To begin, the authors would like to thank the reviewer for their time, attention to detail, and insights on the paper and research. Each comment will be addressed point by point. The * next to line numbers indicates that it is referencing the tracked-changes manuscript. The * next to figure numbers references the supplemental figures in this response to the reviewer and not the original manuscript.

Specific Comments:

l.175: What is the meaning of 'coupled'? Probably only the use of the mineral dust emissions module (but not the transport, mizing, deposition etc.). Please better explain.

Here, the term "coupled" indicates that the meteorology (WRF) and the aerosol module (GOCART) are combined in the model in a way that they can directly impact each other. This is not just the meteorology and land surface part of the code being connected to dust emissions (e.g. wind speed, soil moisture, etc.), but also dust transportation via advection, convection, and turbulent mixing, as well as dry / wet deposition, and aerosol radiation effects. The rates for all of these dust processes are inextricably linked to the meteorology, and are treated such in the code via the direct coupling of WRF to the GOCART model. Additional clarification of "coupled" was added to this paragraph for readers:

> Ln 154-156* [The model is coupled to the Goddard Chemistry Aerosol Radiation and Transport (GOCART) module (Ginoux et al., 2001), which allows for feedbacks between the meteorology and aerosols and is described in more detail in Sect. 2.2.]
>
> Ln 184-186* […it is then transported based on the simulated meteorological fields from WRF, including advection, convection, and turbulent mixing…]

l.177 More details are needed about the schemes used. The paper is a sensitivity study about these schemes and they are not explained. In particular, the way to treat the aerosol for the indirect effects is completely different (the Grell scheme is aerosol aware compared to the others).

There is a paragraph later in the manuscript that points out the major differences between the cumulus schemes tested (Ln 230-240*). The Grell aerosol-aware scheme mentioned by the reviewer is the Grell–Freitas Ensemble Scheme (Grell & Freitas, 2014), whereas the one tested here is the non-aerosol aware version referred to as the Grell 3D Ensemble Scheme (Grell 1993; Grell & Devenyi, 2002). The aerosol aware scheme was not tested because it depends on the modelled aerosol number concentration affecting the CCN number. However, the GOCART aerosol module is a single-moment in mass scheme, which means it carries no number information and cannot alter the CCN number. As such, the GOCART model wouldn't have an effect on the aerosol aware version.

The overarching point of this study is that resolution matters more than the choice of convective parameterization. Thus, the point of including several cumulus parameterization schemes rather than just one was to represent the uncertainty across a spread of different available options in the model, and not to attribute why one scheme produces one solution or another. That is why at no point in the paper are the cumulus parameterizations directly compared to each other. Rather, they are represented as an ensemble mean with uncertainty estimates. Comparing the detailed responses of individual schemes to each other is outside the scope of this paper, but absolutely warrants further study and could be an entire manuscript on its own merit.

l.178: for a mineral dust study "no chemistry" why not. But no initial and boundary conditions for a simulation of 3 days, it is not possible to have realistic results

The only aerosol the authors were interested in for this study was dust. Furthermore, the aerosol burden over the deserts in the Arabian Peninsula is dominated by mineral dust (Heald et al. 2014), and as such, the authors decided that the full atmospheric chemistry code in WRF-Chem (e.g. gas phase and aqueous chemistry, etc.) was not needed and that other aerosol species were outside the scope of this study.

Two additional test cases were performed to address the reviewer's comment relating to the use of initial and lateral boundary conditions. First, a 3 km BMJ simulation was run using both initial conditions (ICs) and boundary conditions (BCs) for dust from the Community Atmosphere Model with Chemistry (CAM-Chem) global model, the output of which can be used for initializing the aerosol and chemistry fields in the mesoscale WRF-Chem model. Second, a 3 km BMJ simulation was run with only the lateral boundary conditions from CAM-Chem.

In the attached supplementary figures (denoted by a * to differentiate them from the figures in the manuscript), it can be seen that these two test cases (labeled as "BMJ-bcs and ics" and "BMJ-bcs only") have very little effect on the dust uplift potential (Fig. 4*), the threshold velocity, surface settling flux, bowen ratio (Fig. 5*), or the mean vertical velocity (Fig. 8A*). This is expected, since all of these fields are dominated by the meteorology and not the dust concentrations in the local environment.

Furthermore, the second test case "BMJ-bcs only," in which only lateral BCs were used to represent dusty air moving across the domain from places like the Sahara, has essentially no effect on the results and is in line with the conclusions from the manuscript where no lateral BCs were used. Including the BCs has minimal influence on integrated dust (Fig. 6A*), vertical dust concentrations (Fig. 7A*), dust fluxes (Fig. 8C*), or dust radiative effects (Fig. 11*). There are two possible interpretations of this result. One is that the dust concentrations being transported laterally into the domain are small for this case study. A second possibility is that the CAM-Chem model underestimates this dust source.

Conversely, including dust ICs ("BMJ-bcs and ics") does have an effect on the dust concentrations in the domain. Looking at the integrated dust plot (Fig. 6A*), there is a decreasing trend starting from the initial timestep throughout the rest of the simulation. The decreasing trend is not seen in the AOD observations from AERONET in the region (Fig 12*). This points to the ICs being significantly higher than what the model produces on its own, and they are out of sync with the equilibrium model solution. Furthermore, comparing modeled AOD to the AERONET stations, there is little added benefit in using the ICs to get better agreement with observations. The model underpredicts dust throughout the simulation compared to observations, which is a finding in Saleeby et al. 2019, where this particular case study simulation is further compared to observations. It would most likely be better to adjust the dust tuning parameter (C) in Eq. 1 than to use ICs that are not in tune across modeling platforms or in keeping with the observations. The result of the added dust in the IC run is higher integrated dust (Fig. 6A*), vertical dust (Fig. 7A*), dust flux (Fig. 8C*), and a stronger radiative effect. While including dust ICs increases the dust load, it does not however change the conclusions of the study.

Saleeby, S. M., van den Heever, S. C., Bukowski, J., Walker, A. L., Solbrig, J. E., Atwood, S. A., Bian, Q., Kreidenweis, S. M., Wang, Y., Wang, J., and Miller, S. D.: The influence of simulated surface dust lofting and atmospheric loading on radiative forcing, Atmos. Chem. Phys., 19, 10279–10301, https://doi.org/10.5194/acp-19-10279-2019, 2019.

l.182 "kept constant" meaning remain the same during the whole simulation?

Correct, and that these physics options don't change across the simulations – the language has been changed in the manuscript to remove any confusion:

> Ln 162* [The following model parameterizations were employed and kept constant across the simulations..]

l.214: The 'Dust Uplift Potential' is a calculation already done in a large majority of dust emissions schemes, by principle of the mechanism to evaluate. Unfortunately, it represents only a small part of the problem and is not really useful. It describes only the link between the friction velocity treshold (using the aeolian roughness length) and the current friction velocity. But other important parameters are not taken into account: the vegetation, the erodibility, the soil humidity, the recent precipitation etc. In addition, the fact to use a constant Ut is not realistic (eq.3): the aeolian roughness length is far to be constant over erodible region. It is the most important varying parameter in mineral dust emissions modelling. The use of three different kind of DUP has a large interest. The message is already contained in one. If the authors really want to use this criteria, only one is enough.

The authors are not sure exactly as to what is being asked here by the reviewer. However, we have done our best to address the questions here as we understand them and hope this will address the reviewer's concern.

Several of the parameters listed here are contained in the varying DUP equations, including the erodibility (Eq. 5 with the variable S) and the soil moisture / recent precipitation (Eq. 4 and 5 depend on U_t - the only varying parameter in Eq. 2 for U_t is soil wetness, w_{soil}). The point of including these different DUP parameters is to tease out which of these processes is the most important without assuming that one is more important than the other for this case study. It has been shown previously in the literature that the soil moisture and erodibility are important for dust uplift (i.e. Gherboudj et al. 2015) in addition to wind speed, which means that using only one parameter doesn't tell the whole story.

Eq. 3, which is the most simplistic of the equations and assumes a constant roughness length, and has been used widely in the literature, especially in offline model dust approximations. To compare our results with that of other studies, it is necessary that we use Eq. 3. However, we point out its limitations and have included the more complicated DUP parameters we think are more useful for this study:

> Ln 205-207* [This simplified equation for dust uplift has been used in previous dust studies, and is useful to include here to place the findings of this manuscript in the context of existing literature.]

l.239: The reference simulation has an horizontal resolution of 3km to enable explicit convection calculation. This simulation has boundary conditions and this is a good point. But these boundary conditions are from the BMJ simulation, i.e one of the studied case. Thus, we can think that the reference case will be very influenced by this case, no? To have a more realistic comparison between scheme, the 'reference' has to be done for each scheme and a first spread can be calculated between all 'high resolution' cases.

This is a good point – especially since there are two competing classes of cumulus parameterizations tested here: BMJ is the only moisture / temperature adjustment scheme, whereas the others are mass flux schemes. To test the sensitivity of the results to which cumulus parameterization scheme is employed in the parent nest, a second 3 km simulation was run. In this test, the Kain-Fritsch cumulus scheme serves as the 15 km parent, which is then nested to 3 km (this run is labeled as "3 km – KF" throughout the supplementary figures) to represent the mass-flux schemes. In none of the figures (Fig. 4-11*) is this difference significant or does it change the conclusions of this paper. Again, because model resolution dominates over the choice of cumulus parameterization, this effect of using a different parameterization in the outer nest has little effect on the results.

> Ln 227-228* [Other combinations of nests were tested, but the results were not sensitive to which 15 km simulation was used as the parent nest, or which lateral boundary conditions, for the 3 km simulation.]

l.263: for long-range transport, 24h of spin-up is not enough. For the time averaged results, it is only th elast two days. But for the time series, it is the 3 days? why this difference?

For the long-range transport part of this comment – see the response and tests from comment l.178. Including the lateral boundary conditions that represent long-range dust transport have little impact on the results and hence we feel that 24h of spin-up is sufficient.

For the time averaged results, we did not want to include the 24-hour model spin up time. However, they were included in the time series to show how the model approaches its equilibrium solution when starting from no dust sources.

l.272: why not use directly the mineral dust emissions fluxes? Please explain this important point.

The emission fluxes convey the same story as the approach utilized here (see Fig. 13*). However, the difference in the magnitude across the simulations is difficult to see in this plot, and it wasn't included. The conclusions are the same from a dust emission standpoint versus a dust concentration perspective. Additionally, the variables that go into the emission flux formula in GOCART (Eq. 1) are very similar to the Dust Uplift Potential (DUP) calculations. The most complete DUP parameter (Eq. 5) includes the same variables as the emission flux, so it would be redundant to include both the DUP calculations and the emission fluxes.

l.280: why the simulation with the coarsest resolution (and not simulation) overestimates the wind speed? Please explain (and I imagine it is the "10-m wind speed", please correct).

Correct – it is the 10-m wind speed. This has been updated in the manuscript:

> Ln 297-298* [The coarsest simulation overestimates the near-surface wind speeds related to the NLLJ mechanism, which...]

There are a few theories regarding why the coarsest simulation would overestimate the near-surface wind speed. Marsham et al. (2011) noted that in their simulations over Northern Africa, the Saharan Heat Low was more pronounced in the coarse simulations. They postulated that cold pool venting in the explicit simulations reduced this thermal low, thereby reducing the horizontal pressure gradients which are responsible for low-level jets in this region. It follows then that the low-level jet is weaker in this scenario, as is the process of mixing of the jet to the surface, and this the near-surface wind speeds. In the Arabian Peninsula case study, this mechanism is certainly quite possible. However, this theory has yet to be tested and is outside the scope of this paper.

l.293: Yes, it is right. And obvious. Of course, a key point in modelling is to try to have a model not sensitive to the spatial resolution. And it seems it is the problem with WRF-chem. In WRF, the principle is to use, for each grid cell, the dominant soil type and landuse. Thus, by principle, the result is very sensitive to the resolution. Some other models are using subgrid scale variability and Weibull distribution for the 10-m wind speed, for example, to avoid this problem. Please see bibliography and replace WRF-chem in the context of all currently used regional dust models.

WRF-Chem is just one of many regional models that can be used operationally and / or in research applications. For instance, The Sand and Dust Storm Warning Advisory and Assessment System (SDS-WAS) includes 12 dust models, with more undoubtedly available to be used in research applications. Each model is unique, and most likely has several options for their cumulus parameterizations as well as other physical representations of meteorological processes. Combining the differences between dust models in this way is a very large undertaking the likes of which are being conducted by organized working groups like the International Cooperative for Aerosol Prediction (ICAP) and is outside the capabilities of a single manuscript.

l.331: it is not sure that there is an interest to have a conclusion such as "resolution increases or decreases the mineral dust emission fluxes". In fact it depends on the studied area, the variability of the orography, aeolian roughness length, soil humidity, vegetation. And, of course, the way to well take into account or not all these processes and their variability.

The reviewer makes an excellent point here. The manuscript has been updated to include more about the uncertainty here:

> Ln 355-360* […dust emissions and airborne dust mass increases in the WRF-Chem simulations in the convection-allowing simulation, which is in closer agreement to the studies of Reinfried et al. (2009) and Bouet et al. (2012) who used COSMO-MUSCAT and RAMS-DPM respectively. Considering each study used a different model and therefore physics, it is unsurprising that the results vary. However, it is not apparent how much of a role the region or specific case study plays in this difference and is an area for future work.]

l.335: I don't understand the discussion with "The rates of gravitational settling are higher in the explicit simulation compared to the coarse simulations, yet Fig. 6.a suggests that this is not enough to offset the higher dust emissions, or the integrated dust quantities would be similar across all the simulations." The dry deposition is proportional to the concentrations, being a velocity applied to the concentrations. How is it possible to have 'enough' settling to 'offset' the higher dust emissions?

If there is more dust aloft, more dust eventually needs to settle out. The point that we are trying to make here is that the missing piece in this process is the higher vertical transport. If dust was transported to the same height, the gravitational settling would offset the higher emissions and there would be no reason for the integrated dust values to be higher. This part of the manuscript has been edited for clarity.

> Ln 364-367* [The rates of gravitational settling are higher in the convection-permitting simulation compared to the coarse simulations because more dust is available aloft to settle out. Nevertheless, Fig. 6.a suggests that this increase in gravitational settling rates in the 3 km case is not enough to offset the higher dust emissions…]

Figure 6: the fact to have difference sbetween resolution is understandable but a factor 2 has to be better explained. Mineral dust emissions mass maps for the common domain (the one with 3km horizontal resolution). The caption is not easy to understand: "Domain averaged integrated dust mass". Please correct with Spatially averaged, vertically integrated.

The difference between resolutions in Figure 6 differ by a factor of 1.5, which we discuss in the previous section with Figure 4 and Figure 5. Using DUP(U,Ut,S) we see that the 3 km has the most potential to loft dust, especially on 04-Aug when there is a convective maximum. This is related to the threshold velocity being lower and soil wetness (Figure 5) and is also explained with the differences in vertical transport, which is covered in the next section of the manuscript. More about the differences in precipitation in convection-allowing versus parameterized simulations affecting soil moisture and the threshold velocity has been included in the text:

> Ln 308-312* [Rainfall is generated differently in parameterized versus convection-allowing simulations, and it has been well documented that parameterized simulations produce more widespread light rainfall, whereas more intense rainfall tends to develop over smaller areas in convection-allowing simulations (e.g. Sun et al., 2006; Stephens et al., 2010). From a domain average perspective, rainfall in the 3 km simulation will cover less area, leading to the soil moisture threshold not being exceeded as frequently compared to the parameterized cases.]

These figures have been updated for clarity and the captions have been changed. Throughout the manuscript anytime there is a reference to “domain averaged integrated dust” it has been changed to the phrase “spatially averaged, vertically integrated.”

> Ln 911* [Figure 6: Spatially averaged, vertically integrated dust mass. Colors and shading are identical to that in previous figures.]

l.346: "the vertical dust profile follows a generally exponentially decreasing function" is it a conclusion of this study? or coming from a reference? These is no reason to have an exponential decrease in the troposphere. Many cases of thin but concentrated dust plumes transports are observed and modelled...

On average, exponentially decreasing aerosol in the troposphere is a good assumption (e.g. Gras 1991; Tomasi, 1982). This type of idealized profile is often assumed for CCN in models (e.g. Fan et al., 2007). You are correct in that individual plumes will change this profile, but here we are looking at a domain average, which regresses to the exponentially decreasing function.

l.369: "The implications for dust transport based on vertical velocities is convoluted." This sentence is difficult to understand.

This part has been further explained in the text to avoid confusion:

> Ln 403-404* [The implication for dust transport based on vertical velocities is convoluted, since updrafts and downdrafts work concurrently to redistribute aerosol.]

l.421: The impact on radiation, with potential heating and cooling, is a process needing more than 2 days of simulation to be significative.

The timescales of interest vary depending on which specific processes are being examined. From a climate perspective, two days is much too short. However, looking at static stability in the lower atmosphere from a mesoscale perspective, including processes like convective initiation or the formation and deterioration of the nocturnal low-level jet, the timescales examined here (or in some cases even shorter timescales) are important and significant.

l.428: there is a sign change. Could you explain why?

The model applies a higher weight (via the refractive index for mineral dust) to dust scattering in the shortwave and cooling compared to the longwave absorption. With more dust in the explicit case, the shortwave effect is amplified.

More explanation has been added to this section for clarity:

> Ln 468-470* [The model has a stronger shortwave effect for dust based on the prescribed index of refraction, but is also related to the timing of dust emissions, considering the SW effect is only active during the daytime.]

General Comments

1. There is no data used in this work: the simulations are compared between them but we have no idea of the realism of the simulations (there is only one reference for a comparison to Aeronet AOD in another paper, under discussion, and no guarantee this is exactly the same model set-up, and which one?). At least, the reference case (dx=3km) should be compared to available data (surface networks such as MIDAS, AERONET, satellite, other data).

The Saleeby et al. 2019 study (referenced above and in the paper) where the 3 km simulation was compared more thoroughly to observations has been published (once again provide the full reference here). In that paper, the exact same model setup was used for the 3 km simulation as was used here, and this point has been added to the manuscript. We have included comparison to the few AERONET sites in this region in the supplementary Fig. 12*, and found similar results (regardless of including or excluding the ICs and BCs in the simulations) with Saleeby et al. (2019) in that WRF-Chem under predicts AOD. However, the model must assume a refractive index for dust to calculate AOD, which may or may not be realistic in itself. Additionally, we have selected dust as the only aerosol present in the model, while in reality there are other aerosol types that may be contributing to the AOD. Thus, making one-to-one comparisons here with observations is difficult. None of the continuous observational networks provide dust concentration, which is what is actually needed for a true validation. Nevertheless, if WRF-Chem is underpredicting dust concentrations, this doesn't change the conclusions of the study.

2. The studied case extended from 2 to 5 August 2016: there is no spin-up time, important when studying transport of aerosol such as mineral dust. Time series are presented for the three days, but some average are done only for the last two days, explaing that the first day is spin-up. But, viewing the domain size, the minimum spinup time should be at least one week.

See response to comment on l.178 above.

3. There is no boundary or initial conditions. These missing background values may have a large impact on the results, in particular knowing that the model couples the meteorology and the aerosol concentrations: direct and indirect aerosol effect may be long-term and it is required to have correct boundary conditions to have realistic effect of aerosol on meteorology. For the 'reference' domain, the boundary conditions are extracted from one of the studied case, biasing the results.

See response to comment on l.178 above.

4. The convection schemes used are not explained. The paper is a sensitivity study about these schemes but there is no explanations about their real differences, how they take into account aerosol or not, thus no conclusion about why results may be different depending on the scheme.

See response to comment l.177 above.

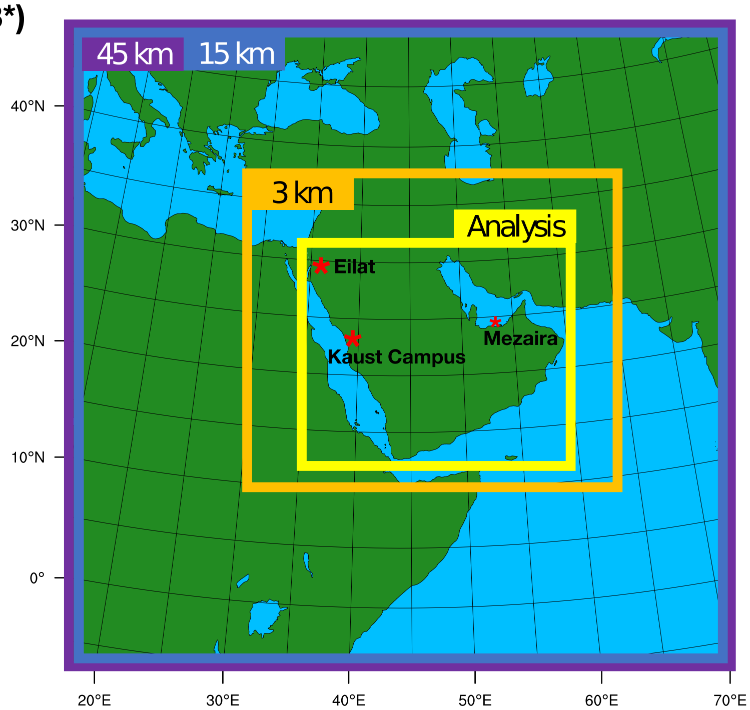
5. The paper deals with the sensitivity to the model resolution. But since the schemes are not well implemented (no wind speed distribution, no subgrid scale variability), there is a large sensitivity but not for realistic and physical reasons: the differences are not due to the convection schemes in general but just to the fact that the problem of the resolution is not well designed in this model: it is not possible to describe a threshold problem (such as mineral dust emissions) without taken into account disstributions of input parameters. Results are linked to this model only and are not useful for other modellers

Regardless of how successfully or unsuccessfully these schemes have been implemented into WRF-Chem, it is still a very widely used model for air quality, atmospheric chemistry, and more relevant for our manuscript - dust research and forecasting. A list of some of the current forecasting centers using WRF-Chem can be found on the WRF-Chem users page (https://ruc.noaa.gov/wrf/wrf-chem/Real_time_forecasts.htm) and is one of the dust models included and evaluated in the SDS-WAS real-time forecasts.

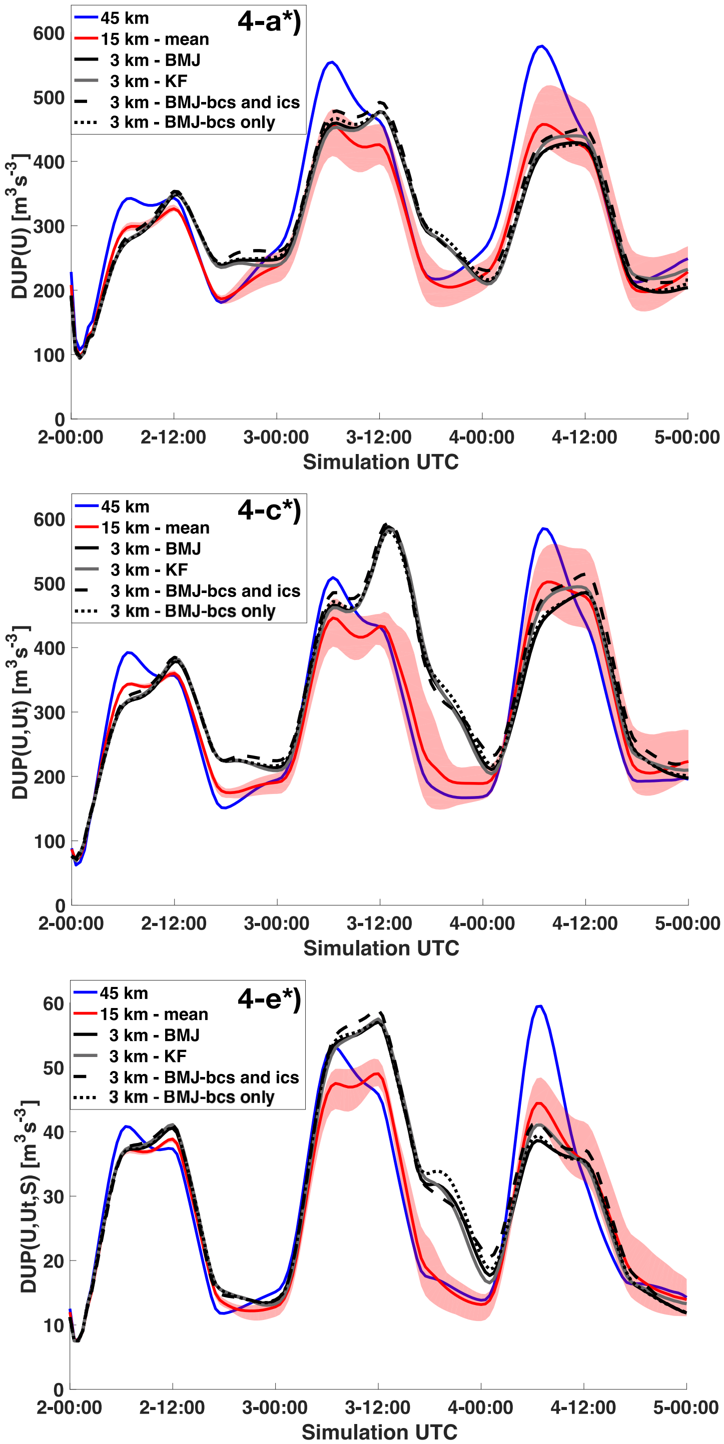
WRF-Chem users need to be aware of its limitations and its sensitivity to resolution when designing numerical experiments, and readers should be cognizant of this when interpreting results from both past and future studies that use this model. Furthermore, some of the results we found here are similar to other studies that have used different regional models, such as Reinfried et al. (2009), while other manuscripts are in disagreement with our findings, such as Heinhold et al. (2013) and Marsham et al. (2011). Clearly, we have not reached a consensus and more work is needed. Between the user base for the WRF-Chem model and the spread in results between our findings and previous literature, there is a broader community of interest for this paper.

Supplementary Figures:

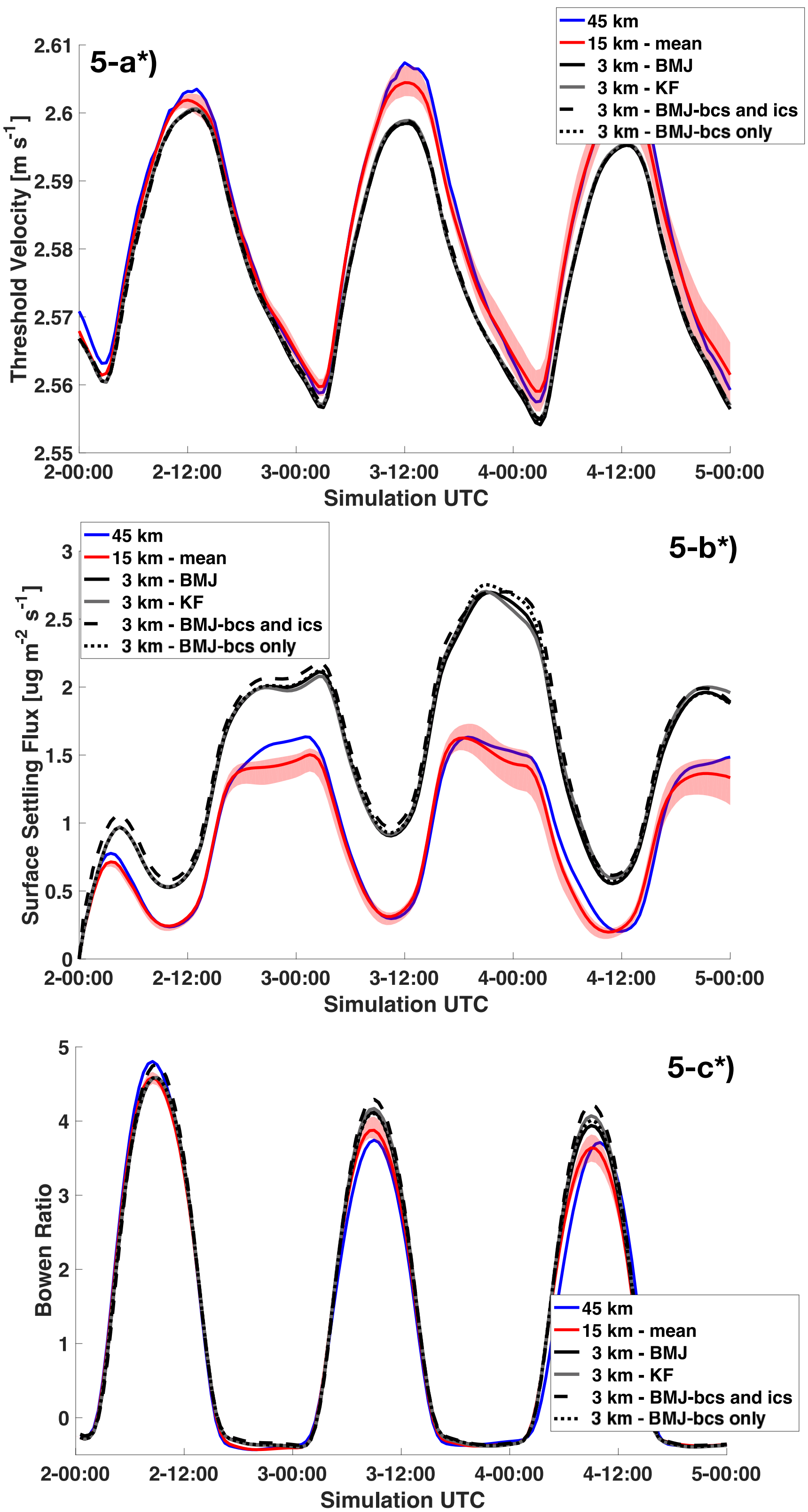
3*)



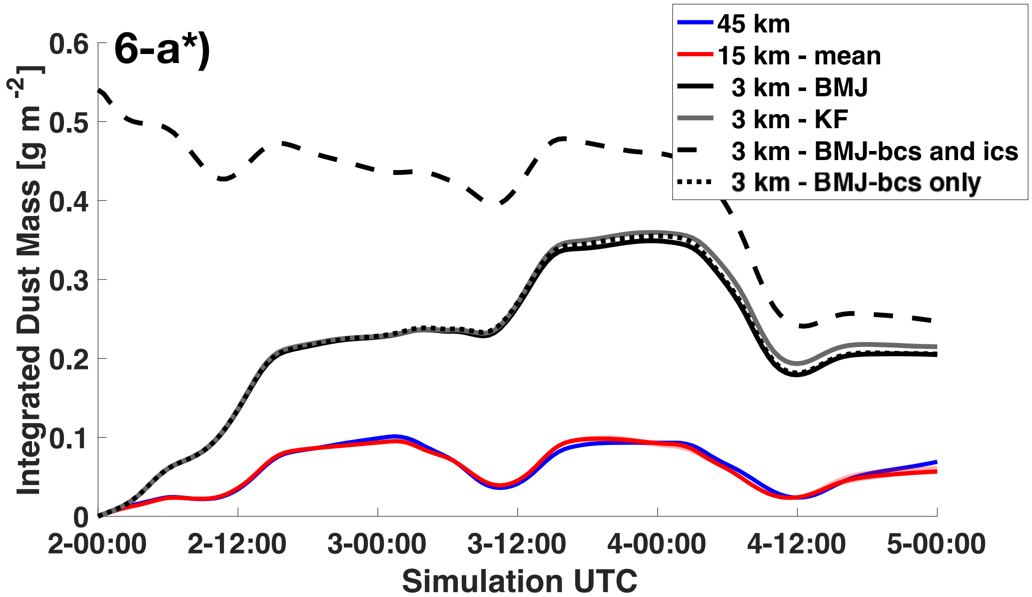
Supplementary Figure 3) Same as in Fig. 3 in the manuscript, but the location of the 3 AERONET sites in the analysis have been added.



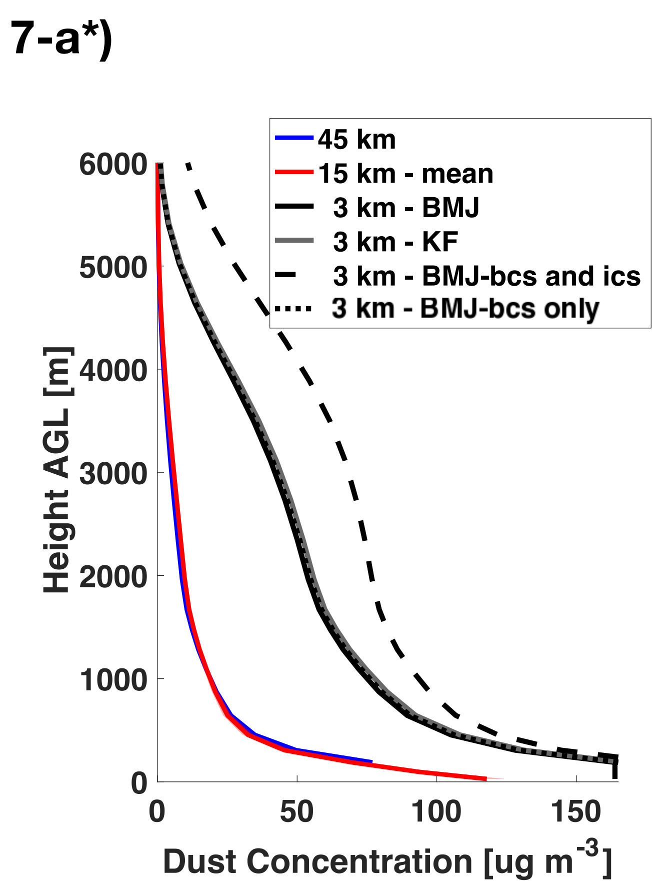
Supplementary Figure 4A, 4C, 4E) Same as panels A, C, and E in Fig. 4 in the draft, but with 3 additional test cases: a 3 km inner nest which used the KF cumulus parameterization in its outer nest for initialization (gray solid line), a 3 km simulation with the BMJ cumulus parameterization with both initial and lateral boundary conditions for dust from the Community Atmosphere Model with Chemistry (CAM-chem) global model (black dashed line), and a 3km BMJ simulation with only the lateral boundary conditions for dust (dotted black line).



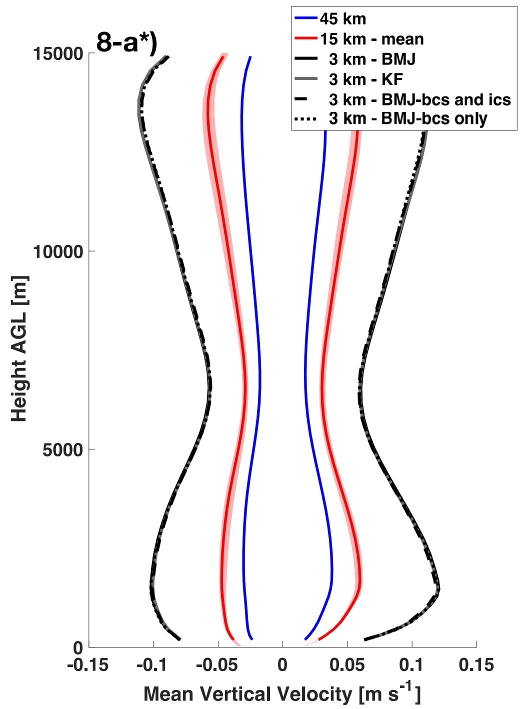
Supplementary Figure 5) Same as Fig. 5 in the draft, but with the 3 additional test cases as in Supplementary Figure 4.



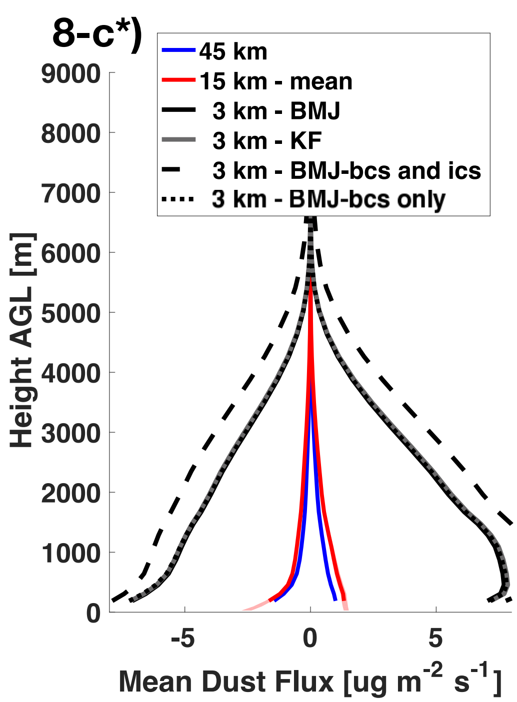
Supplementary Figure 6A) Same as panel A in Fig. 6 in the draft, but with the 3 additional test cases as in Supplementary Figure 4.



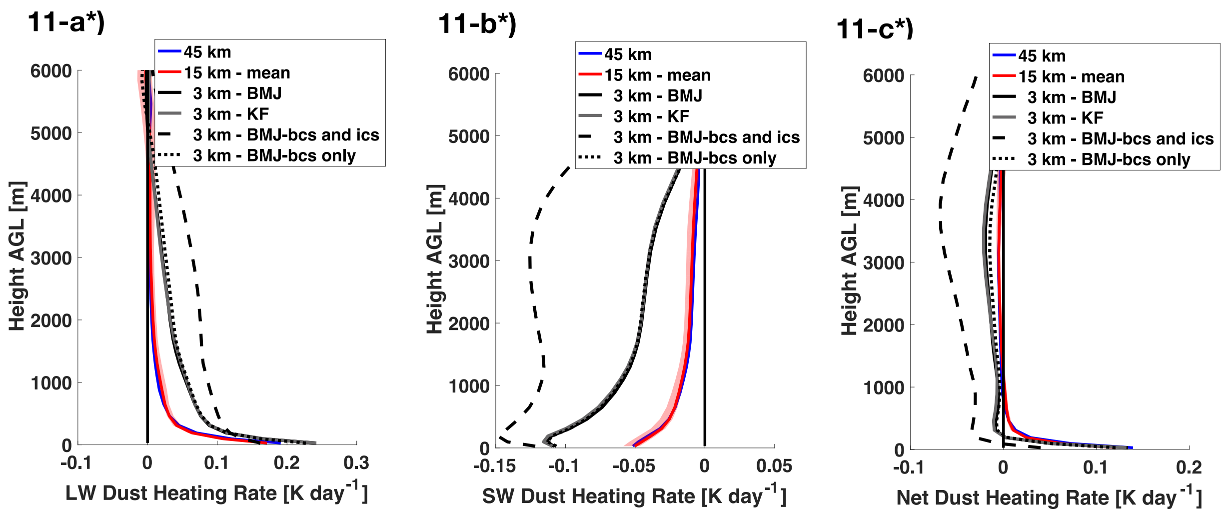
Supplementary Figure 7A) Same as panel A in Fig. 7 in the draft, but with the 3 additional test cases as in Supplementary Figure 4.



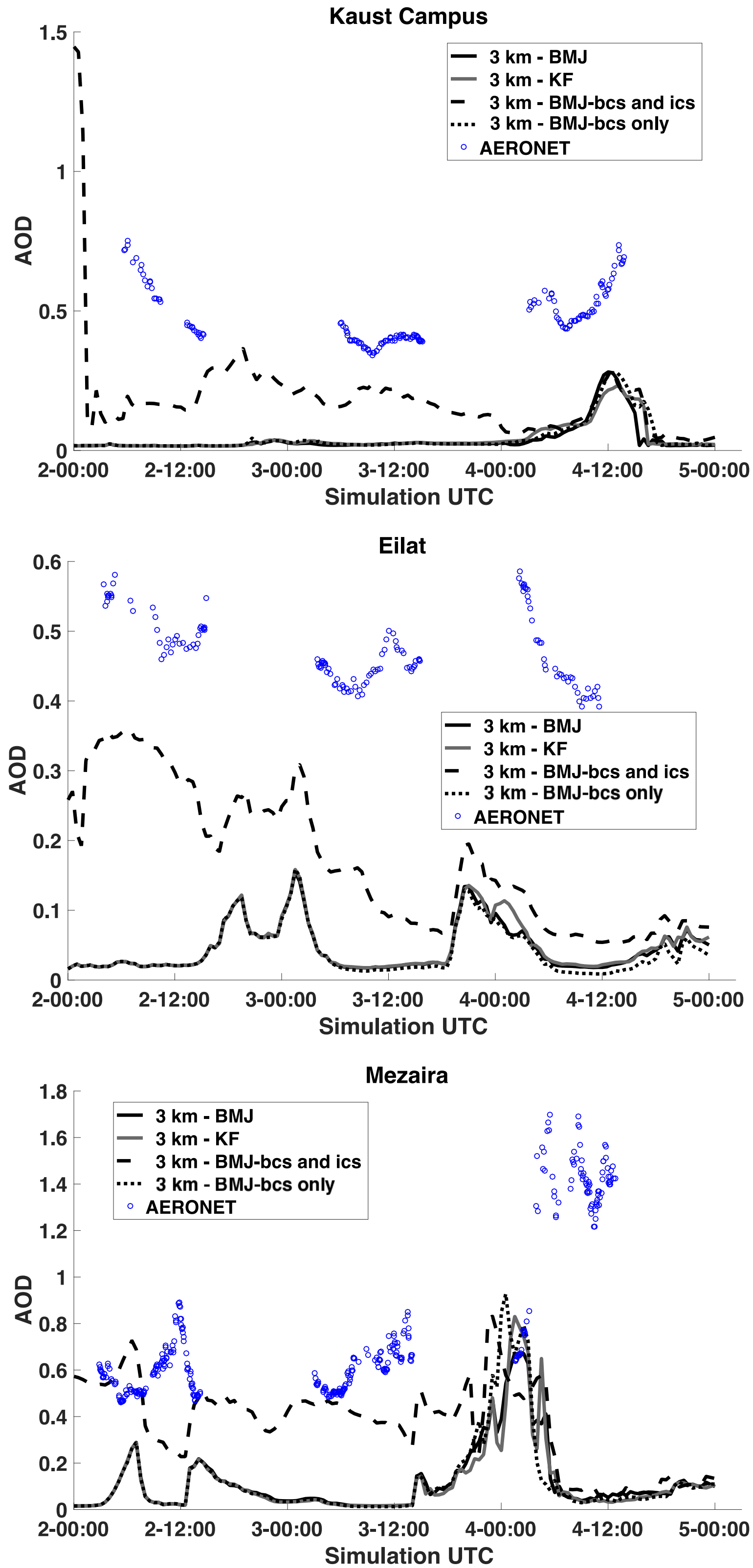
Supplementary Figure 8A) Same as Fig. 8A in the draft, but with the 3 additional test cases as in Supplementary Figure 4.



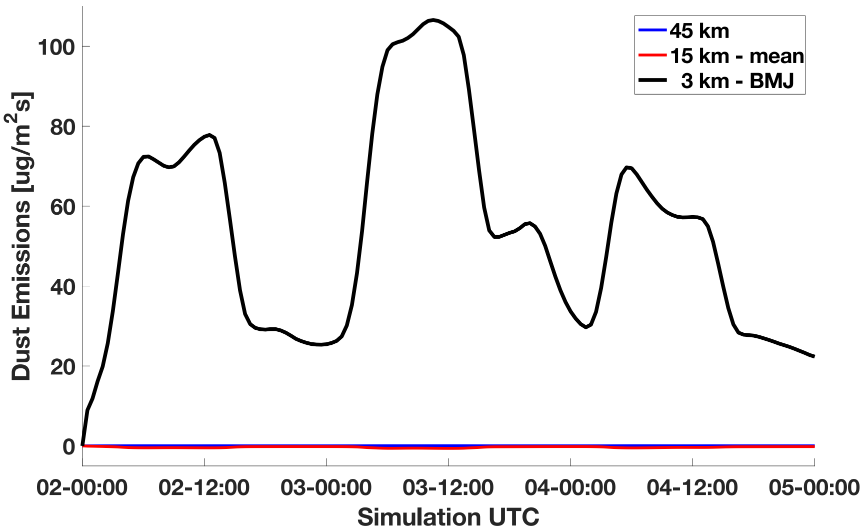
Supplementary Figure 8C) Same as Fig. 8C in the draft, but with the 3 additional test cases as in Supplementary Figure 4.



Supplementary Figure 11) Same as Fig. 11 in the draft, but with the 3 additional test cases as in Supplementary Figure 4.



Supplementary Figure 12) Comparison of the 3 km simulations modeled AOD with AERONET AOD for 3 different observational sites.



Supplementary Figure 13) Time series of dust emissions from the surface to the atmosphere.

Additional Citations

Emmons, L. K., Walters, S., Hess, P. G., Lamarque, J.-F., Pfister, G. G., Fillmore, D., Granier, C., Guenther, A., Kinnison, D., Laepple, T., Orlando, J., Tie, X., Tyndall, G., Wiedinmyer, C., Baughcum, S. L., and Kloster, S.: Description and evaluation of the Model for Ozone and Related chemical Tracers, version 4 (MOZART-4), Geosci. Model Dev., 3, 43-67, https://doi.org/10.5194/gmd-3-43-2010, 2010. *We acknowledge use of NCAR/ACOM CAM-chem global model output available at* https://www.acom.ucar.edu/cam-chem/cam-chem.shtml.

Gherboudj, I.,S. N. Beegum, B. Marticorena, and H. Ghedira: Dust emission parameterization scheme over the MENA region: Sensitivity analysis to soil moisture and soil texture, J. Geophys. Res. Atmos., 120, 10,915– 10,938, doi:10.1002/2015JD023338, 2015.

Gras, J. L.: Southern hemisphere tropospheric aerosol microphysics, J. Geophys. Res., 96 (D3), 5345– 5356, doi:10.1029/89JD00429, 1991.

Tomasi, C.: Features of the Scale Height for Particulate Extinction in Hazy Atmospheres, Journal of Applied Meteorology, 21, 931-944, doi:10.1175/1520 0450(1982)021<0931:FOTSHF>2.0.CO;2, 1982.

Fan, J., R. Zhang, G. Li, W.-K. Tao, and X. Li: Simulations of cumulus clouds using a spectral microphysics cloud-resolving model, J. Geophys. Res., 112, D04201, doi:*10.1029/2006JD007688*, 2007.