Response to Anonymous Referee #2 interactive comments on “Can Explicit convection improve modelled dust in summertime West Africa?”

Reviewers comments are given in black and responses are in green.

The paper’s goal is to understand whether including explicit convection improves the modeled dust in summertime West Africa. In the model used, there is an improved diurnal cycle, but the average dust aerosol optical depth is only slightly modified with explicit convection because the increased evening dust is balanced by a reduction of morning dust (associated with the breakdown of the low-level jet). The results show increases in the frequency of the strongest winds but they are still weaker than observed. Finally, the authors conclude that their study is limited due to other model problems such as the poor representation of the land surface condition in the Sahel, where haboobs are frequently generated in summer.

Although some of the results of the study are interesting, I have some concerns with respect to the formulation of the research question, the experimental set up and the interpretation of the results. I summarize these concerns below.

1) Research question: In my opinion, the content of the paper cannot respond to the question posed in the title: “Can explicit convection improve modelled dust in summertime West Africa?” What the study shows is that errors from neglecting explicit convection are of second order compared to other model errors, and the conclusion can only apply to the limited area version of the UM of the UK Met Office used in the paper. There may be other models with a similar behavior but this is not shown in the paper, and extrapolating would be speculative. It would be convenient to modify the title, otherwise it can be confusing for the reader (in terms of what the reader expects by reading the title). The same happens in the abstract and the paper: it has to be clear that these results are specific to the UM.

Firstly, we would like to thank the reviewer for their time and the helpful comments they have made on this manuscript.

The title of the paper is meant to indicate that the expectation we had on starting this work. We had thought that by representing convection explicitly we would see a marked improvement in the dust aerosol fields when comparing with simulations with parameterised convection and satellite retrievals of Aerosol Optical Depth (AOD). We tried to make this clear in the “story” of the paper with section 3.1 outlining the lack of effect of representation of convection on the overall AOD and the subsequent results sections investigating the differences in near surface wind and modelled storms. In that respect we don’t believe the title to be misleading, the answer to the posed question is “no”, and we have then attempted to explain the reason that this is the case.

However, we are also aware that the results shown in this work are limited to a single setup of the Met Office Unified Model (UM) running in a limited area setup without reinitialisations. It is plausible that simulations performed using models from different centres would produce different responses to a change in the representation of convection. **Therefore, we have modified the title of the work to indicate that this work focusses on the UM in a limited area framework. We have also endeavoured to make sure that in the discussion of**

results throughout the paper that it is made clear that other models might respond differently.

2) Figures 2 and 6 show a very poor behavior of the model (regardless of explicit convection) when compared to observations. For example, there is a factor 3 to 4 difference in the AOD and an uncorrelated seasonality when comparing the model to the MODIS observations. I have several questions in this respect:

a. To what extent the retrieved AOD form MODIS is reliable over land? Some measure of uncertainty in the observation is needed when using AOD products over bright surfaces.

Over bright surfaces the combined product relies on the Deep Blue algorithm. The implementation of the Deep Blue algorithm has provided a much improved technique for the retrieval of AOD values over bright surfaces compared to the Dark Target algorithm. By using maps and libraries of surface reflectance in the blue part of the spectrum the Deep Blue algorithm is able to retrieve AOD values that compare well with AERONET measurements. The technique is described in detail in Hsu et al., (2013) and has an estimated error better than 0.05 + 20 %, with 79 % of the best AOD data falling within this range. **Detail about the accuracy of the product have now been included in the manuscript.**

This concern has also been addressed in the comments to reviewer 1 (see below).

"The MODIS combined product utilises both Dark Target and Deep Blue algorithms to be able to retrieve aerosol optical depth over both ocean and land (including arid, high albedo surfaces). Sayer et al., (2014) gives a detailed analysis of the performance of both algorithms with the MODIS AOD comparing well globally with AERONET AOD retrievals (R= 0.92) with the MODIS and AERONET AOD distribution being centred around the 1:1 line.

In dusty regions there is a tendency for MODIS to underestimate AOD values. However, the spatial distribution of dust will be well represented and allow for a good comparison with the spatial distribution of modelled dust. Also, given the very low values of AOD in the simulations compared to retrievals, the absolute value of AOD in MODIS becomes less important for analysis."

b. Have the model outputs been spatially and temporally collocated with the MODIS data in order to perform the comparison (i.e. did you select the modeled times and places corresponding to the availability of the MODIS data?) If not, to what extent your comparison can be affected? My experience is that it matters a lot. Would this make sense with your current model set up, i.e. a regional climate run only fed by the analysis data though the boundaries?

We would like to apologise for not making this clear. In all comparisons between MODIS and models, simulations are sub sampled temporally to limit the potential errors introduced by the diurnal variations and spatially to account for missing MODIS data (due to clouds). The comparison of mapped monthly averages and, box averages also reduces the introduction of erroneous analysis that would result from using a single model grid box or observation site and interpolating a pattern across a wider region. **The manuscript has been updated to make this treatment of model data clearer.**

c. Why AERONET stations were not used? There are quite a few stations in the domain for 2011. AERONET is reliable and is the main tool used to evaluate model performance. Without the AERONET evaluation is difficult to judge the performance of this model compared to other models. Nowadays many regional models represent reasonably well the seasonality of dust in AERONET stations (daily correlations between 0.6 and 0.8 when reinitialized daily and without dust data assimilation). There are also available high resolution PM10 surface observation concentrations for the Sahelian Transect (Marticorena et al 2010) that would really help evaluating the model.

Even without AERONET or surface PM10 measurements we are able to see that, with respect to dust, the simulations are doing a poor job. The location where dust is being raised and where it reaches its greatest AOD values do not match well with satellite observations, therefore we didn't feel it was necessary to include further (single point) observations of dust. Were the simulations doing a better job of replicating the spatial and temporal variations of dust we think that AERONET or the Sahelian transect would have been useful in further validating the simulations, however, given the large simulation/retrieval discrepancies already shown we cannot see how this could add extra insight.

Operationally, dust models will be able to be extensively tuned for overall dust loading to match observed values, that is not the case here. One of the reasons for the experimental design was to have conditions similar to a true, free running simulation and allow insight into the kind of behaviour that is produced by the model in climate simulation conditions. It is imperative that we try to understand the factors that dominate dust uplift in models constrained in this way and try to understand the limitations of simulations in representing real physical processes. **This sentiment has been strengthened in the paper.**

d. Concerning the previous point: in the introduction the authors claim that both winds and dust should be explored together with observations. It is surprising that the authors do not use the most reliable resources of dust measurements besides the more uncertain satellite products.

It was felt that there was a greater value in having a product with a broader spatial coverage, that allowed the regional pattern of atmospheric dust to be compared with simulated dust fields. As mentioned above, had the simulations shown a better spatial distribution of dust then it is likely that other observations (for model validation) would have been required.

e. Concerning the general decrease of dust in the model from May to September (compared to the observations showing a peak around July): Given that the model is not reinitialized every, has the humidity in May 1 been warmed up for at least 1 year? If not, this could be a reason for such behavior (the model could be showing a trend in dust because of a drift in the soil humidity). Has the model been evaluated for the same time period reinitializing the atmosphere and the soil every day from the parent domains?

We believe that the month on month reduction in AOD can most convincingly be explained by the reduction in u* values (especially in the Sahara and Northern Sahara boxes). This has been discussed in the reply to reviewer #1 (see below).

"There is generally a decrease in u* in the Sahara and Northern Sahara from May to September. However, the pattern in the Sahel is less clear as shown in Figure 6. Given that emission is a cubic function of u* associated with exceedance of a u* threshold, it is no surprise that with dropping u* values (possibly associated with a weakening Harmattan after the start of the monsoon onset and developing Saharan Heat Low) that AOD values would also fall. We see that there is only a weak growth in afternoon and evening winds and an unrealistic drop in nocturnal low level jet winds in convection permitting simulations. As discussed in the paper, this can be attributed to cold pools that are too small and too weak. Cold pools are known to be key to dust uplift in the Sahel and southern Sahara from observations (Marsham etal., 2013, Allen et al., 2013, 2014) and given the recirculation of dust around the Saharan Heat Low would go a long way to explaining the peak dust location growing through June, July and August in MODIS AOD."

The greatest control on soil moisture in the simulations is rainfall. The Northern Sahara box shows drying throughout the modelled period (making uplift more likely), the Sahara box shows little change and there is a large increase in the Sahel (associated with the monsoon rains; all shown on figure 6). We think that this kind of behaviour is to be expected and doesn't constitute a "drift" in soil humidity that would reduce uplift across the region over the simulated period.

f. More details should be given on the emission scheme. Do the authors use a preferential source? Do they use estimates of aerodynamic roughness length? This may also at least partly explain such a mismatch with observations.

The dust emission scheme is that with the Met Office Coupled Large-scale Atmosphere Simulator for Studies in Climate (CLASSIC) and is described in detail in Johnson et al., (2011). Dust emission is calculated at each time step using prognostic model fields. The widely used approach of Marticorena and Bergametti (1995) is employed where the horizontal flux of sediment in 9 bins is calculated (bin sizes are 0.0316, 0.1, 1.0, 3.16, 10.0, 31.6, 100.0, 316.0, 1000.0 μm radius). This is given by

$$G_i = \rho B U^{*3} \left(\frac{1 + U_{ti}^*}{U^*} \right) \left(1 - \left(\frac{U_{ti}^*}{U^*} \right)^2 \right) M_i \frac{CD}{g}$$

Where i refers to bin number, G is the horizontal flux, ρ is the air density at the surface, B is the bare soil fraction, U* is the friction velocity over bare soil, U_t* is the threshold friction velocity, M is the mass fraction of soil particles in the bin, C is the constant of proportionality, D is a tunable parameter and g is acceleration due to gravity. Emission is also inhibited if snow is present, the ground is frozen, on steep slopes, if soil moisture is too high and at costal grid points with fractional land cover.

The U* value used for emission calculations is generated in the Joint UK Land Environment simulator (JULES) scheme. U* is largely a function of changing wind speed with height. Therefore the surface roughness (but not orographic roughness e.g. vegetation) is an important factor in the calculation of dust emission. M is dependent on the soil characteristics in a particular grid box (from ancillary files) and takes into consideration soil clay, silt and sand fractions. The fraction of clay in a particular soil also impacts how soils respond to moisture by modifying the threshold friction velocity depending on clay fraction and soil moisture according to the method of Fécan et al., (1999).

The vertical transport of dust away from the surface is linked to the horizontal sediment transport by

$$F = 10^{(13.4 F_c - 8.0)} \sum_{i=1}^{9} G_i$$

Where F is the vertical flux of dust away from the surface, Fc is clay fraction. Emission and transport away from the surface only takes place in the smallest 6 size bins.

As such, there is no explicit preferential dust sources used, however, where soil characteristics and surface roughness conditions are favourable, the threshold U* value for emission can be reduced, allowing for favoured emission in particular regions. This could go some way to explaining erroneously high-uplift regions such as that along the Atlantic coast, however, this would not explain the near zero emission in the Sahel. **Statements to this effect have been added to the paper to clarify.**

3) A major question: because convection seems to be a second order error in this study, can we really respond to the question posed in the title?

While we appreciate that assessing the impact of convection permitting simulations on dust uplift cannot be fully investigated, we believe that the initial question is still valid. The answer to the question "Can Explicit convection improve modelled dust in summertime West Africa?" is "not currently" (at least in the UM). we think that is a well understood problem that parameterised convection in models limits their ability to represent the strongest winds for important dust uplift regions. However, just because the results from this work were unexpected should not be a reason to alter the initial question.

4) What can explain that the 12 Km explicit and 4 km explicit are so similar?

We would not expect there to be a large difference between the 12 km and 4 km explicit simulations. Previous work as part of the Cascade project (Birch et al., 2014) showed that the largest differences in this region were the result of changing the representation of convection, not modifying grid-scale. Over West Africa it seems that the scale of the convective storms (hundreds of km) means that they can be represented in simulations that would generally be considered to have too coarse a grid-spacing for the representation of convection. There are key differences between different explicit simulations and observed storms, some of which are likely to be the result of simulation grid-scale, however, those results are beyond the scope of this paper.

5) In Figure 7, the 12-km explicit has a more prominent tail of high winds compared to the 4-km explicit. This behavior is surprising to me. What does explain this behavior? Is the frequency at a specific location comparable using different model resolutions? That is not really clear to me.

There are obviously likely to be some differences due to the differing grid-box scales between different simulations, however, the fact that all the simulations (from 40 km to 4 km) show remarkably similar wind speed distributions suggests that the grid-scale effect is relatively small. I'm not sure that the differences between the 12 km and 4km explicit simulations at the high wind speeds tail are so great that we would consider them to be an important feature. In fact the only clear differences that we think are important for high wind speeds is the shortfall in frequency of all simulations when compared to observed winds at the Saharan stations.

6) Figure 12: Does it make sense to compare the model for a specific day for this experimental set up? Reproducing a specific episode requires (recent) and accurate initial conditions and the model is running in a regional climate mode only constrained through the boundaries.

The aim of this figure was to show several behaviours in the simulations: (1) that simulations with explicit convection are capable to generating large MCS storms, (2) the large MCSs (when present in simulations) are capable of generating convective cold pools that spread into the Sahara, (3) that when cold pools are present they do produce a maxima in wind speed (or friction velocity) that we would expect to be associated with the raising of dust and

(4) that the emission of dust from such an event is lower than might be expected when compared to similar observed events.

In order to highlight these points the storms chosen from observations and simulations did not have to be on the same day or time. However, it so happened that the storm generated on 23rd August 2011 in the 4 km simulation was a good match for an actual storm that occurred on that date. This suggests that information from the lateral boundary conditions (hourly global UM simulation reinitialised every 6 hours using ERA-Interim data) was able to produce conditions favourable to convection near the middle of the model domain, despite being such a long time into the simulation.

Even if these events had been on different days it would still be possible to show that points 1 to 4 mentioned above are all true. The weak uplift from event in the simulation highlights a real issue in the simulation of dust emission. However, it is important to note that whether the cause is weak winds or a problem of surface characteristics is beyond the remit of this study.