## Topic editor comment

The manuscript presents a promising approach with thoughtful revisions that have addressed many of the initial concerns. However, a key issue remains unresolved and was raised by one of the reviewers regarding the justification for using ERA5-Land predictors. To ensure the robustness of the study's conclusions, an additional experiment excluding ERA5-Land is necessary to clarify whether performance improvements are driven by the modeling approach or the input data quality. A major revision is therefore warranted to address this methodological uncertainty.

Thank you for your time and helpful feedback. We have conducted the ablation experiment by excluding the ERA5-Land SM as suggested by the reviewer. Please refer to our detailed response to the reviewer for the experimental results and discussion.

## Report #2

Thank you for the authors' thoughtful responses to my previous comments. I agree with most of the clarifications and revisions made.

However, the justification for using ERA5-Land as a predictor remains insufficient. It is still unclear whether the improved model performance is primarily due to the advanced methods employed or the quality of the ERA5-Land dataset itself. To disentangle these effects, I strongly suggest that the authors conduct an additional experiment testing model performance without the ERA5-Land predictors. This would help assess the degree to which ERA5-Land contributes to the overall results and improve the robustness of the conclusions.

We sincerely appreciate your valuable suggestion. In response, we conducted an additional experiment by excluding the ERA5-Land SM from the input features and re-evaluated the performance of the AtLSTM model on the test set. The results show that the $R^2$ decreased from 0.987 to 0.954, and the RMSE increased from 0.011 $m^3\ m^{-3}$ to 0.020 $m^3\ m^{-3}$. Similarly, when the GLASS-AVHRR albedo and LST were removed from the inputs, the $R^2$ dropped to 0.968 and the RMSE increased to 0.018 $m^3\ m^{-3}$. These results demonstrate that both ERA5-Land SM and GLASS-AVHRR albedo and LST are critical to the performance of our long-term SM estimation model. By integrating multi-source datasets and leveraging their complementary strengths, the model achieves substantially improved accuracy on the test set. As data-driven approaches, the performance of ML and DL models is highly dependent on the quality of the input datasets. Nevertheless, the comparative analysis conducted in this study indicates that utilizing temporal information and optimizing model architectures also

play an important role in further enhancing the accuracy of the SM estimation model.

Additionally, we have added the following sentences to the revised manuscript: “To further investigate the importance of multi-source datasets for the performance of the AtLSTM model, we conducted ablation experiments by individually removing the ERA5-Land SM and the GLASS-AVHRR albedo and LST products from the input datasets. The results show that the AtLSTM model’s accuracy on the test set decreased significantly, with $R^2$ dropping to 0.954 and 0.968, and RMSE increasing to 0.020 $m^3$ $m^{-3}$ and 0.018 $m^3$ $m^{-3}$, respectively. These results demonstrate that by integrating multi-source datasets and leveraging their complementary strengths, the AtLSTM model cam achieve substantially improved accuracy in long-term SM estimation.” (P16, L400-406)