Reply to RC3

We thank the referee for the valuable comments which contributed to an improved manuscript. In the following please find the point-by-point reply.

> General Comments:
> This paper presents a deep neural network model that predicts global surface soil moisture from precipitation, temperature, and humidity outputs from a climate model. The model was trained on daily satellite retrievals of soil moisture. The authors suggest two uses for the model: 1) to provide modeled soil moisture inputs for related applications, and 2) to fill missing values in satellite retrievals. The authors demonstrate an application by simulating threshold surface friction velocity for mineral dust emission in the Arabian Peninsula and Mesopotamia.
> I found this paper to be rigorous, complete, and convincing. While I have some questions, I think the overall quality is very good. The conclusions are well-founded, and the future research questions are well discussed.
> This paper could use some help from an English language editor. Some sections of the paper are very well written, and some have grammatical and language flaws. Even so, the paper is easy to read and understand.

Language copy-editing will be applied during production.

> Specific Comments:
> The authors present a simulation of threshold surface friction velocity for mineral dust emissions in the Arabian Peninsula and Mesopotamia as an application of the model. They simulate the threshold friction velocity using both observed and DNN-modeled soil moisture with good agreement. This leads me to ask: why use the DNN here at all? Why not just use the observations directly?

The observations have pixels with missing data, Figure 6 reduces their effect by considering the regional mean (and linearly interpolating over the remaining gaps), but a direct use of the observation in a dust simulation would require a proper gap-filling strategy. Secondly, the DNN can be used not only for nudged simulations of periods where observations are available, but also for free running simulations. We mention these motivations in the introduction.

> The authors say that Figure 6 shows "The results based on the observed and predicted soil moisture show good agreement and a strong seasonal cycle". More detail should be given here. The model appears to overpredict the threshold surface friction velocity a bit in the summer. Then, "whereas the result based on the EMAC soil water has little variability" – they could also compare the DNN to EMAC soil water directly.

We added a discussion of the slightly too high summertime thresholds. Another direct comparison of the EMAC and DNN soil moisture would add limited information, because Figs. 5 and Fig. 6 already demonstrate that the deviations are quite substantial. A major reason for these deviations is that the EMAC soil moisture also represents water in deeper soil layers, whereas the DNN was trained on the surface soil moisture, which is the relevant variable for surface processes. One purpose of Fig. 6 is to compare the threshold estimate that has been available in EMAC simulations with the new estimate based on the DNN which we propose to use instead. The advantage of a better representation of the moisture at the surface by the DNN is now emphasized.

> The authors should provide more information about the model selection criteria they used for the DNN input variables. They used a set of 18 input variables, all of which are intuitive. However, it would be interesting to see which of these input variables drive the predictive power of the model. This might be a particularly interesting question for the mineral dust application: what are the most important drivers of soil moisture in the Arabian Peninsula and Mesopotamia and what does this mean for vulnerability to dust storms? In regions where the temporal correlation is weaker, are the cos(2*pi*t/a) and sin(2*pi* t/a) terms dominating to impose the observed seasonal cycle?

We have added the motivation for the choice of predictors and show the sensitivity of example time series to the different predictors in the supplement (Figs. S6 to S8). While generally the analysis of the relevance of different parameters using DNNs is a promising approach and will certainly be focus of future studies, with the present setup one has to be careful because the predictors are not independent. For example, the results in the Middle East are not sensitive to a precipitation reduction, even though a causal relation surely exists. But the sporadic precipitation events hardly coincide in model and reality and make precipitation an unreliable predictor in this region, which the DNN compensates by making more use of the other predictors.

> Figure 5 shows a time series comparison of predicted and observed soil moisture at a single pixel during the test period. This pixel is located in Germany, where the model is reported to have strong temporal correlation (Fig. 4). What does the time series look like in a pixel with a poorer temporal correlation? What does is look like in a pixel in the poorly correlated region of the Arabian Peninsula?

We added a plot of a time series on the Arabian Peninsula with poor temporal correlation to the supplement (Fig. S1). The sparse and uncertain satellite observations are typical in the driest regions.

> Figure 7 shows the global distribution of observed and predicted volumetric soil moisture on two days in the training period. It would be very interesting to see similar plots for the test period.

With post-processing of observational data in mind, where both training and prediction are performed within the same time period, we present data from within the training period in Fig. 7. Note that even though the DNN is evaluated in the training period, the predictions of interest, i.e., predictions for grid cells without observations, are naturally not part of the training data. Of course the trained DNN can also complete observations in the test period and we provide example distributions in Fig. S9 in the new the supplement. However, the model is expected to perform poorer outside the training period so that this is not recommended to obtain best results for production data sets.

> Technical Corrections:
> There is quite a bit of model evaluation work in the “Applications” section. In particular, I think that the presentation and some of the discussion of Figures 5 and 7 could be moved up.

We moved the paragraph discussing the DNN result in Fig. 5 (and the new time series plots in the supplement) to the evaluation in section 3. The discussion of Fig. 7 remained in the application section because it illustrates the gap-filling application but does not rigorously compare predictions and observations.