Review Huneeus et al.: Forecasting the North African dust outbreak towards Europe in April 2011: A model intercomparison

General Comment:

The paper is comparing 5 state-of-the-art dust forecast models with respect to their capability to forecast a major dust event that took place in April 2011. The forecast time is 24, 48 and 72 hours and the authors compared and validated the model skills in terms of dust related parameters as well as meteorological parameters that led to the dust event. The full suite of available observational data was used within the known limits of their applicability for model comparison studies. They found that model choice matters more than lead time. They couldn't rule model deficiencies in representing the synoptic conditions, as opposed to a variety of known limitations with regard to the parameterisation of the dust cycle (emission scheme, size distribution, dry and wet deposition). Those limitations are causing considerable quantitative differences in dust AOD, emission flux, surface concentration and vertical dust distribution as a function of model choice.

While the reason for choosing the 5 selected models remains unclear, it is certainly a worthwhile exercise in light of the absence of such studies. Sure, it would be desirable to have a full-blown model inter-comparison exercise with all state-of-the-art models available; the paper provides a useful framework for future such studies. Ideally, the authors can provide the community with stringent guidelines as to how a quasi-operational model validation exercise should look like. For example, given that there already exists an operational forecast evaluation project within the SDS-WAS framework (http://sds-was.aemet.es/forecast-products/forecast-evaluation/model-evaluation-metrics), it seems fairly straight-forward to extend this effort beyond the current setup (perhaps introducing sub-regions to facilitate dust event evaluation). Binietoglou et al 2015 could be added in this context as well.

Equally desirable, yet beyond the scope of this study, would be an extension of this validation exercise to different types of dust events. In particular, it would be interesting to see whether there are systematic forecast model biases with regard to the breakdown of the low-level jet or is the forecast skill sufficient to predict convectively triggered haboobs with some lead time. Admittedly, the latter depends on the model resolution and might not work with the selected set of models (or at least not at the chosen horizontal model resolution) to start with, but it would be worth putting such suggestions for follow-up work in the discussion/conclusion section. Also, a method to quantify the impact of imprecise forecast of synoptic conditions upon the dust emission flux would help to detect the key aspects of future work. Based on my own work with the HadGEM3 model at 12x12km grid size, the surface winds are very well reproduced (compared with direct observations at 10m height) even when allowing for considerable lead time (unsurprisingly, the MetUM used in this study shows similarly good results for all lead times). This suggests that future work should focus on improving the emission schemes, which is something I wish the authors of this paper could confirm.

That said the study provides insights into the fidelity and skill of the models to represent observed synoptic conditions and resulting dust emissions and transport. Metrics for model evaluation have been chosen carefully including detailed discussion of the results. I therefore recommend publishing the paper after some minor revisions which I have detailed below.

Specific Comments:

Introduction

p.26666, lines 4/5: A short justification or explanation why those 5 (and only those 5) models have been chosen for the analysis would be desirable

Data and Models

p.26667, lines 5/6: The orange dots in Fig 1 are really hard to identify. I suggest to put all station information in a separate plot in order to facilitate identification

lines 17-26: MODIS AOD is also biased towards the time of satellite passage. Do you account for this potential source of error when you validate the model results? If not, how much of an on the results of the analysis could it have? Appropriate reference needed.

p.26668, section 2.3: I would suggest to introduce the MERRA reanalysis here as well (as you are using its wind data). Could be put into the model section as well. A short paragraph of known issues with reanalysis data in general and MERRA in particular should also be added. NCEP as well as ERA40/Interim reanalysis considerably overestimate nighttime wind speeds and underestimate higher wind speeds in general (e.g. Haustein et al 2012; Largeron et al. 2015; more to come soon from Engelstaedter et al. in Review)

Results

p.26672, section 4.1 and p.26673/74, section 4.2: What is the main reason that the MetUM overestimates the dust emission flux and the surface concentrations so consistently (a feature which is also apparent in the operational forecast)? Is the preferential source map (based on topography) switched on in their operational model setup? I recommend to add a paragraph in the discussion section that deals with this noticeable problem in this model. Ideally, it can be established what the likely cause for the overestimation is (e.g. strong tuning due to poor parameterisation of deposition). I note that the emission/deposition ratio is briefly mentioned at p.26680, lines 20-24. Perhaps this is where the discussion fits best.

p.26673, line 23: NNMB —> NMMB

p.26674, lines 5/6: Again, it would help to have a short discussion of the potential causes for the large range of model outcomes wrt emission flux in the corresponding section.

p.26676, lines 18/19: Are there any known issues with BSC-DREAM8b (e.g. with regard to the PBL or soil moisture scheme) that could be causing such discrepancies? Could be revisited in the discussion section

p.26678, lines 4/5: See earlier comment on MERRA uncertainties

Discussion

p. 26680, lines 18/19: Recent findings (e.g. from Allen et al. 2013; Ryder et al. 2013) suggest that larger dust particles can indeed be found in higher levels of the atmosphere, suggesting that omission of larger particles (or their treatment in terms of deposition, respectively) in models is a potential source of error.

p.26682, lines 23-27: Could go into the conclusions.

On a more general note, as alluded to in my general comments already, what would be most useful for the modelling community to have is a quantification of the impact imprecise capturing of synoptic conditions in general and surface wind speeds in particular would have upon the resulting model emission flux. Or in other words, we need an assessment which tells us what spatial model resolution is required to reproduce observed wind speeds (and wind gusts) good enough to exclude it as a major source of error when it comes to testing the performance of the individual components of the dust scheme in the model. I do think this study can already provide some clues in that regard (albeit not in a strictly quantified analytical sense) which is why I would appreciate a slightly more in depth discussion of this crucial subject. If the authors don't feel comfortable to go out on a limb on that, I would recommend to put it at least as a major short term research goal in the conclusion section in order to draw the readers attention to what appears to be the most pressing issue (in my humble opinion that is).

Conclusion

p.26683, lines 1-6: Repetition of what has already been said in the discussion section (—> delete)

The conclusions are generally a bit too repetitive wrt the previous discussion section. While I tend to structure things the same way myself, the conclusion section should focus more on the impact/repercussions of the findings/results which have been discussed before. For example, the topic of separating meteorological/synoptic and dust cycle parameterisation related problems would fit the conclusion section perfectly. This goes along with an outlook of follow up research of this particular paper and suggestions where future research on the subject should focus on in general. Therefore I recommend to overhaul (and shorten) the conclusion section as recommended. I am convinced that it can help to wrap up this otherwise very well written and well thought-out paper in a neat and concise fashion.

References:

Allen et al: Dust emission and transport mechanisms in the central Sahara: Fennec ground-based observations from Bordj Badji Mokhtar, June 2011. JGR Atmospheres, 118, 6212–6232, doi:10.1002/jgrd.50534, 2013

Binietoglou et al: A methodology for investigating dust model performance using synergistic EARLINET/AERONET dust concentration retrievals. AMT, 8, 3577-3600, 2015

Haustein et al: Atmospheric dust modeling from meso to global scales with the online NMMB/BSC-Dust model – Part 2: Experimental campaigns in Northern Africa. ACP, 12, 2933–2958, 2012

Largeron et al: Can we use surface wind fields from meteorological reanalyses for Sahelian dust emission simulations? GRL, 42, doi:10.1002/2014GL062938

Ryder et al: Optical properties of Saharan dust aerosol and contribution from the coarse mode as measured during the Fennec 2011 aircraft campaign. ACP, 13, 303–325, 2013