Anonymous Referee #1:

Major comments:

1. *Section 2.1: The authors concentrate on a narrow (100km) coastal plain with complex circulation (e.g. sea breezes) but use a model resolution that is too coarse (10km). They use the same resolution also when testing varying emission resolutions. Even if this is due to computational constraints, surely short tests of specific dust events can be conducted at higher resolution to properly test dust dispersion. Can the authors identify such events and complement the paper with higher resolution model runs?*

In this study, we focus on the effect of the spatial resolution of the surface characteristics on the dust generation. It is especially important over the coastal plain where the surface is heterogeneous. We fixed the meteorological forcing to isolate the impact from high-resolution datasets. The text is modified to make this point clearer.
Following the reviewer's advice, we have compared our offline emissions with those calculated in the one-month WRF-Chem simulation of dust storm event in January 2009. The CLM4 results are shown to be in good agreement with WRF-CHEM, producing similar amount and spatial pattern of dust generation. The text is modified to reflect this comparison.

2. *The authors allude to a year-to-year strong variability (Sec.2.3, p.7 l.18), yet only simulate the short period of 3 consecutive years. Are the results meaningful and how are any resulting systematic uncertainties constrained?*

Actually, in this study, we are more concerned to evaluate the effect of spatial changes than temporal variability, which, as we mentioned in the paper, is not very strong and the coastal plane is a steady dust source every year. The phrase on (Sec.2.3, p.7 l.18) is related to flash floods that differ from year to year but have about decadal frequency. The individual flash floods are small-scale events that cannot make significant changes to vegetation and other surface characteristics. They cause short-term local reductions of dust emission due to their impact on soil moisture. In the long run, however, flash floods are responsible for bringing the alluvial material that is a source for dust generation, but this process is important on much longer timescales. Following the comment, we have expanded and improved the text.

3. *The authors do not address dust aging and its potential effects. They should comment if it's relevant before deposition. Does short atmospheric residence time make it not significant?*

We mention the importance of chemical and physical processes that occur during atmospheric transport (dust aging) when discussing dust deposition to the Red Sea. However, the study is aimed at quantifying dust emission processes and does not consider any transport processes. The detailed analysis of dust transport, lifetime and aging is beyond our current scope and requires a separate modeling framework. But, because of the close proximity of the coastal plain to the Red Sea, the transport is short, and atmospheric processing is not very important for dust generated from the coastal plain. Following the comment, we have made this statement clearer.

4. *Section 2.2 describes mostly the dust generation mechanisms that I understand is not the authors own work and is available in the literature. The whole section should be shortened, replacing the details of model parameterisations with references.*

Thank you for pointing it out. Following the suggestion, we have shortened Section 2.2 to make it more focused. However, we want to keep this section, as it is useful to introduce different parameters we use in the paper.

5. *Section 2.4,5: Given the very coarse resolution and the relatively small size and high complexity of the region under study of the MERRAero grid, is it advisable to scale to the emission total? What is the overall point of this section? Sections 2.4 and 2.5 should be merged.*

We paid a special attention to model calibration, as we believe this is one of the key points of the study. The reanalysis products with atmospheric aerosol component have just recently become available. The spatial resolution of MERRAero reanalysis is enough to capture the dust generation hot spot area in the western Arabian Peninsula. Although it is too coarse to represent the spatial structure of dust emission, the integral multi-year estimate of total dust generation from approximately 150000 km^2 area is a reasonable reference point for model calibration. Following the reviewer's suggestion, we have merged Section 2.4 and Section 2.5.

6. *Section 3.1: Given the same forcing in all experiments, and differences of a few percent between emission resolutions, what is the point of sensitivity tests? Surely the sensitivity to meteorological conditions and climatic variability is more important and should be studied. The authors should comment and better motivate their study.*

The sensitivity tests are aimed at quantifying the impact of high-resolution satellite datasets on spatial patterns of dust generation. In order to perform these tests, the meteorological forcing should be fixed. Although the integral dust generation does not differ strongly, it does not mean that the impact of input datasets is negligible. Locally, dust emissions may change up to 100 % (Fig. 2). We do not analyze the dust emission long-term variability, as it is beyond the scope of the current paper and might be a topic of the further studies. However, we point out that the spatial pattern of dust hot spots is stable (Fig. 6 a-c, Fig. 7) and thus high-resolution inventories should add value regardless of meteorological forcing. Following the suggestion, we have expanded and clarified the corresponding section.

7. *Section 3.2: Surely statistical testing can't involve changing thresholds until significance is reached. Also are monthly values used when hourly are available? Given the strong diurnal variability, isn't the latter advisable?*

To test the temporal variability of modeled dust emission, we process the hourly values to produce monthly time series of dust frequency and intensity, as seasonal variability was the primary focus of the study. Applying the thresholds for model data cannot be avoided, as the station data also have a "threshold": the visibility and weather code is only reported when it drops below 10000 m. Due to this fact, there is no actual diurnal statistics in visibility data, as reported visibility reductions usually only last for several hours, and diurnal cycle cannot be validated. Depending on station's location, meteorological and environmental conditions, the same visibility reduction may be caused by the different level of dust loading; therefore, changing thresholds could not be avoided as well. We show that for the most of the stations, the results are statistically significant in the wide range of threshold values. The statistical metric of model spatial performance, however, does not use thresholds.

8. *Section 4: Again given varying vegetation and land type, isn't 3 years to short a time span to produce an emission climatology?*

The focus of the study is on the spatial structure of dust emission with special attention on generation regime in the hot spots. The hot spot areas are the stable structures that are conditioned by land surface characteristics, whereas the amount of generated dust is mostly driven by wind velocity. We agree with the reviewer that wind circulation regime changes, associated with climate variability, lead to changes in dust emission. However, based on results from MERRAero reanalysis we show that interannual variability of dust generation from the coastal plain during 2003 – 2015 is relatively small (7.2 ± 0.5 Mt a^{-1}) and the seasonal cycle is stable (see fig. 7). The three-year simulation period was enough to capture the significant improvement of the dust emission spatial pattern with the use of high-resolution satellite data. We agree with the reviewer, however, that referring to climatological time scale should be replaced by "multi-year estimate".

Minor comments:

1. Using the study-specific acronyms FineALL, HighALL, LowALL, etc. is confusing to the reader. Propose to just change with quoting model resolution and if necessary input data (ie. 1km, 10km, 50km, etc.)

Following the suggestion, we have changed the experiment acronyms.

2. *Please include the span of years modelled in the abstract.*

The modeling period have been added to the abstract.

3. *Abstract l. 26: appears to be -> is estimated to be [...] suggests -> shows*

Thanks, changed.

4. *P.9 2nd paragraph: First it's stated that it is not rare to have no dust reports, thus the visibility measurement is used, then that the authors prefer sampling as most station visibility are complemented with weather codes. Aren't these statements contradictory? Please re-write the paragraph for more clarity.*

The referred statement means that there were no visibility reduction reports at all, not that visibility reductions came without weather code reports. We have re-formulated these statements.

5. *P.9 l.11 current month -> that particular month*

Changed.

6. *P.13 l.9 Please elaborate how Yu et al is questioned based on the results of the present study - not just by quoting another paper.*

Thanks for the comment. We have re-formulated the statement to address the questioned issue in more details.

7. *P. 19 l.8 Paragraph/line incomplete? Missing full stop?*

Full stop has been added.

Anonymous Referee #2:

***1. Abstract, main part, and conclusions**: The amount 7:5 Mt a^{-1} of total dust emission as presented in the abstract, the successively derived magnitude of dust emission from different locations, and the quantification of the amount of iron oxides and phosphorus are all predetermined by the calibration of the model. For the lack of measured dust emission rates, the calibration is done by assuming the same emitted dust amount in the land model as in the MERRAero reanalysis over the investigated time period, which is not based on measured values, but calculated using a dust module.*

This is correct, our total dust generation and other bulk estimates (e.g. mineralogical composition) are based on calibration with respect to the best, in the absence of observations, assimilation product. However, the main focus of this paper is on the spatial heterogeneity and seasonal variability of dust sources from the coastal plain, and their dependence on the surface data sets. These results are invariant with respect to total bulk emission. Our bulk emission estimates could be easily recalibrated when better estimates of total emission are available.

This approach rests on the assumption that the magnitude of the dust emission in MERRAero is a reliable estimate of the true dust emission from this region. To my knowledge, no evaluation has been published with respect to the dust emission from this region in MERRAero. The authors themselves acknowledge that the resolution of the reanalysis is too low to provide reliable estimates of the dust emission from this region, and they show with their own analysis that the magnitude of dust emission increases with refinement of the horizontal resolution.

See our answer above. In addition, we have to mention that the MERRAero resolution is too low to resolve the spatial heterogeneity of the emissions, but is less deficient for estimating the bulk emissions from the entire area

Thus, this suggest the conclusion that the magnitude of the presented total dust emission, the emission amount from individual locations, and the amount of iron oxides and phosphorus are highly uncertain. This uncertainty should be addressed. Perhaps, one could use the variability of the emission in MERRAero from the whole time period 2003-2015, which already has been used in the manuscript by the authors, to provide a first estimate of the lower and upper range of the emission related quantities presented in the paper, especially for the ones presented in the abstract, even though that still wouldn't address any possible bias in the MERRAero dust emission. The issue of the uncertainty and its sources for the estimates provided in the paper should also be thoroughly discussed in the conclusions. Also, when providing the absolute quantities in the abstract, it should be pointed out that the values are just first estimates that are still highly uncertain.

The uncertainty mentioned by the reviewer is unfortunately, a state of the art problem. However, the proposed approach allows easy recalibration when better estimates are available. We have updated the abstract and the text mentioning that the presented values are the first estimates, and indicated the uncertainty range based on MERRAero variability through the paper, as suggested. We have to mention that although MERRAero dust emission has not been validated yet, the recent paper (Ridley et al., 2016) reports better seasonality of dust AOD in

MERRAero compared to other datasets, and points to potentially better dust emission due to finer spatial resolution and representation of surface winds.

2. Page 4, lines 10–13: *The assumption that the mineral composition of dust aerosols and the mineral composition of the soils in Claquin et al. (1999) and Nickovic et al. (2012) were close does not (always) hold just because of changes during the life cycle of dust from emission to deposition, even more importantly, it does not hold because the measurements of the soil mineral fractions were done for soils that had been wet-sieved. Wet sieving is a technique that strongly disperses soil aggregates (Shao, 2001), which is not realistic for dust emission from the parent soils of the dust sources. This caveat to the assumption made by the authors should be added to the manuscript.*

Having said this, the authors are mainly interested in the amount of iron oxides and phosphorus. Nickovic et al. assume the same iron oxide fraction in the clay and silt size range, and phosphorus is provided only for the clay and the silt-size range together. Therefore, the fractions of these minerals are less affected by the wet-sieving problem, based on these assumptions. Also, in the present manuscript, only the integrated amount over all size bins defined in the dust module is presented. Thus, other sources of uncertainty probably affect the calculated iron oxide and phosphorus amount more than the wet-sieving problem. The wet-sieving issue still may be are more relevant source of error for the other minerals presented in Figure. 9, though.

We thank the reviewer for this detailed suggestion. We have updated the text with the remarks on the additional source of uncertainty linked with wet sieving technique.

3. Page 7, lines 4–7*: How the choice for the threshold value for the statistical source function was made should be explained in detail. It is not clear for the reader from simply stating, "The threshold value is chosen with respect to the temporal frequency of the SEVIRI instrument".*

Thanks for pointing this out. The threshold has to be applied to filter out background dust and is usually chosen empirically (Schepanski et al., 2012). We have tested the thresholds in the range of 0.8 – 1.15 and found the spatial patterns of the source functions to be very similar. The chosen threshold value of 1.12 is larger than the one used in (Ginoux et al., 2012) but comparable to other studies (Schepanski et al., 2012). The choice of relatively large threshold was motivated by several reasons. First, the background dust AOD in Arabian Peninsula is much higher than globally observed one. Second, SEVIRI was shown to overestimate AOD under moist conditions and low dust loadings that are the case for the Red Sea coastal plain (Banks et al., 2013). Overall, this larger threshold allows us to better represent intensive dust sources, in contrast, e.g. to (Ginoux et al., 2010; Ginoux et al., 2012) that aimed at capturing and classifying smaller sources. The text has been updated to account for these remarks.

4. Page 15, line 5*: Do not say "Dust emission climatology", since the analysis is done only for three simulates years. Name the section "Multi-year dust emission" or similar.*

Thanks for this important suggestion, the section has been renamed.

5. Page 14, line 23*: The unit of the total dust emission in the text should be the same as in Fig. 4a, i.e., g m a.*

Changed.

6. Page 17, line 29*: "All of the quantities have a pronounced diurnal cycle, …" should be phrased more precisely as "Total dust generation, frequency, and maximum emission rate have a pronounced diurnal cycle, …". The authors themselves discuss the exception for the intensity further below.*

Changed.

7. Page 18, line 20*: Add Scanza et al. (2015).*

The reference has been added.

8. Page 31, Table 1*: The used individual components of the WRF model configuration should be presented in a way that is friendly to the reader who is not an insider of the WRF model. That is, not just by using acronyms, but fully spelled out, with references added and information how these components can be accessed.*

Thanks for the suggestion. The corresponding information has been added.

References:

1. Banks, J. R., Brindley, H. E., Flamant, C., Garay, M. J., Hsu, N. C., Kalashnikova, O. V, Klüser, L. and Sayer, A. M.: Intercomparison of satellite dust retrieval products over the west African Sahara during the Fennec campaign in June 2011, Remote Sens. Environ., 136, 99–116, doi:http://dx.doi.org/10.1016/j.rse.2013.05.003, 2013.
2. Ginoux, P., Garbuzov, D., and Hsu, N. C.: Identification of anthropogenic and natural dust sources using Moderate Resolution Imaging Spectroradiometer (MODIS) Deep Blue level 2 data, J. Geophys. Res., 115, DOI:10.1029/2009jd012398, 2010.
3. Ginoux, P., Prospero, J. M., Gill, T. E., Hsu, N. C., and Zhao, M.: Global-scale attribution of anthropogenic and natural dust sources and their emission rates based on MODIS deep blue aerosol products, Rev. Geophys., 50, 1-36, DOI:10.1029/2012RG000388, 2012.
4. Ridley, D. A., Heald, C. L., Kok, J. F. and Zhao, C.: An observationally-constrained estimate of global dust aerosol optical depth, Atmos. Chem. Phys. Discuss., 1–31, doi:10.5194/acp-2016-385, 2016.
5. Schepanski, K., Tegen, I. and Macke, A.: Comparison of satellite based observations of Saharan dust source areas, Remote Sens. Environ., 123, 90–97, doi:10.1016/j.rse.2012.03.019, 2012.