Reply to RC2

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

> This study explores DNN-based soil moisture predictions that are trained on remote sensing data and forced by a climate model, which are used to replace low-quality soil moisture predictions in that model for an improved prediction of mineral dust emissions. Overall, the topic is relevant, the methodology novel and sound, and the manuscript concise, well structured, and was a pleasure to read. I just have a few concerns and suggestions which I hope the authors will consider before publication.
> -) To me, hardware developments do not seem to be the most pressing issue and motivation for the presented study. I would recommend to start the introduction outright with the more relevant issue, i.e., the low-quality internal soil moisture predictions of the climate model which could potentially be replaced by a DNN-based module.

The performance and hardware developments are not the most pressing issue for the soil moisture example, where the immediate benefit is a more realistic representation of surface soil moisture. However, we believe the challenge of adapting climate and atmospheric chemistry models to current and future hardware is serious enough to justify the short digression on this aspect. Building on the soil moisture example, computational more demanding model components can be addressed to achieve more significant performance gains.

> -) Maybe a better way to achieve independence between training and test data than doing a standard 80/20 splitting could be to take the characteristics of the ESA CCI SM product into account, that is, splitting the data at a time where there is a change in the underlying satellite instruments that are being merged. Or perhaps better yet: training the model on the "active-only" ESA CCI SM data set, and evaluating it on the "passive-only" data set (but still in a different period of course). That's probably too much to ask for this paper, but perhaps something to keep in mind for future studies.

Taking advantage of the details of the dataset is a very useful proposal for future refinements. So far, we deliberately chose a "naive" approach by just using the most complete data set as is without manually adding additional knowledge about the data (which might not be available in other applications). The split between training and test data was used to obtain more than one year of test data while keeping as much training data as possible. Additionally, we required the training and test data to be in chronological order so that DNN predictions in the test period represent forecasts. For optimal results, future refinements could certainly exploit detailed characteristics of the satellite data. We added the motivation of the data split in section 3.

> -) What motivates the selection of the predictor variables? Are they taken from another study, or did you try out different combinations and evaluated variable importance? A little elaboration on that would be helpful.

The predictors where chosen based on the availability in the simulation output on the one hand

and physical relevance on the other hand. We tested using time lags of more than one week without obtaining a clear improvement of the validation performance. However, since we did not systematically test all possible predictor combinations, better choices are certainly possible. We now elaborate on the predictor choice in section 3.

> -) On a related note: I am a bit concerned that the DNN predicts mostly a spatial and temporal climatology. The correlations of >0.9 are arguably unusually high for typical soil moisture data sets. This argument could be settled, for example, by showing anomaly correlations as well. As for the temporal correlations: The values of ~0.5 indeed appear to be more realistic, but I don't understand the reason for not showing the acutal values in Figure 4. A discussion on how these temporal correlations compare to values typically found for other data sets could also be beneficial. For example, Figure 7 in Dorigo et al. (2017) shows correlation coefficients of modeled versus ESA CCI SM soil moisture, which may serve as a good benchmark for putting the attained values into perspective. Another way to appease the reader may come from an elaboration on the selection of predicture variables, as noted above... For example, how would the predictions look if merely coordinates and the time of the year would be provided? Or if they would be omitted? Also, the time series shown in Fig. 5 do not really convince me. For example, it is pointed to "irregular features" in, for example, October 2014. But to me it seems that there are distinct dry-downs in the prediction in both October 2014 and October 2015, while this dry-down is only visible in the observation in 2014.

The high overall correlation coefficient is dominated by the realistic spatial distribution and can almost be achieved by a climatology, but the temporal correlation is the more challenging contribution and good results are only obtained including the meteorological predictors (Figs. S4 and S5 in the new supplement). We added a discussion of the importance of meteorological vs coordinate/time predictors in section 3 and updated Fig. 4 to show the actual correlation coefficient. The features of the time series in Fig. 5 mentioned in the text are magnified in Fig. S3 in the new supplement.

> -) Having said that, for the application shown, it seems that this may not be relevant at all. The improvements in dust emission predictions shown in Figure 6 are quite remarkable, and it seems that the DNN-predicted values, even it they would be only a climatology with possibly a little short-term variability signal added on top, are doing already much better than the model-internal soil moisture representation. The taken approach of validating soil moisture predictions using a downstream application is a great way to show their actual utility, which is much more useful than the more common approach of simply looking at rather meaninglesss correlation values alone.

Describing the DNN predictions as "climatology with added short-term variability" is not entirely wrong and would indeed be sufficient for many applications. But the added variability is based on meteorology where, e.g, decreased humidity results in lower soil moisture. Therefore, persistent meteorological anomalies result in persistent soil moisture changes (Figs. S6 to S8 of the new supplement) so that also the long-term variability goes beyond a pure climatology. Besides, already the construction of a globally consistent soil moisture climatology based on satellite data is not trivial, but the DNN obtains the climatological contribution automatically during training.

> -) I'd change the title of Sec. 4 to singular since only a single application is shown.

We propose two independent applications: replacing soil moisture parametrisations in climate models (on which we elaborate using the example of dust emission schemes) and reprocessing observations.

> -) I recommend to add a note of caution to the argument that the DNN could be used to fill gaps in the remotely sensed data. For example, tropical rain forest are masked out entirely in the ESA CCI SM. I am not sure if the DNN is able to properly learn the relation between the predictor variables and soil moisture in such a distinct regime if it is not represented in the training data set at all. A similar argument can probably made for the winter. If soil moisture cannot be retrieved because the soil is frozen or covered with snow, then I would not expect a DNN to properly turn precipitation, which is most likely in the form of snowfall, into accurate predictions of soil moisture.

We have added a note of caution to section 4.