Reply to referee 2

We thank the referee for the thorough review and the helpful advices to improve the article. They have all been considered in a revised manuscript and a new supplement. In the following please find our replies to the individual comments.

> General comments This study presents updates of the dust emission scheme implemented in the global atmospheric chemistry model EMAC based on the previous work of Astitha et al. (2012). The land cover, vegetation topography and clay fraction maps are updated to more recent versions using higher spatial resolution. Changes are also imposed to the dust emission scheme directly. The updated dust emissions are evaluated with AOD measurements from AERONET, MODIS and IASI for the year 2011. The title, flow and structure of the paper are appropriate. All updates are well received and long needed, given the importance of quality input data to accurately parameterize physical processes that cannot be described by first principles. However, the authors keep the evaluation part largely on the qualitative side, which does not help the reader and the community to fully understand why these changes were impactful and significant. The conclusions are also very brief for a model development/improvement paper. The authors miss a great opportunity to discuss the very interesting aspects of each revision and inform the community of which one should be considered more impactful (if not all). The specific comments below will help the authors revise the paper so it can be accepted for publication with GMD.

We have added a section discussing the effects of the individual modifications. The evaluation - comparing numeric results for AOD, DAOD, correlation coefficients and skill scores - has been extended by even more quantitative comparisons of dust concentrations and deposition rates.

> Specific comments/suggestions
> Section 2.2 (Vegetation): Please elaborate on the calculation of fveg (Eq.1): What is the role of 0.35 and what is the meaning of fveg being 1 or less than 1.

This vegetation factor, also used by Astitha et al. 2012, interpolates linearly between full emissions for no vegetation and entirely suppressed emissions for LAI > 0.35 which was introduced as threshold by Mahowald et al. 1999. We have added this information.

> Section 2.3 (Clay fraction): provide a map of the updated clay fraction in comparison to the one previously used in the model. It will provide context on the significance of changes that later affect the parameterization scheme.

The map has been added to the supplement.

> Section 3 Page 4 (soil moisture): The soil moisture term in Astitha et al. (2012) and Eq. A1 in this paper is omitted from the threshold friction velocity. However, the authors correctly describe the dependence of dust emission on soil moisture at the end of this paragraph. What is not clear is if the statement "we consider a detailed parametrisation of the soil moisture effect to be essential to capture the observed trends in future simulations. This will require a comprehensive soil model providing accurate moisture values for the topmost surface layer" refers to an action already taken for this study or a future goal. In any case, a discussion on how the exclusion of soil moisture correction influences the simulations is important here.

The more comprehensive soil model is a future goal, we have clarified the statement and added a figure illustrating the effect of omitting the factor to the supplement.

> Page 5, (Surface friction velocity limit): a note must be placed that Eq.2 holds only when u* > u*t. Also, choosing to limit the threshold velocity to a maximum value of 0.4 m/s seems arbitrary and needs to be elaborated. What led the authors to this specific value? Some context and rationale must be provided.

We have clarified the equation and included results for different limits in the supplement.

Page 5 (Topography factor): I am not sure of the role of the normalization factor 5.3 and how it conserves global emissions. It sounds like a tuning factor to me, so please elaborate on the role of the factor and the method used to estimate it.

Using the topography factor S_{topo} as given in Eq. (3) has two effects: the desired effect is that it adjusts the spatial distribution of the emissions, but since by definition $0 \leq S_{\mathrm{topo}} \leq 1$ and usually $S_{\mathrm{topo}} < 1$ an undesired side effect is the reduction of the emissions globally. We quantified this reduction in a one month simulation obtaining a ratio between the global emissions without and including the factor S_{topo} of 5.3. Consequently, we include $5.3 \times S_{\mathrm{topo}}$ instead of just S_{topo} and thereby conserve the global emissions. In practice, this normalisation factor can be combined with the empirical scaling factor c, hence it introduces no additional tuning factor. We have expanded the corresponding corresponding text.

Page 5 (Mode mapping): This is not a strict update of the emission scheme but rather an alteration in order to use the GMXE aerosol model compared to M7 used by Astitha et al. (2012). A brief note must be included in this section to clarify that the original scheme (as well as the reference simulation herein) used a different aerosol module thus a different approach to particle size distribution. The omission of the eight transport size bins is surely a change from the original version.

In this study, we use the GMXE submodel for both, reference and validation simulation. Since GMXE is based on M7 and uses the same modal concept, the question of how to map the three emission modes to the aerosol submodel modes is unaffected by this choice. Since in the original scheme the dust was not further processed while in the "transport" bins but directly mapped to the GMXE/M7 modes, skipping this step is in fact an implementation detail and when aligning the threshold between accumulation and coarse mode with the bin boundary at radius 0.6 um yields identical results.

Figure 6: what is the higher value in this scale (above 0.1)? When we see 0.1 fraction of Ca++, does that indicate the mass, volume (or else) fraction of the total particles within each specific grid cell? A better explanation of the mineral cations fraction could be included also in page 6 (last paragraph of section 3).

The upper limit of the scale is 0.12, which is reached by the Ca++ fraction in the Kalahari and Taklamakan Desert; we have added a tick mark. The fractions shown in Figure 6 are mass fractions; we have added this missing information in the caption.

Section 4.1, page 7: 1. "On the other hand, dust events observed by AERONET in January and December are reproduced by the validation simulation, but not by the reference simulation": this comparison is not at all discernible in the plot as it is. If this is an important argument, the plot must be revised somehow to make the statement visible.

We have marked the events we are referring to in the plot.

2. Given that Izana and La Laguna are within the same model grid cell, an average of the AOD from both sites could be an alternative way to compare with the model value. In addition, when evaluating numerical model simulations one can employ the nearest neighbor (as done here) or a bilinear interpolation between the observation and the model value from the four closest grid points.

Averaging the AOD values of all stations within the same grid cell generally is a reasonable strategy, in this case, however, even La Laguna station at 568 m altitude is not representative for the grid cell which is mostly covered by ocean. Since the neighbouring cells are also predominantly covered by water, bilinear interpolation would not make a big difference in this regard. Better agreement could be obtained by computing the model AOD at station altitude rather than at model surface height. Generally, such distinctive sub-grid topographies and shore lines reveal the limitations of the model resolution.

> 3. The main criticism I have for the evaluation using the skill score (as with any other statistical metric) is the qualitative determination of which configuration provided the best results. Characterizations such as “slightly better” or “marginally larger” do not show robustness in the performed evaluation. My immediate question is: are these differences statistically significant? Are they statistically different? This is the only way to prove or convince the audience that a, say, 0.05 change in the skill score is significant enough.

Measures like the skill score are supposed to quantify agreement. The significance of the skill score improvement has only been indicated by the dominance of green bars in Figure 8. In the revised manuscript we have amended the figure to depict error estimates for the ΔS values.

> 4. What about using AOD of coarse vs. accumulation modes from AERONET?

The fine/coarse mode AOD product is more sparse than the AOD data, moreover it is only available at 500 nm and we do not have corresponding model output available. We will extend the evaluation to other observables instead (see below).

> 5. How about using total PM concentrations wherever available (and for cases of high dust concentrations) to evaluate model performance? An additional means of quantitative evaluation needs to be included.

The evaluation will be extended by comparisons with dust concentration and deposition data.

> Figure 8: Please consider replacing “Regions A, B,” etc. from the figure with the names of the regions as it is not convenient to go back and forth between Fig. 7 and 8 to identify the regions.‘

A more convenient naming has been introduced.

> Sections 4.2 and 4.3: The scarcity of desert dust concentration measurements is a well-known problem in the modeling community when assessing model and parameterization scheme performances. This is when satellite and remote sensing observations come into play and are important tools of assessment. Nevertheless, leaving the comparison in the qualitative state only, influences the robustness of the conclusions. Looking at the IASI zoomed plots (Fig. 13), I would not immediately say that the validation is better than the reference simulation. They are different and somewhat both incorrect in my view. If the authors presented a quantitative assessment of the performance, there would be no doubt on the comparison. Also, why is the zoomed area over Middle East only? What is the special interest for this specific region? This has to be explained thoroughly so it will not be seen as “cherry picking”.

To guide the eye in Figs. 11 and 13 we have marked the region where we see considerable improvements, which is the Arabian Peninsula including Iraq, Syria and Jordan. To corroborate the improvements, we have evaluted the spatial correlations and skill scores in this region which significantly increase as shown in the new Figs. S6 and S7 in the supplement. The Middle East is of special interest because there the original emission scheme clearly suffered from outdated input data, as mentioned in the introduction. To avoid the impression of cherry picking we now provide seasonal global plots in a supplement.

> Conclusions: the conclusions are quite brief (two sentences in the end of the section). More discussion should be invoked on how the changes influence a better model performance (as long as there is a robust determination of “better” or “worse”). There are very interesting aspects in this study and it would be very useful for the community to understand how the changes that were implemented individually affect dust emissions. I believe that adding such discussion would greatly strengthen the paper.

The conclusions now reflect the additional aspects covered by the revision.

Appendix: There are a couple of things missing in the depiction of the emission flux jemis (Eq A2): 1) I don't see the mass fraction (source to transport bins, M in Eq.4) that should be multiplied in the right side of the equation. 2) In Astitha et al. (2012) they used the relative surface area covered from particles with diameter D (Srel) to calculate the horizontal flux H. Did the authors omit this calculation in their revisions?

The emission flux given in Eq. (A2) is the total emission flux. In the revised manuscript we have multiplied the mass fraction to present the flux for each mode instead. The factor S_{rel} is only utilised in the emission scheme variant DU_Astitha2, not in the scheme DU_Astitha1 used in this study.