Reply to RC2

We thank the referee for the constructive review. We have addressed the comments with a revised manuscript and the point-by-point reply below.

> Review of "Data-driven aeolian dust emission scheme for climate modelling, evaluated with EMAC 2.54" By Klingmüller and Lelieveld
> The manuscript presents a data-driven (machine learning) dust emission scheme that is trained together with other data-driven components (transport, deposition,...) to reproduce dust aerosol optical depth retrievals by the Infrared Atmospheric Sounding Interferometer on board the MetOp-A satellite. The resulting data-driven scheme is then applied in a climate chemisty model (EMAC) and compared to the standard dust emission scheme in that model. The results show improvement in the representation of the dust cycle when using the data-driven scheme in comparison to AERONET measurements.
> This is a very interesting and novel contribution that is well-written and deserves to be published. I have some comments that would need to be addressed before the paper is accepted for publication. (I have to say that I am an expert on dust emission and modeling and I have limited knowledge on machine learning. Therefore, the paper would benefit from an additional review by a machine learning expert.)
> ■ The application of data-driven (learning) methods to dust emission and modeling is quite unprecedented. There is almost no literature on the subject and generally dust modelers (like me) are not specialized in machine/deep learning methods. I find that the description of the method lacks detail, both general/conceptual and specific, at least the kind of details that would benefit the future main readership of this article (dust and climate modelers like me). It would therefore be very valuable that the authors add more detail to the background and the methodology. Why they selected such learning framework and not others? Is this a deep learning algorithm, neural network? What does it bring to the table compared to other machine methods for this specific problem? What functionalities of Pytroch were used? This is really not totally clear and it is not explicitly discussed.

We have expanded Section 3 to better explain and motivate the methodology.

> ■ The derived emission scheme depends upon the following inputs: friction velocity, vegetation, clay fraction, soil moisture, snow cover, topography and land use cover. All of them depend upon location and some of them on location and time. However, if I correctly understood, with this method, emission will depend upon grid-cell (location) even if all the inputs are the same. Can you elaborate on that? I am assuming that this is a feature of the (deep) learning method used, but maybe just I misunderstood the method (sorry if that is the case). This question goes back to my main point: more details are needed to understand the data-driven model and its implications. This issue is potentially important: I wonder about the determinism of the method. Do you obtain different emissions in different grid cells even if the input parameters are exactly the same? Particularly, in climate model simulations (future and past climates), with changes in climate regime, the lack of determinism could be problematic. If dust emission is not deterministic, could you provide some data analysis to show to what extent this could be

> (or not) problematic. In other words, could you show the degree of determinism of your model? One way would be to run one time step of the model with constant inputs everywhere.

The emission model depends on the local values of the above mentioned input variables and does not explicitly depend on the location and time, which we have clarified in the revised Section 3 (only the topography factor additionally considers the terrain in surrounding grid cells). Consequently, if evaluated at a different location and time but with the same input values, it yields identical emissions. This makes the emission scheme well suited for climate scenario simulations because it does not rely on memorised spatial or temporal patterns that might change with climate.

> ■ What is the relative importance of the different features/inputs used? Can that be derived from your method? All in all, as you see I am really interested on this work but there is a general lack of diagnostics to understand the behavior of the system.

In the revised manuscript, we have included estimates of the importance of each input variable (Fig. 6). These estimates should be interpreted with caution, in particular because the input variables are to some extent correlated, so that the effect of one variable may be reflected in the importance estimate of another variable. A more detailed analysis of the importance of different factors is subject of a follow-up study.

> ■ There are many choices that seem a bit arbitrary or at least not justified. Some examples follow. Why the atmosphere goes up to 4 km and there are only 4 vertical layers? Why the dust concentration in the four layers is increased by the emission flux? The dusty PBL in the Sahara and the Saharan air layer reach easily 6 km in Summer.

In section 2 we now explain the choice of the vertical levels: "Mineral dust can reach higher altitudes, but since most of the dust mass remains within these layers, this approximation is a reasonable compromise between a realistic representation of dust transport and an acceptable computational burden." Future versions, especially when utilising new hardware, will consider more and higher levels. The dust emission flux increases the concentration in the lowest layers directly, the concentration in the higher levels increases only by vertical transport. We have clarified the sentence of Eq. (3).

> ■ Line 218: The model seems think that dry deposition is unimportant. Can you comment on that? There is strong dry deposition over sources. Is it possible that the method is just adjusting/providing a net emission (emission – deposition)? That could be the case as dry deposition scales with friction velocity (like emission). Comment on the implications.

It is plausible that the emission flux we obtain is to some extend a net emission flux because deposition contributions that are directly linked to the conditions causing the emissions, i.e., that have the same spatial and temporal distribution, cannot be distinguished from a reduction of the emissions. However, since the distribution of deposition in general is not identical to that of the emissions, the model should be able to learn deposition and emission parameters independently. We have rephrased line 121 to "[...] indicating that the current implementation of our approach is not sensitive enough to identify dry deposition. The reason is likely the limited atmospheric residence time discussed below, which represents an unrealistic additional removal process. Moreover, a fraction of the deposition that always occurs collocated with emissions might be represented by a reduced emission flux."

> ■ Have you analyzed the trainable parameters? It would be very informative to understand whether the trainable parameters make physical sense. In my previous comment, clearly the fact that there is no dry deposition outlines this problem.

Besides creating an emission scheme for the application in climate models, obtaining a better understanding of dust mobilisation by studying the trained model is a main goal of our efforts. However, the latter requires careful analysis. One reason is that, as mentioned above, our present input variables are not independent, which is not a problem with regard to good model predictions, but complicates studying the effect of the individual variables. Therefore, in this article we focus on the practical application of the model and address a comprehensive model analysis in a separate article. Nevertheless, we have added plots of the effect of the surface friction velocity and the soil moisture, which, according to our preliminary analysis, are particularly important.