contribute” etc.

We revised the conclusion and made it more concise.

Detailed Comments:

1. Abstract: Please use abbreviations only after providing their full form at the first use (e.g. ME). Abstract does not do justice to the manuscript. Some more key finding should be added.

Added full form for ME (Middle East). More key findings were added.

2. Line 6: WRF-CHEM code was modified but this is not described in detail in this manuscript. Authors wish to publish it as separate manuscript and hence this does not fit to abstract.

We agree and have removed this sentence from the abstract.

3. Line 15: rich – reach

Fixed

4. Lines 15-16: Contribution of both organic matter and black carbon are negligible. Is is important to mention this comparison?

The PM speciation is vital to plan air quality mitigation measures. There are few observations available, so model results that provide spatially resolved information are valuable for understanding the effect of different types of aerosols on air quality in the ME.

5. Line 35: Essential – Important/crucial

Fixed

6. Line 46: PM10 and PM2.5 are defined with respect to “aerodynamic diameter”.

Yes

7. Lines 61-63: Is it justified to compare the 21 days’ mean with air quality regulation standards for 1 year. Please note that some of the measured mean PM10 concentrations are smaller than 24 hours’ limit.

Due to the modification of the introduction, this part was removed from the text.

8. Lines 79-85 : What are the conclusions of these comparisons?

The text was rephrased to: “*These data assimilation products adequately reproduce AOD and PM concentrations at different regions of the world (Provençal et al., 2017; Buchard et al., 2017; Cesnulyte et al., 2014; Cuevas et al., 2014).*”

9. Lines 89-92: Given that mineral dust contributes 75-95% of the PM, how much discrepancy is caused by outdated emission inventories in MERRA-2 and CAMS-OA?

The anthropogenic emissions certainly make an essential contribution in the air pollution in the cities and this information is important for air quality control in urban centers. To clarify this issue we have added the following sentence (p.4 line 75): “*E.g.,SO2 emissions used in MERRA-2 and CAMS-OA differ by 45-50% in some ME regions (Ukhov et al., 2020)”.* In the 2nd paragraph of the Sec. 4.1 we also mentioned that 14 previously unaccounted SO_2 point sources located in the ME were included in the new OMI-HTAP dataset.

10. Line 113: What are CIMEL and PREDE?

To clarify the text the sentence is updated to read "*AERONET comprises more than 1000 observation sites equipped with CIMEL sunphotometers and PREDE skyradiometers manufactured in France by CIMEL and in Japan by PREDE.*"

11. Lines 118-120: Authors should also provide a statistical comparison for the case when only cloud screened and quality-assured data are used in the results and discussion.

In our analysis now we use "Utilizes AERONET AOD, which is pre- and post-field calibration applied, automatically cloud cleared and manually inspected (Level 2.0 AOD)." The text (Sec 2.1, 1st paragraph) is clarified to read: *"We utilized level 2.0 (cloud screened and quality assured) AERONET AOD data."*

12. Line 119, 122: Angstrom – Ångström

Fixed

13. Line 139: MAIAC also provides AOD at 470nm.

Added

14. Section 2.2: Please mention the Quality assurance filter criteria if applied!

We did not use a quality assurance filter.

15. Line 153? quarterly refers to what?

"Quarterly" refers to the calibration audit. Sentence has been rephrased to read (Sec 2.3. Last sentence) "*...audit is conducted quarterly by Ricardo-AEA Ltd...*".

16. Line 164 (DMS)

Fixed

17. Line 173: This line is not clear to me.

CAMS-OA is the operational analysis, not reanalysis. The model, its horizontal resolution, and assimilation routine are improving on the way, so we always use the best available product. E.g., the important upgrade of the horizontal resolution of the operational system from T255 (80km) to T511 (40 km) was accomplished on 21.6.2016.

18. Line 179 and later in the text: dustbins – dust-bins.

Fixed

19. Line 2019: OH is hydroxyl radical and not "Hydroxide radical".

Fixed

20. Line 219: I had difficulties understating the treatment of PM, BC and OC emissions. Black carbon, organic carbon and dust, these are already included in PM. So if the emission of both PM and its constituents are specified separately, this would end up in doubling of certain constituents of PM.

Reviewer meant line 229 not 219. We agree that it sounds confusing, because we followed the emission categories used in the WRF-Chem. It meant that the "PM" emissions comprise the additional aerosol biogenic and fossil components. Now the text reads as (Sec 4.1, 2nd paragraph): "*All other constituents (other PM from biogenic and fossil components, black and organic carbon, etc.), ...*"

21. Equation 3: Use of S in both LHS and RHS are confusing. I would suggest using Smod or S' or something different.

We agree, this is confusing. We defined S' as a modified topographic source function.

22. Lines 250-253: How is the value of C=0.5 achieved? The tuning of C with respect to measured AOD should be discussed in more detail.

WRF-Chem is tuned to reduce the seasonal mean AOD biases with respect to AERONET observations. The value of *C*=0.5 obtained in the course of multiple WRF-Chem runs with different values of *C* gives the best AOD fit. Three references with detailed description of the tuning procedure were provided in the original text (page 10 line 248): (Kalenderski et al., 2013; Jish Prakash et al., 2015; Anisimov et al., 2017).

23. Section 4.2: How are the diagnostic output of PM are different from those calculated in section 5.3?

As we mention in the original text, if we would use the default WRF-Chem v3.7.1 code we would overestimate the PM10 and underestimate the PM2.5 surface concentrations. For typical Middle East conditions, *diagnostic output of PM2.5 surface concentration could be underestimated by 7% and PM10 surface concentrations could be overestimated by 5%.*

24. Section 5.1 Lines 271-275 fit better for methods/domain description.

Sorry, we prefer to concentrate on the ME climate description in section 5.1.

25. Figure 2: What is the physical significance of the topographic source function? Do the high values represent higher dust emission potential?

The topographic source function defines a spatial pattern of emission. The factor *C* - controls the total amount of emissions. The topographic source function has been built under an assumption that low-land areas accumulate fine-scale material (Ginoux et al., 2001). The areas with the higher values of source function generate higher dust emission flux, see eq. (2).

26. Line 283: Missing ")".

Fixed

27. Lines 305-309: Higher R and lower RMSD for V are not specific only for summers.

We agree. The text has been corrected accordingly.

28. Table 3: How are the statistics for Autumn and Spring

The dry subtropics have essentially two seasons, warm Winter, and very warm Summer. The intermediate seasons are not so essential. We prefer not to spend much time on their discussion.

29. Table 3,4 and 5: Slope/Bias should also be provided in addition to the R and RMSD. These quantities provide an idea about overestimation/underestimation/trend.

Table 4 shows bias for the AOD time series. We added the scatter plots for the AOD time series in Fig. 6. The bias has been added for the spatial distributions of AOD in Table 5. Figure 3 shows that the seasonal mean wind field in WRF-Chem and both reanalyses do not have systematic differences. We believe it is not needed to add bias for the wind in Table 3.

30. Line 315: Aerosol content is also characterized by other quantities apart from AOD.

We talk here about the satellite observed quantities. Of course, one needs aerosol size distribution or mass extinction coefficient to convert AOD to mass loading.

31. Lines 327-330: It would be nice to see the underestimation/overestimation with default sp fraction and its magnitude as a figure (at least in appendix).

In the revised paper, we reiterated the sensitivity of the dust size distribution 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). Below we compare the size distributions obtained in the simulations with this updated and the default s_p=(0.1; 0.25; 0.25;0.25;0.25) values for summer of 2015. Using the updated s_p values improves the size distribution fit (see Figure below). We have added the new Appendix 3 to the paper to discuss these points.

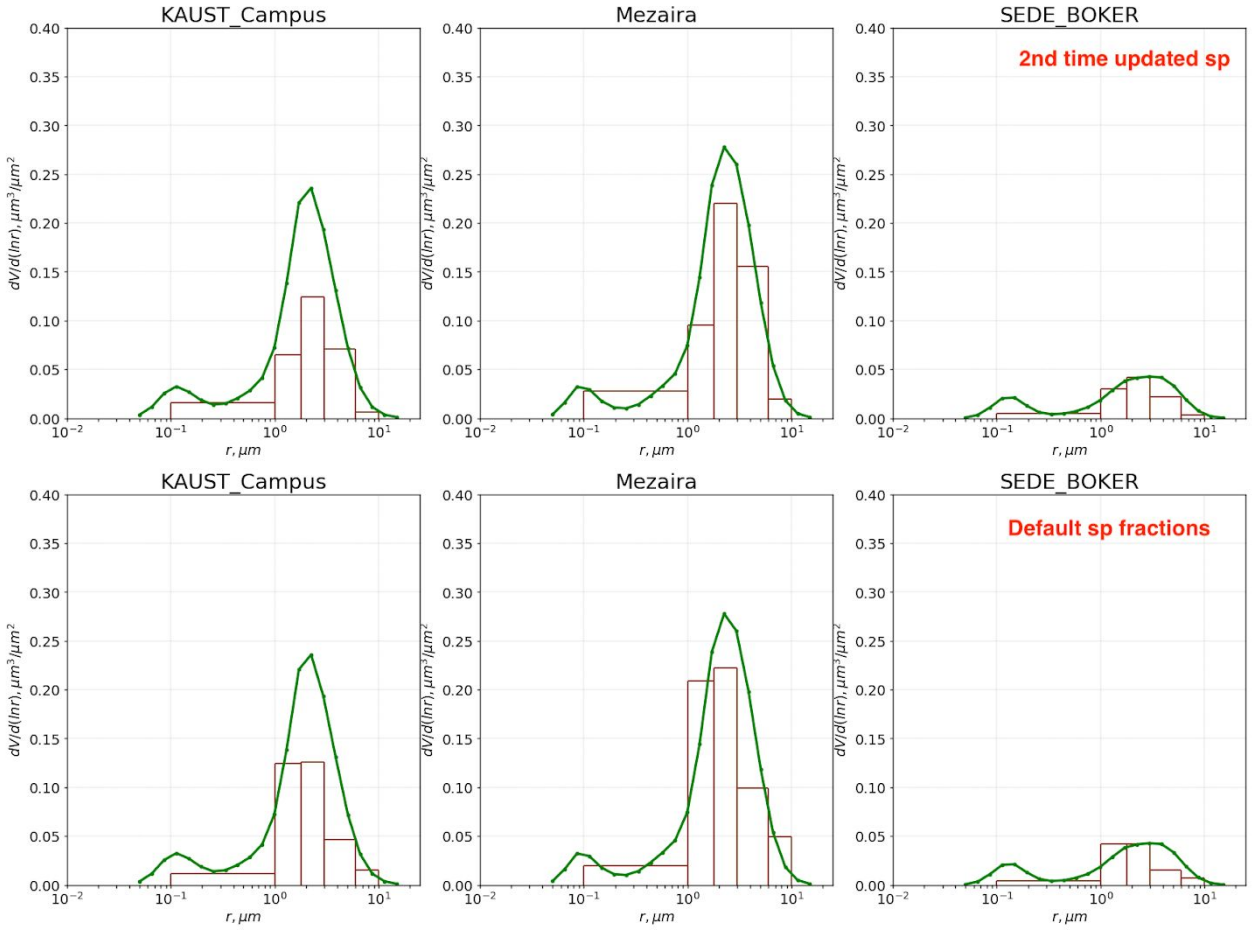


Figure C. Volume size distribution at KAUST AERONET site averaged for JJA of 2015 from WRF-Chem simulation with default sp={0.1,0.25,0.25,0.25,0.25} fractions (bottom) and updated sp={0.15; 0.1; 0.25; 0.4; 0.1} (top).

32. Line 345: This line should only be kept if the evaluation of updated CAMS-OA is shown in the manuscript.

Figure 4 compares the volume size distributions from WRF-Chem, MERRA-2, and CAMS-OA with the AERONET retrievals. The paper evaluates the operational CAMS product, CAMS-OA. So, for any given time, only the forecast and analyses of the current operational version are available. Further, rerunning the CAMS system (with data assimilation at the full resolution) is quite expensive. So it can not be easily redone. The 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 comparison can not be made in the current study.

33. Figure 6: Please mention that panel A corresponds to 2015 and B correspond to 2016.

The caption has been changed to address this issue.

34. Line 366: At a given location, up to 4 measurements are possible on several days due to overlap of two orbits each for TERRA and AQUA.

Thanks. The text is corrected.

35. Section 5.2: I was surprised to see that MAIAC underestimates AOD with respect to AERONET. The evaluation of MAIAC by Lyapustin et. al., shows overestimation at all the three AERONET sites shows in this study. Authors should address, why even for a similar dataset, an underestimation is observed in this study by MAIAC. Authors could also refer to the finding of Liu et. al., 2019, where they have found that applying a QA filter significantly reduces the Deep blue (over land) AOD from MODIS over China. There are other evaluation studies (e.g. Liu et al., 2019, Mhawish et. al., 2019), which have found MAIAC to be more accurate than Deep blue and Dark Target. Authors should address, why for their domain this is not the case.

According to (A. Lyapustin personal communication, April 2020), MAIAC underestimates AERONET in the ME (at *KAUST_Campus* and *Mezaira* sites). So, our results are consistent with this. We do not apply a QA filter in our calculations.

36. Figure 7, I wonder how there are NAN values at around 40 ◦N 40 ◦E in MODIS DB&DT products in the annual mean but there are no NAN values in MAIAC annual mean. If the seasonal NAN values are removed by annual mean, this should hold valid for both the MODIS data products. I would recommend the authors to recheck the calculation of spatial means. Please also indicate the location of three AERONET site in Figure 7. This would help the reader to follow the discussion.

There are some undefined pixels in the MAIAC product that we referred to as NANs. This confusing terminology has been corrected in the revision. In Figure 7 in the original manuscript, we interpolated MODIS and MAIAC AODs to the MERRA-2 grid. That caused some discrepancies, e.g., led to an artificial increase in undefined areas. Now model outputs and satellite products are plotted in its original resolution (see Figure 7 in the manuscript and figure below). We have recalculated all statistical characteristics in Table 4 using MAIAC on its original grid. Table 4 shows that MAIAC now compares better with AERONET than MODIS DB&DT in terms of bias and correlation coefficient. Locations of 3 AERONET sites are now shown on the plots, as requested by the reviewer. We also fixed an error in the calculation of spatial means.

Figure D. MODIS 100x100km^2 and MAIAC (interpolated on MERRA-2 grid) and error in calculation of seasonal means.

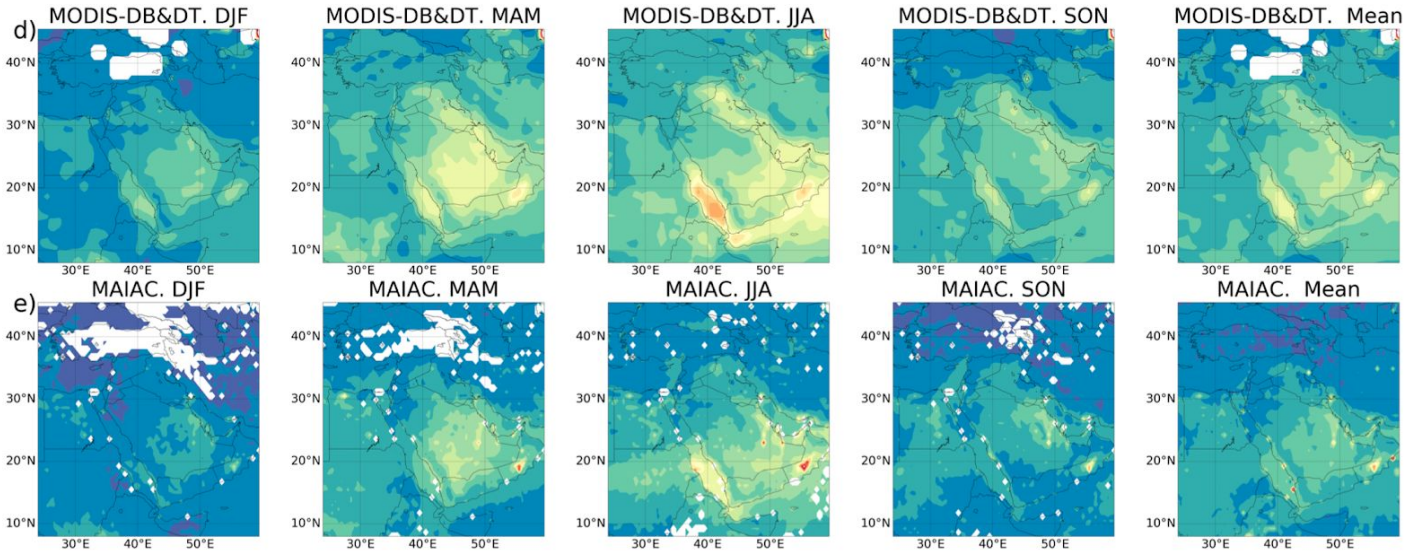
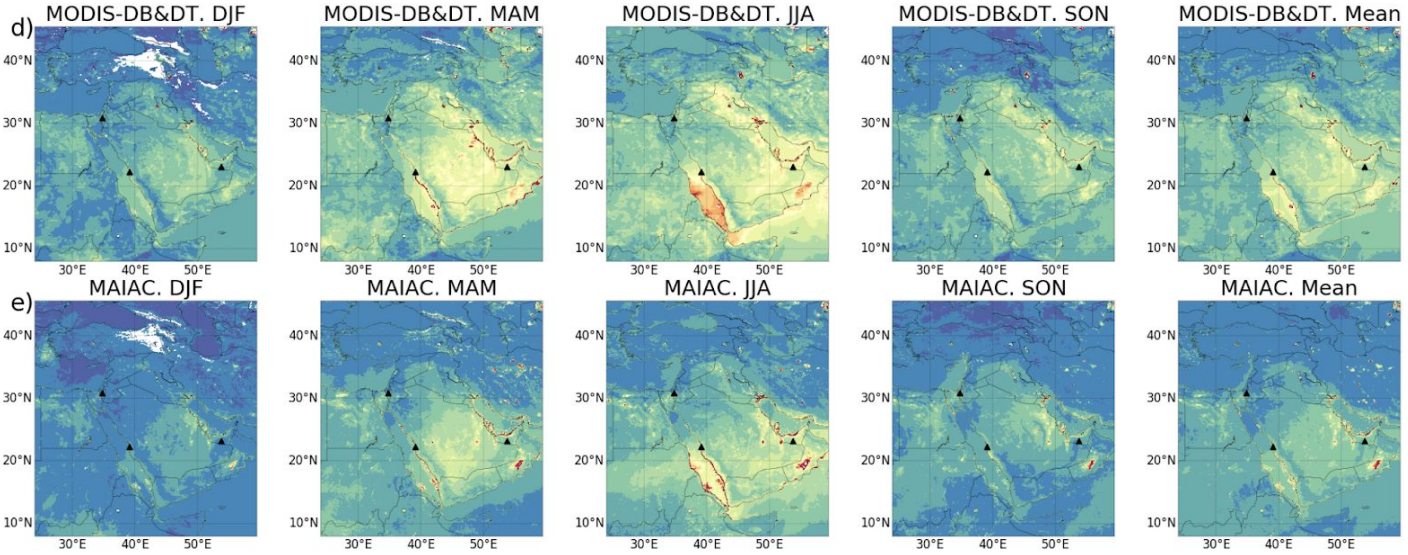


Figure E. MODIS 10x10km^2 and MAIAC (on its original resolution). Fixed error in calculation of seasonal means.



37. Section 5.3: Please provide references from where the formulas for calculation of PM2.5 and PM10 are adapted. What is the rationale behind the choice of the coefficients used in equation 4 and 5?

These formulas are taken from the WRF-Chem source code and Copernicus knowledge base (https://confluence.ecmwf.int/display/CUSF/PM10+and+PM25+global+products). The coefficients in those formulas account for the contributions of dust and sea salt bins to PM2.5 and PM10. Dust and sea salt have different bin sizes in WRF-Chem and CAMS-OA; therefore, those coefficients are different for WF-Chem and CAMS-OA. Both WRF-Chem and MERRA-2 use the GOCART aerosol module with the same bin sizes; therefore, the coefficients for WRF-Chem and MERRA-2 are the same.

38. Sections 5.3, 5.3.3, 5.3.4 and 5.4: Air Quality and Air pollution are very broad terms which also include trace gases in addition to the aerosol. Hence, the subtitles of these sections should be made more specific.

The reviewer is generally correct and we changed the titles 5.3, 5.3.3, 5.3.4 (there is now 5.4) to "*PM Air Pollution*", "*Spatial patterns of PM air-pollution*", "*PM air-pollution in the ME major cities*", correspondingly.

We have to mention here that according to the US EPA the air quality index is defined by the leading pollutant, which in the ME, almost exclusively is PM2.5.

39. Line 443: How does the calculated concentration of 298 µg/m3 compare against the measurements?

It is not possible to compare because there are no observations of total dust concentration, only PM2.5 and PM10 are available from MODON observations. Daily average PM10 surface concentration on 8 July 2016 registered by Jeddah AQMS is 184 µg/m3.

40. Figure 8 and 9: Please provide the uncertainty marks in the histogram which represent the variability over the mean.

Uncertainty marks are shown now in both figures. PM2.5 and PM10 measurement error is +/- 5%. This information was added to the text (Sec. 2.3).

41. Lines 496-503 Authors evaluate the PM2.5/PM10 ratio to evaluate the dominance of coarse/fine particles. A more quantitative evaluation would be PM10-PM2.5, which provides a more exclusive number for larger particles.

Both PM10-PM2.5 and PM2.5/PM10 are informative. The PM2.5/PM10 ratio is widely accepted in air pollution literature, e.g., see Gehrig et al., 2003; Parkhurst et al., 1999; Querol et al., 2001. Therefore we prefer to use this ratio in this study.

42. What are the major non-sulfate constituents in total PM2.5 non-dust aerosol?

In coastal areas it is sea salt and organic matter, over inland only organic matter. BC has a very little effect (see Table 6).

43. Lines 587-588: In addition to the AOD retrieval uncertainty, there are several other differences e.g. Spatial resolution, Quality assurance filter which contribute the observed difference.

The discussion is expanded to add more detailed explanation.

44. Line 600: Please use the same convention for the naming of seasons. “Fall” season is nowhere discussed in the text and appears for the first time in the conclusion.

Changed to autumn.

45. Line 609: Air quality should be replaced with PM air quality.

According to the US EPA the air quality is evaluated based on the concentration of the most significant leading pollutant, which are PM2.5 and PM10 in the Middle East, so PM air quality and air quality terms are almost equivalent in the ME.

References:

[1]. Liu, N., et al. (2019), Evaluation and comparison of multiangle implementation of the atmospheric correction algorithm, Dark Target, and Deep Blue aerosol products over China, Atmos. Chem. Phys., 19(12), 8243-8268, doi:10.5194/acp-19-8243-2019.

[2]. Mhawish, A., et al. (2019), Comparison and evaluation of MODIS Multiangle Implementation of Atmospheric Correction (MAIAC) aerosol product over South Asia, Remote Sensing of Environment, 224, 12-28, doi:10.1016/j.rse.2019.01.033.

[3]. Lyapustin, A., et al. (2018), MODIS Collection 6 MAIAC algorithm, Atmos. Meas. Tech., 11(10), 5741-5765, doi:10.5194/amt-11-5741-2018.

References used in the response:

Anisimov A, Axisa D, Kucera PA, Mostamandi S, Stenchikov G. Observations and cloud-resolving modeling of Haboob dust storms over the Arabian Peninsula. Journal of Geophysical Research: Atmospheres. 2018 Nov 16;123(21):12-47.

Gehrig R, Buchmann B. Characterising seasonal variations and spatial distribution of ambient PM10 and PM2. 5 concentrations based on long-term Swiss monitoring data. Atmospheric Environment. 2003 Jun 1;37(19):2571-80.

Ginoux P, Chin M, Tegen I, Prospero JM, Holben B, Dubovik O, Lin SJ. Sources and distributions of dust aerosols simulated with the GOCART model. Journal of Geophysical Research: Atmospheres. 2001 Sep 16;106(D17):20255-73.

Kalenderski S, Stenchikov G, Zhao C. Modeling a typical winter-time dust event over the Arabian Peninsula and the Red Sea. Atmospheric Chemistry & Physics Discussions. 2012 Oct 1;12(10).

Querol X, Alastuey A, Rodriguez S, Plana F, Ruiz CR, Cots N, Massagué G, Puig O. PM10 and PM2. 5 source apportionment in the Barcelona Metropolitan area, Catalonia, Spain. Atmospheric Environment. 2001 Dec 1;35(36):6407-19.

Prakash PJ, Stenchikov G, Kalenderski S, Osipov S, Bangalath H. The impact of dust storms on the Arabian Peninsula and the Red Sea. Atmospheric Chemistry & Physics Discussions. 2014 Dec 2;14(13).

Parkhurst WJ, Tanner RL, Weatherford FP, Valente RJ, Meagher JF. Historic PM2. 5/PM10 concentrations in the southeastern United States—Potential implications of the revised particulate matter standard. Journal of the Air & Waste Management Association. 1999 Sep 1;49(9):1060-7.

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 SO2 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.