

Geosci. Model Dev. Discuss., referee comment RC2
<https://doi.org/10.5194/gmd-2021-423-RC2>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on gmd-2021-423

Anonymous Referee #2

Referee comment on "Evaluating dust emission model performance using dichotomous satellite observations of dust emission" by Mark Hennen et al., Geosci. Model Dev. Discuss., <https://doi.org/10.5194/gmd-2021-423-RC2>, 2022

Review of Hennen et al. Evaluating dust emission model performance using dichotomous satellite observations of dust emission

The paper proposes a framework for dust emission model evaluation based on dichotomous (presence=1 or absence=0) observations of dust emission point sources (DPS) derived from satellite data. To show the potential of the framework, a so-called albedo-based dust emission model (AEM) using a smooth entrainment threshold fixed over space and time is evaluated. The paper uses 9 DPS datasets using different sensors and approximations depending on the region. The results are that: 1) dust emission is rare (1.8 %) even in North Africa and the Middle East, which would indicate extreme, large wind speed events, 2) The AEM overestimates the occurrence of dust emission by between 1 and 2 orders of magnitude. It is then concluded that 1) the smooth entrainment threshold is typically too small and needs to vary over space and time, 2) there is incompatibility between threshold and friction velocity scales, 3) False positives are linked to the absence of any limit in sediment supply in the model and therefore 4) new schemes are needed to the smooth threshold and account for restrictions in sediment supply. Finally, the authors state that the DPS data provide "a consistent, reproducible, and valid framework for the routine evaluation of dust emission models and potential model optimisation, and that the study emphasizes the "growing recognition that dust emission models should not be evaluated against atmospheric dust".

While the overall approach could represent a valuable complement to current evaluation capabilities of dust models, its current implementation and interpretation has, in my opinion, several conceptual flaws and limitations that very likely bias the conclusions of the paper and largely limit the applicability of the approach. In addition, in view of these limitations, statements such as "the study emphasizes the growing recognition that dust emission models should not be evaluated against atmospheric dust" are just not supported by evidence provided in the paper.

Below are my general comments/concerns. Based on them I do not recommend publication of this manuscript as I cannot see how those fundamental limitations and their impact on the conclusions and perspectives of the paper can be easily amended.

On the evaluation of models with observations of dust emission point sources vs atmospheric dust:

Satellites do not observe dust emission but atmospheric dust. Estimates of dust emission point sources are retrieved or inferred from atmospheric observations. In fact, the same applies to in situ measurements: emission cannot be observed directly and can only be inferred from airborne measurements. This is an important conceptual nuance, and it must be clear that the proposed framework relies on a DPS dataset, which infers emission point sources based on many assumptions and potentially important limitations as I will describe below.

Even if the implementation of the proposed evaluation approach would be sound, why dust emission models “should not be evaluated against atmospheric dust”? Why is it incompatible? The statement is just not justified, particularly given the limitations highlighted in my next comment. Evaluation efforts of dust models (with embedded emission and dust cycle) typically include a variety of observations (in situ, satellite, remote sensing) of different variables at different spatial and temporal scales, and all are very welcome and helpful to characterize the behavior of a dust model including its emission. Why not seeing different approaches as complementary? In any case, the statement is just an opinion and is basically not supported by the results of the paper.

On the use of DPS datasets to evaluate dust emission models:

At present there are well known limitations in the retrievals used to infer dust emission that are very likely strongly biasing the comparison with the dust emission model. Take for example SEVIRI Dust RGB product used over North Africa and the Middle East. It is well known that the product can detect the particularly high-concentration dust storms but fails in detecting thin, low level and/or low-medium concentration dust clouds/events, which can come from frequent low emission events that are widespread in North Africa and the Middle East (partly due to the high availability of saltators). This important limitation invalidates to a large extent the proposed dichotomous evaluation approach as dust emission from the model contains all type of dust emission events ($u^* > u^*_{ts}$) and dust emission from the DPS is strongly biased towards high emission events, which makes the proposed framework currently inconsistent, and the conclusions likely flawed. For example, the overestimation of the occurrence of dust emission in the AEM by 1 – 2 orders of magnitude and the rarity of dust emission in the DPS for North Africa and the Middle East point towards a problem in this sense.

Another potential problem for the evaluation of global models is the inconsistency of the DPS among regions. Using MODIS in some regions and SEVIRI in others with their

different sensitivities could further bias the conclusions on the behavior of a model in different regions. A nice exercise would be to compare a DPS in North Africa and the Middle East based on SEVIRI with another one based on MODIS, and see how the evaluation and the conclusions are impacted.

There are no easy solutions to circumvent the problem of the low-medium dust emission events. Also, this problem clearly evidences that quantitative AOD products (along with their quantified uncertainties) over sources regions can and should at least complement dust emission model evaluation efforts.

In addition to these inherent biases in the DPS and the associated comparison, the difference between simulated wind scale, and the DPS scale makes the interpretation of the results very complex. I acknowledge there is a section in the discussion about this problem but in my opinion the issue should already be considered in the basic design of the evaluation framework. In other words, models should be evaluated as much as possible at consistent spatiotemporal scales, otherwise conclusions can be fundamentally flawed.

All in all, the previous highlighted limitations are very likely biasing the results obtained and the derived conclusions. For example, the overestimation of the frequency of dust emission is partly attributed to the absence of any limit to sediment supply.