

CC2: 'Comment on essd-2024-553', SHAOBO SUN

General comments

The authors developed a consistent and seamless global soil moisture product over 1982-2021 using deep learning models and existing SM and auxiliary data. Validation showed that the SM data well capture temporal variability of global SM. Thus, the newly developed, long-term data show potential in wide applications. While, the manuscript needs to be improved.

We sincerely thank you for the positive and encouraging comments on our dataset and its potential applications. We also appreciate the suggestion to improve the manuscript. In response, we have revised the text to enhance its clarity and overall quality. Detailed responses and corresponding changes are provided below.

Specific comments

1. The authors did not use SM measurements to establish a DL model and develop global SM product. I think the developed GLASS-AVHRR SM product is a reprocessed SM product from ERA-Land SM and previous GLASS MODIS SM products. In the abstract, the authors said "we generated a consistent and seamless global SM product from 1982 to 2021 using deep learning (DL) by integrating four decades of Advanced Very High Resolution Radiometer (AVHRR) albedo and land surface temperature products with multi-source datasets."; did not mention use of ERA-Land SM which is much important for generating the SM data (Fig. A2). I think the authors should clearly stated how they generated the GLASS-AVHRR SM product with existing datasets in Abstract.

Thank you for pointing out the lack of clarity regarding the data sources used in generating the GLASS-AVHRR SM product. We have revised the abstract to more explicitly state the role of ERA5-Land SM and clarify the datasets used. The updated sentence in the abstract now reads:

“Here, we generated a consistent and seamless global surface SM product (0–5 cm) spanning 1982 to 2021 using a deep learning (DL) model. The model was trained with the GLASS-MODIS SM product and was designed to integrate four decades of Advanced Very High Resolution Radiometer (AVHRR)-derived albedo and land surface temperature, ERA5-Land SM, as well as terrain and soil texture datasets as input features.”

2. Page 3: "including land cover mapping, data fusion and downscaling, and environmental parameter retrieval (Yuan et al., 2020). " Please add citations for each

research area.

We have added citations as suggested. In particular, we now cite Huang et al. (2018) for land cover mapping, Wang et al. (2021) for data fusion and downscaling, and Ma and Liang (2022) for environmental parameter retrieval.

Huang, B., Zhao, B., and Song, Y.: Urban land-use mapping using a deep convolutional neural network with high spatial resolution multispectral remote sensing imagery, *Remote Sens. Environ.*, 214, 73–86, <https://doi.org/10.1016/j.rse.2018.04.050>, 2018.

Wang, F., Tian, D., Lowe, L., Kalin, L., and Lehrter, J.: Deep Learning for Daily Precipitation and Temperature Downscaling, *Water Resour. Res.*, 57, e2020WR029308, <https://doi.org/10.1029/2020WR029308>, 2021.

Ma, H. and Liang, S.: Development of the GLASS 250-m leaf area index product (version 6) from MODIS data using the bidirectional LSTM deep learning model, *Remote Sens. Environ.*, 273, 112985, <https://doi.org/10.1016/J.RSE.2022.112985>, 2022.

3. Table 1: Many other available long-term SM products were not included, such as the GLEAM SM.

Thank you for the suggestion. We have added two model-based long-term global SM products—GLEAM4 and SiTHv2—in Table 1. In addition, a brief description of these products has been added to the manuscript as follows:

“Recently, models that focus on the dynamic simulation of evapotranspiration and SM, such as the fourth generation of the Global Land Evaporation Amsterdam Model (GLEAM4) (0.1°, 1980–2023) and the Simple Terrestrial Hydrosphere model, version 2 (SiTHv2) (0.1°, 1982–2020), have also provided long-term global SM products by integrating multi-source satellite data and hydrometeorological variables (Miralles et al., 2025; Zhang et al., 2024a)”

Miralles, D. G., Bonte, O., Koppa, A., Baez-Villanueva, O. M., Tronquo, E., Zhong, F., Beck, H. E., Hulsman, P., Dorigo, W., Verhoest, N. E. C., and Haghdoost, S.: GLEAM4: global land evaporation and soil moisture dataset at 0.1° resolution from 1980 to near present, *Sci. Data*, 12, 416, <https://doi.org/10.1038/s41597-025-04610-y>, 2025.

Zhang, K., Chen, H., Ma, N., Shang, S., Wang, Y., Xu, Q., and Zhu, G.: A global dataset of terrestrial evapotranspiration and soil moisture dynamics from 1982 to 2020, *Sci. Data*, 11, 445, <https://doi.org/10.1038/s41597-024-03271-7>, 2024a.

4. Section 2 datasets: introductions on the datasets used in this section were verbose. I suggest the author simply simplify this section.

Thank you for the suggestion. We have simplified Section 2 by removing some less important sentences to improve clarity and conciseness.

5. Section 3 methods: please also simplify introductions on the DL/machine learning models, and pay much attentions on discussing results.

We have also simplified Section 3 by removing some less important sentences to improve clarity and conciseness.

6. Figure 9: In northern high latitudes, permafrost, snow and ice distribute widely, SM values in these regions are invalid in non-growing seasons. Thus, the SM maps in Jan. should masked these regions.

Thank you for the comment. Figure 9 is intended to present the spatial distribution of various SM products in their original forms for a direct visual comparison. Regions with permanent snow and ice as well as water bodies, where SM values are invalid, have now been masked out in all SM products using the MODIS land cover type product (MCD12C1).

However, masking regions covered by permafrost as well as seasonal snow and ice in the January SM maps would require reliable monthly land cover datasets, which are currently difficult to obtain. Therefore, we have clarified in the figure caption that “SM values in northern high latitudes in January should be interpreted with caution due to the widespread presence of permafrost, snow, and ice.”

7. Line 615: "model based on deep learning" - using DL models

Revised as suggested.

8. Page 28: “a topic of ongoing debate” - citations are needed.

Citations are added as suggested.

9. Discussion: I suggest the authors pay more attention on discussing their results, including uncertainties in the their developed SM data, and importance of the input variables (Fig. 2A). Specially, Fig. 2A shows that elevation exhibits largest influence on predicting SM. While, the ERA-Land SM had small important value. Why?

Thank you for raising this important point. In response, we have added a discussion on the uncertainties associated with our developed SM product in the Discussion section (Section 5). Specifically, we included the following statement:

“Meanwhile, as a data-driven product, the quality of the GLASS-AVHRR SM product largely depends on the selected input features, their accuracy and consistency, and the representativeness of the training data. Potential uncertainties may arise from biases or errors in the satellite and reanalysis inputs. In particular, the reduced model accuracy observed in the high SM range is likely due to the inherent imbalance in the

numerical distribution of SM samples and increased uncertainty in the accuracy of input features under wet surface conditions. In terms of feature selection, due to constraints such as record length, spatiotemporal completeness, and accuracy requirements, some informative but less consistently available variables may have been excluded, further contributing to the uncertainties in the final SM product. Moreover, as the ERA5-Land reanalysis SM was used as one of the input features, the generated product cannot be considered entirely independent. Future research could explore developing a fully independent, long-term, and seamless global SM product with sufficiently reliable accuracy.”

Regarding the importance of input features, we have added the following explanation to the manuscript to clarify the observed differences:

“In particular, although elevation is a static variable, it plays a critical role in shaping the spatial distribution of SM by influencing precipitation, temperature, vegetation type, and evaporation processes. Its impact on the spatial variability of SM tends to be more stable and consistent over time. In contrast, the contributions of dynamic input features such as ERA5-Land SM may fluctuate across time and space and can be diminished by inherent uncertainties and biases in the input data. Moreover, their importance may also be influenced by correlations with other input features.”

RC1: 'Comment on essd-2024-553', Anonymous Referee #1

This study creates a consistent long-term and high-resolution global soil moisture product with comparable accuracy to previous deep learning approaches and demonstrates a clear superiority of the AtLSTM approach over other compared approaches. The findings and datasets are valuable to the community. The reporting is clear. Overall I recommend it for publication after minor revisions. Please see below for comments:

Thank you very much for your positive feedback and valuable suggestions. Your insightful comments have helped us address several aspects of the manuscript that were previously unclear, greatly enhancing its clarity and rigor. We truly appreciate your time and effort in reviewing our work and improving its quality.

1. The depth of the developed dataset should be noted in the abstract.

Thank you for your suggestion. We have now included the depth of the developed dataset in the abstract to provide clearer information.

Specifically, we have revised the abstract to mention that: “Here, we generated a consistent and seamless global surface SM product (0–5 cm) from 1982 to 2021...”

Additionally, the depth of the dataset is also emphasized in other parts of the manuscript to ensure consistency.

2. The training target is a previously developed 1km dataset, and the newly developed dataset mainly provides the advantage of being longer-term and better winter coverage. This would be clearer if the temporal coverage of the Zhang et al. (2023) dataset is stated in p4 lines 117-118.

Thanks for the reminder. We have now added the temporal coverage of the GLASS-MODIS SM product (2000–2020) to distinguish it from the newly developed dataset.

3. p5 line 146: "8-day temporal resolution interpolated to daily" Please specify which interpolation method was applied.

We applied linear interpolation and have emphasized this in the manuscript.

4. It would be nice to add 1-2 sentences about uncertainty in the sampling strategy of the training target (p9 lines 242-p10 ln 244).

We appreciate your advice and have included an analysis of the uncertainty of the sampling strategy, as follows: “While the three years were selected to represent different periods within the available time span (2000–2020), this selection may

introduce some uncertainty, as climate and environmental conditions can vary annually, and extreme weather or climate events in certain years may affect the representativeness of variables such as LST and SM. Nevertheless, this approach was adopted to control the sample size while ensuring the representativeness of samples across different years.”

5. p13 Table 4: Should the number of layers be "2" for Bi-LSTM and 1 elsewhere?

Thanks for pointing out this mistake. Upon rechecking the code and experimental notes, we found that the “number of layers” for all DL models is 1. This has been corrected in Table 4.

Following the standard convention in deep learning frameworks (e.g., PyTorch), in our Bi-LSTM model, “Number of layers = 1” refers to a single Bi-LSTM layer, which internally includes both forward and backward LSTM units.

We realized that the relevant sentence in the manuscript caused some confusion and have revised it from “The Bidirectional LSTM (Bi-LSTM) extends the LSTM network by using two separate LSTM layers to process the input sequence from both forward and backward directions, and then concatenating the outputs of both layers” to “The Bidirectional LSTM (Bi-LSTM) extends the LSTM network by incorporating both forward and backward LSTM units within a single layer, allowing the model to capture contextual information from both directions before concatenating their outputs.”

6. The sequence lengths in Table 4 span whole year. The testing on p17 only spans 0-29 days and demonstrates stabilized performances at much shorter lengths than 29 days. Why the drastic increase in sequence lengths in production runs? Also, since the input data are 3 discrete years, please specify what values are used to pad the 60 days before the start and after the end of the whole year.

Given the need to generate daily SM products over 40 years, using an input sequence length of 10–30 days per run and generating SM estimates for such short periods would significantly increase the time required for data preprocessing. To balance accuracy and efficiency, we extended the sequence length to 425 days, enabling the generation of a full year's product in a single run.

During both the training and production phases, the first and last 30 days of each 425-day sequence were padded with actual data from adjacent years to ensure consistency. We have clarified this point in the manuscript.

7. p14 lines 366-367: A bit confusing because this reduced-sample testing was not described in the Methods. Could you add a description?

We have added a description of the purpose of this reduced-sample experiment in the

manuscript and revised the relevant sentence as follows: “The fairly high overall accuracy of the benchmark XGBoost model may be attributed to the large number of training samples, specifically 135,360 pixels per day over 3 years, evenly distributed across the globe on a daily basis. To evaluate the impact of sample size on model performance, we conducted an experiment by reducing the number of training samples. When the sample size was reduced by a factor of 100, the accuracy of the XGBoost model dropped considerably, with an R^2 of 0.96 and RMSE of $0.017 \text{ m}^3 \text{ m}^{-3}$ on the test set. This highlights the importance of having sufficient samples to achieve high accuracy with XGBoost and indicates the advantage of using the GLASS-MODIS SM product as training target, which can provide much richer samples than the sparse in situ ISMN SM dataset.”

8. p15 Table 5: Is there a specific reason for choosing $0.4 \text{ m}^3/\text{m}^3$ as a threshold for large SM values?

The threshold of $0.4 \text{ m}^3 \text{ m}^{-3}$ was chosen because soil moisture values exceeding this level are generally considered high in most global land areas, except for wetland and irrigated regions. This threshold enables a more detailed assessment of model performance under high soil moisture conditions, where sample availability is relatively limited. This pattern can also be observed in the density scatter plots shown in Fig. 3, where the red dots, indicating the highest density, are primarily concentrated within the $0\text{--}0.4 \text{ m}^3 \text{ m}^{-3}$ range.

9. Fig. 6: COSMOS dataset is not directly comparable to the developed dataset due to significant depth differences. It is okay as an auxiliary comparison, given that other comparable independent validation datasets are used (1982-1999 ISMN and SMAP validation sites). However, for informative purpose, it would be good to provide boxplots of the depths of the COSMO observations used in the last three columns of Fig. 6 in supplementary info.

Following your suggestion, we have summarized the sensing depths of all sensors across the three COSMOS networks and added a boxplot (Fig. A2) in the Appendix. We have also included relevant descriptions in the main text.

“The distribution of sensing depths for each station across the three COSMOS networks is presented in Fig. A2. While COSMOS sensors measure SM at relatively deeper layers, they have been used to validate microwave and modelled surface SM products and show good correlations with them (Montzka et al., 2017; Peng et al., 2021b).”

Montzka, C., Bogena, H. R., Zreda, M., Monerris, A., Morrison, R., Muddu, S., and Vereecken, H.: Validation of Spaceborne and Modelled Surface Soil Moisture Products with Cosmic-Ray Neutron

Probes, Remote Sens., 9, <https://doi.org/10.3390/rs9020103>, 2017.

Peng, J., Tanguy, M., Robinson, E. L., Pinnington, E., Evans, J., Ellis, R., Cooper, E., Hannaford, J., Blyth, E., and Dadson, S.: Estimation and evaluation of high-resolution soil moisture from merged model and Earth observation data in the Great Britain, Remote Sens. Environ., 264, 112610, <https://doi.org/https://doi.org/10.1016/j.rse.2021.112610>, 2021.

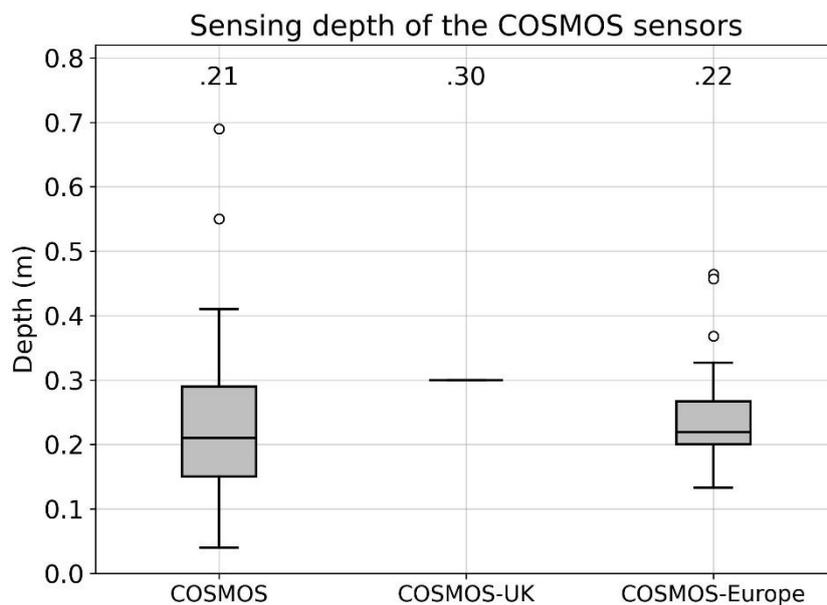


Figure A2 Boxplots of sensing depths across the three COSMOS networks used for validation

10. More on Fig. 6. The currently used 1982-1999 ISMN set provides spatiotemporally independent comparison. The 1982-1999 data at the 715 representative stations would still be temporally independent comparison. Could you include the performance metrics on this subset of data?

Thank you for your suggestion. However, the vast majority of ISMN SM monitoring stations were established after 2000. Among the 715 representative stations, only 23 had observations before 2000. Comparing the metrics obtained from these two groups of datasets may not be sufficient to draw robust conclusions. Therefore, we opted not to include this comparison in Fig. 6. Nevertheless, we assessed the validation metrics of the GLASS-AVHRR SM product across the 23 representative stations from 1982 to 1999, yielding a median R of 0.76, a median bias of $-0.03 \text{ m}^3 \text{ m}^{-3}$, and a median ubRMSE of $0.029 \text{ m}^3 \text{ m}^{-3}$.

RC2: 'Comment on essd-2024-553', Anonymous Referee #2

This study used several deep learning models to estimate SM with some optical remote sensing indices, soil properties and DEM. Even though it's a big work, it lacks sufficient innovation and employs inappropriate evaluation method. Despite the overall scale of the work, the methodological design raises serious concerns.

We sincerely thank you for your time and effort in reviewing our manuscript and for recognizing the scale of our work. With due respect, we would like to express our disagreement with the comment that the study lacks sufficient innovation and adopts inappropriate evaluation method.

In terms of methodological innovation, our study explored state-of-the-art deep learning models to generate a seamless, four-decade global SM product at 5 km resolution. By integrating multi-source datasets, this product achieves complete spatiotemporal coverage, reliable accuracy, rich spatial details, and low biases. Given the limited research on developing long-term global SM products based on deep learning, along with the notable strengths of our product, we believe this work meets the journal's requirements for innovation and high-quality datasets.

Regarding the evaluation strategy, we carefully designed the validation process by incorporating multiple methods to assess the performance of our model and product. During the model development phase, we used the GLASS-MODIS SM product—also employed as the training target—as a reference to evaluate the model's estimated SM on the test set. The high accuracy achieved indicates the strong agreement between the estimated and target SM values, and demonstrates the reliability of the GLASS-AVHRR product as a temporal extension of the GLASS-MODIS product. During the product accuracy evaluation phase, we fully evaluated the product's accuracy using independent in situ SM datasets across different spatial scales, including point-scale ISMN stations, field-scale COSMOS networks, and 9-km upscaled SMAP core validation site, as detailed in Section 4.2. In addition, time series comparisons and spatial consistency analyses with other long-term global SM products provided further evidence of the robustness and reliability of our product.

We hope these clarifications help to alleviate your concerns and highlight the novel aspects and methodological soundness of our work.

Major Comments:

Line 135, Table 2: I question the innovation and scientific contribution of using ERA5-Land SM as an input variable in a model that aims to estimate soil moisture. This

approach may introduce circular reasoning and undermines the novelty of the proposed method.

We sincerely appreciate your insightful comment and understand your concern regarding potential circular reasoning—specifically, the risk that using ERA5-Land SM as an input while also estimating SM might lead the model to simply reproduce the ERA5-Land SM, rather than genuinely learning from multi-source datasets.

We would like to clarify that ERA5-Land SM was not used as a target variable, nor was the intention to replicate its pattern. Instead, it served as one of several input features—alongside satellite observations, terrain, and soil properties—aimed at improving the model’s ability to capture the complex spatiotemporal variability of SM. Furthermore, our final product was fully validated using independent in situ SM datasets (ISMN, COSMOS, and SMAP CVSs), and the results showed that the biases in our product were significantly lower than those in the ERA5-Land SM product. As such, our approach focuses on multi-source data fusion, leveraging the strengths of various datasets, rather than merely “using SM to predict SM.” Therefore, we believe that the use of ERA5-Land SM as an input does not undermine the novelty of our approach or lead to circular reasoning.

Accordingly, we explicitly clarified the objective of this study in the manuscript:

“(1) To develop a DL-based global SM estimation model by integrating multi-source datasets and leveraging their complementary strengths, so as to derive a seamless and reliable long-term global SM product.”

Line 380, Table 5: It is inappropriate to use GLASS-MODIS SM as a reference to evaluate the model's estimated SM, as they are not independent. Since variables such as LST and albedo were also used in the model estimation, the evaluation becomes biased, leading to inflated performance metrics (e.g., $R^2 > 0.98$). This dependence compromises the reliability of the validation.

Thank you for pointing out this important concern. We would like to clarify that the use of GLASS-MODIS SM as a reference in Table 5 was solely part of the model development process—specifically, to assess how well the models could reproduce the training target on a held-out test set. This evaluation is a standard practice in deep learning studies to verify model learning behavior before proceeding to independent validation.

Table 5 aims to compare the performance of different DL models on the test set in order to select the model with the highest accuracy. The high R^2 values reported here reflect the models’ fit to the training target under consistent input settings, but they do

not represent the final product's accuracy. In contrast, the accuracy of our final product was thoroughly evaluated in Section 4.2 using independent in situ SM datasets across different spatial scales.

To avoid any potential misunderstanding that the model's accuracy on the test set represents the actual accuracy of the GLASS-AVHRR SM product, we have removed the reference to model accuracy in the abstract and revised the relevant text as follows: "Our results show that all four DL models outperformed the benchmark XGBoost model, with the AtLSTM model achieving the highest accuracy on the test set, particularly at high SM levels ($> 0.4 \text{ m}^3 \text{ m}^{-3}$)."

Line 20: The five deep learning models yielded very similar results, with R^2 values ranging from 0.982 to 0.987. Therefore, the claim that these models "...effectively enhance..." SM estimation is not well supported.

Thank you for your valuable comments. We fully understand your concern regarding the similarity in overall accuracy among the five models. However, it is important to point out that the AtLSTM model significantly improved SM estimation accuracy under high SM conditions. Specifically, on the test set, the XGBoost model exhibited notable underestimation at high SM levels ($> 0.4 \text{ m}^3 \text{ m}^{-3}$), yielding an R^2 of 0.413 and RMSE of $0.022 \text{ m}^3 \text{ m}^{-3}$. In contrast, by incorporating both temporal information and attention mechanisms, the AtLSTM model effectively reduced this underestimation and achieved an R^2 of 0.621 and RMSE of $0.016 \text{ m}^3 \text{ m}^{-3}$ within the same range, demonstrating superior modeling capability under these challenging conditions.

Therefore, although the overall R^2 values across these models appear similar, we believe that this significant improvement at high SM levels by the AtLSTM model supports the claim that temporal and attention-based modeling can effectively enhance the estimation accuracy of SM. To improve clarity, we have revised the relevant statement in the abstract to: "These results suggest that under some challenging conditions, utilizing temporal information as well as adding an attention module can effectively enhance the estimation accuracy of SM."

Minor Comments:

Line 60: There are additional remote sensing-based SM products that should be referenced, including but not limited to: Cheng et al. (2023); Guevara, Taufer, & Vargas (2021); and Zheng, Jia, & Zhao (2023).

Thank you for the valuable information. We have added citations to Cheng et al. (2023) and Guevara et al. (2021) in Page 3 of the revised manuscript, respectively. Zheng et al. (2023) had already been cited in the same page.

Line 190: The time period referred to as "period1" is 2000–2018. Why were the years 2019–2021 excluded from the analysis?

The period 2000–2018 was intentionally selected to ensure consistency with the validation period used in our previous paper on the GLASS-MODIS SM product, as illustrated in Fig. 5 of Zhang et al. (2023). In this work, we also used the GLASS-MODIS SM validation results at ISMN stations as a benchmark for comparison. Aligning the time period helps avoid potential confusion for readers regarding any differences in the reported performance of the same product between the two studies.

Zhang, Y., Liang, S., Ma, H., He, T., Wang, Q., Li, B., Xu, J., Zhang, G., Liu, X., and Xiong, C.: Generation of global 1 km daily soil moisture product from 2000 to 2020 using ensemble learning, *Earth Syst. Sci. Data*, 15, 2055–2079, <https://doi.org/10.5194/essd-15-2055-2023>, 2023.

Line 355: The statement “two widely used long-term global SM products” should specify which products are being referred to for clarity.

Thank you for the comment. We have revised the sentence to specify the two SM products, namely the ERA5-Land and ESA CCI.

Cheng, F., Zhang, Z., Zhuang, H., Han, J., Luo, Y., Cao, J., . . . Tao, F. (2023). ChinaCropSM1 km: a fine 1 km daily soil moisture dataset for dryland wheat and maize across China during 1993–2018. *Earth System Science Data*, 15(1), 395–409. doi:10.5194/essd-15-395-2023

Guevara, M., Taufer, M., & Vargas, R. (2021). Gap-free global annual soil moisture: 15 km grids for 1991–2018. *Earth System Science Data*, 13(4), 1711–1735. doi:10.5194/essd-13-1711-2021

Zheng, C., Jia, L., & Zhao, T. (2023). A 21-year dataset (2000–2020) of gap-free global daily surface soil moisture at 1-km grid resolution. *Sci Data*, 10(1), 139. doi:10.1038/s41597-023-01991-w

RC3: 'Comment on essd-2024-553', Anonymous Referee #3

The authors compared multiple deep learning methods and selected the most suitable approach to generate a 5 km-resolution soil moisture product based on AVHRR data, spanning four decades. Overall, this is a well-written manuscript, and the scope of the study aligns well with the aims and scope of Earth System Science Data. However, the current version of the manuscript appears overly technical in its presentation. The authors should address the following major and specific comments before the manuscript can be considered for publication.

We sincerely thank you for the positive feedback and constructive suggestions. We have revised the manuscript to reduce technical complexity and improve readability. Below are our point-by-point responses to all major and specific comments.

General comments

1. While the description of the deep learning methods is sufficiently detailed, the treatment of the predictor variables requires substantial clarification and improvement. It remains unclear how these variables were selected, particularly the rationale for using ERA5-Land data as model inputs. In my view, the relatively strong performance of GLASS-AVHRR in capturing seasonal variations, especially for the R metric, may be partially attributed to this. The authors should consider producing an independently driven soil moisture product, rather than relying on predictors that may introduce redundancy or circular reasoning.

These input variables are widely used in ML and DL-based SM estimation studies. Their selection was guided by prior literature as well as our experience from previous SM product development. Due to practical constraints such as the need for long temporal sequences (40 years), consistent spatiotemporal coverage, and reliable accuracy, some potentially informative variables were excluded.

The ERA5-Land SM product was selected because of its long-term temporal coverage, physical consistency, and relatively high validation accuracy (e.g., in terms of R). Moreover, The ERA5-Land SM product has been widely adopted in previous SM downscaling studies, which supports its credibility and applicability within our modeling framework.

We acknowledge that the relatively strong performance of the GLASS-AVHRR SM product may be partially attributed to the inclusion of ERA5-Land SM as an input. As shown in the feature importance ranking (Fig. A3), ERA5-Land SM contributes significantly to the model's predictions. However, other variables—such as elevation,

black-sky visible albedo, and slope—also play important roles, indicating that the model integrates diverse sources of information beyond any single input.

Our goal was to generate a seamless and reliable long-term global SM product by integrating multi-source datasets and leveraging their complementary strengths. While we acknowledge that some degree of redundancy may exist, the AtLSTM model is capable of effectively handling this through its attention mechanism, which allows it to dynamically focus on the most relevant features over time.

We understand the concern regarding potential circular reasoning—specifically, the risk that using ERA5-Land SM as an input while also estimating SM might lead the model to simply reproduce the ERA5-Land SM, rather than genuinely learning from multi-source datasets. We would like to clarify that ERA5-Land SM was not used as a target variable, nor was the intention to replicate its pattern. Instead, it served as one of several input features—alongside satellite observations, terrain, and soil properties—aimed at improving the model’s ability to capture the complex spatiotemporal variability of SM.

We appreciate the suggestion to explore the development of an independently driven SM product and will consider this direction in future work. However, we believe the current design is appropriate for our objective of generating a long-term global SM product with complete spatiotemporal coverage, reliable accuracy, rich spatial details, and low biases.

In response, we have revised the manuscript to provide a clearer explanation of our variable selection rationale and the role of ERA5-Land SM. Specifically, we added the following sentences:

“These input features are widely used in ML and DL-based SM estimation studies.”

“In terms of feature selection, due to constraints such as record length, spatiotemporal completeness, and accuracy requirements, some informative but less consistently available variables may have been excluded, further contributing to the uncertainties in the final SM product.”

“(1) To develop a DL-based global SM estimation model by integrating multi-source datasets and leveraging their complementary strengths, so as to derive a seamless and reliable long-term global SM product;”

“Moreover, as the ERA5-Land reanalysis SM was used as one of the input features, the generated product cannot be considered entirely independent. Future research could explore developing a fully independent, long-term, and seamless global SM product with sufficiently reliable accuracy.”

2. I remain concerned about the generalization capability of the developed deep learning model, particularly considering the strong spatial autocorrelation inherent in the current training, validation, and test data split. Furthermore, the results presented in Table 5 indicate that model performance deteriorates when soil moisture exceeds $0.4 \text{ m}^3/\text{m}^3$, which the authors attribute to a lack of training samples. This observation indirectly suggests that the model performs poorly in unseen regions or conditions, highlighting its limited generalization ability. The authors should explore alternative data splitting strategies to rigorously assess the robustness and generalizability of their approach.

We acknowledge your concern regarding the generalization capability of our DL model. In this study, the training, validation, and test datasets were constructed by randomly sampling from 135,360 pixels distributed across the globe, each representing a time-series sample with spatial independent locations. Specifically, these pixels were selected at an interval of five grid cells (i.e., 25 km), ensuring that the spatial distance between any two samples exceeds 25 km. Given that the spatial autocorrelation of soil moisture typically diminishes beyond this scale, the spatial dependence among samples is sufficiently minimized. Therefore, we believe that the current data splitting strategy can ensure a reliable evaluation of the model's generalization performance.

We agree that the model's performance declines when SM exceeds $0.4 \text{ m}^3 \text{ m}^{-3}$. This is mainly due to the inherent imbalance in the distribution of SM observations, as samples at high SM levels are naturally scarce across both time and space. In addition, satellite observations under wet conditions tend to be less reliable because of saturation effects, cloud contamination, and decreased sensitivity in albedo, all of which affect the quality of the input features. Despite these challenges, the AtLSTM model still achieves much better performance than XGBoost in this range, as demonstrated by significantly improved R^2 and RMSE values. In general, the model performs well across the majority of the SM range, with only slight differences in overall accuracy between the spatially independent training and test sets, indicating relatively robust generalization capability.

In response, the following sentences have been added to the revised manuscript:

“These samples were then randomly divided into training, validation, and test datasets in the ratio of 7:2:1 based on their locations, ensuring spatial independence with distances between any two samples exceeding 25 km, thereby minimizing the influence of spatial autocorrelation.”

“In particular, the reduced model accuracy observed in the high SM range is likely due to the inherent imbalance in the numerical distribution of SM samples and increased uncertainty in the accuracy of input features under wet surface conditions.”

3. Figure 2 in the Methods section is unnecessarily technical and lacks clear justification.

A concise textual explanation that hyperparameter tuning was performed would be sufficient. Moreover, both Figures 1 and 2 fail to clearly illustrate how the spatiotemporal training was implemented, especially considering that LSTM and related models are specifically designed to capture temporal dependencies. The authors should provide a more explicit and structured explanation of how time-series characteristics were incorporated into the model training, without ignoring the spatial autocorrelation issue raised above.

Thank you for your valuable comments. In response, we have added a new schematic diagram (Fig. 2) to explicitly illustrate how spatiotemporal sampling was conducted from the input dataset to construct time-series input samples that are evenly distributed across the globe. Specifically, spatial sampling was performed at an interval of 5 pixels, ensuring that the minimum distance between any two samples exceeds 25 km. As a result, the training and test samples are spatially independent, and the potential influence of spatial autocorrelation is expected to be minimal.

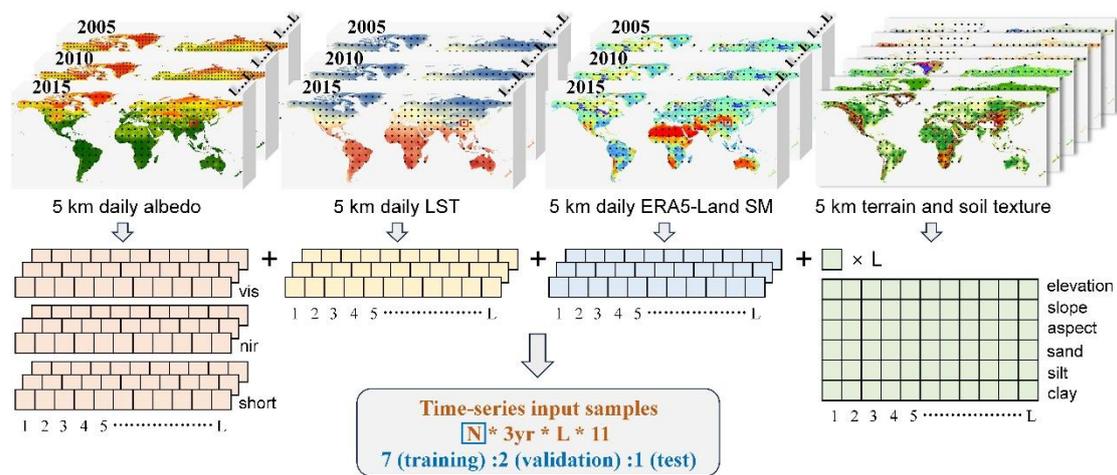


Figure 2. Schematic diagram illustrating the construction of time-series input samples for the DL models. N denotes the total number of samples, and L represents the sequence length, a hyperparameter that needs to be tuned.

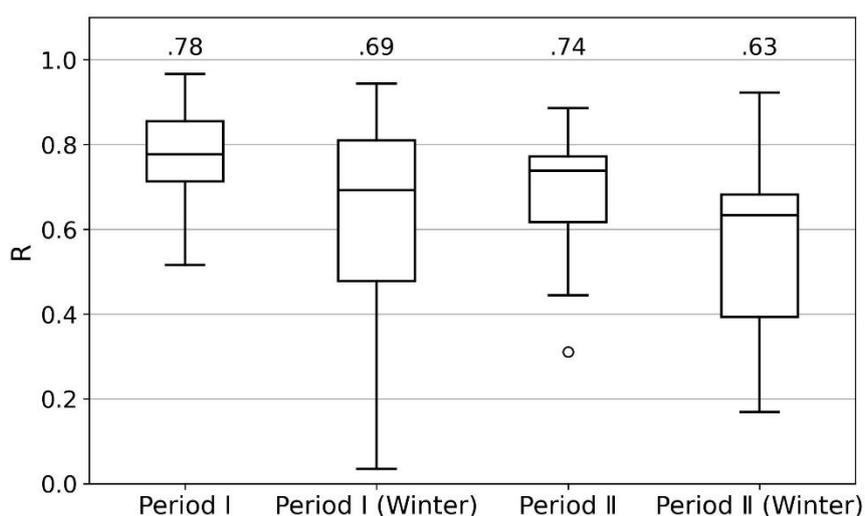
4. In Figure 9, I noticed that the GLASS-AVHRR estimates provide coverage during the winter season, unlike the CCI product. The authors should clarify what measures were taken to ensure the reliability of these estimates under winter conditions, particularly in areas covered by snow, ice, or water. An additional seasonal evaluation focusing specifically on winter performance would be welcome and could strengthen the credibility of the product. In addition, all products in Figure 9 should be masked over permanent water bodies, where soil moisture retrieval is not valid. This issue is especially noticeable in Quebec, Canada.

Thank you for the comment. The current algorithm is not able to reliably estimating

SM under areas covered by snow, ice, or water. We have masked permanent snow and ice as well as water bodies in the final GLASS-AVHRR SM product. However, seasonal snow and ice were not masked due to the lack of high-quality global products. Users are encouraged to apply additional filters based on their specific application needs—for example, by masking pixels with LST below 0 °C during the winter season.

In response to your suggestion, we conducted an accuracy assessment of the GLASS-AVHRR SM product during the winter season, using in situ measurements from ISMN stations. The COSMOS SM dataset was not included in this evaluation due to its relatively high uncertainty under wetter soil conditions, which are prevalent in winter. Specifically, we selected ISMN stations located above 30°N latitude and extracted daily observations from December to February. These observations were matched with the GLASS-AVHRR SM estimates, and only stations with more than 100 matched records were retained for analysis. The R values between the two datasets at these stations were then calculated, and the results are presented in the boxplot below.

Based on the station selection criteria described above, the GLASS-AVHRR SM product achieved a median R value of 0.78 at 374 representative ISMN stations during Period I (2000–2018), and 0.74 at 19 stations during Period II (1982–1999). In the winter season (December to February), the corresponding median R values were 0.69 for Period I and 0.63 for Period II. While the product’s performance is relatively lower in winter, it still offers valuable estimates—particularly given the substantial data gaps in the CCI SM product during the winter season in regions above 30°N latitude. This result has been incorporated into Section 4.3 of the manuscript (Page 26).



We have also revised Figure 9 as suggested. In the updated figure, permanent snow and ice as well as water bodies have been masked out in all products using the MODIS land cover type product (MCD12C1). These changes can be clearly seen from the

ERA5-Land and GLASS-MODIS SM products. However, due to the coarse display scale of Figure 9, many masked lakes (such as those in Quebec) are not visually discernible. As shown in the zoomed-in maps in Figure 10 over the Tibetan Plateau, the lakes have been effectively masked.

5. The authors should provide more reasonable explanations for some of the evaluation results. For example, why does GLASS-AVHRR perform worse than ERA5-Land in terms of the R metric over COSMOS-Europe? In addition, I do not fully agree with the explanation provided for the strong dry bias observed in COSMOS-UK. ERA5-Land shows a consistently wet bias across different COSMOS networks, which raises the question of whether the developed product itself suffers from instability in different regions.

ERA5-Land is a reanalysis product developed by ECMWF that assimilates and calibrates against a large number of in situ observations worldwide, with particularly dense and high-quality observational networks in Europe. As a result, it achieves notably high accuracy over the COSMOS-Europe network, with a median R value of 0.85. The slightly lower performance of the GLASS-AVHRR SM product in this network may be attributed to two main factors. First, our model is trained globally and thus may have learned generalized patterns that are not specifically optimized for European conditions. Second, while integrating multiple data sources, such as AVHRR LST, albedo, and topographic information, can improve overall global performance, it may introduce noise or redundancy in regions where ERA5-Land already provides highly accurate estimates, leading to reduced correlation metrics.

We realized that the original sentence—“Besides, the biases of the GLASS-AVHRR SM product across the COSMOS networks were much larger than those observed across the ISMN network, particularly on the COSMOS-UK network, where the median bias reached $-0.09 \text{ m}^3 \text{ m}^{-3}$ ”—is inappropriate. In fact, the large bias was primarily observed on the COSMOS-UK network. We examined the sensing depths across the three COSMOS networks and found that the COSMOS-UK network has a median sensing depth of 30 cm, which is much deeper than those of COSMOS (21 cm) and COSMOS-Europe (22 cm) (Fig. A2). This greater sensing depth likely contributes to the large bias observed in the COSMOS-UK network. Moreover, COSMOS measurements in organic soils or humid regions—conditions more prevalent in the UK—are subject to greater uncertainty, as also noted by Zheng et al. (2024). This may explain the higher ubRMSE values observed for both the GLASS-AVHRR and ERA5-Land SM products on the COSMOS-UK network.

Therefore, we have revised the relevant sentences to: “The biases of the GLASS-

AVHRR SM product on the COSMOS-UK network were much larger than those on the other two COSMOS networks, with the median bias reaching $-0.09 \text{ m}^3 \text{ m}^{-3}$. This is likely due to the greater sensing depth of the COSMOS-UK network, which has a median depth of 30 cm, compared to 21 cm and 22 cm for the COSMOS and COSMOS-Europe networks, respectively (Fig. A2). Moreover, both the GLASS-AVHRR and ERA5-Land SM products exhibited larger ubRMSE values on the COSMOS-UK network. This may be related to the increased uncertainty of COSMOS measurements in organic soils or humid regions, which are prevalent in the UK, as also reported by Zheng et al. (2024).”

Zheng, Y., Coxon, G., Woods, R., Power, D., Rico-Ramirez, M. A., McJanet, D., Rosolem, R., Li, J., and Feng, P.: Evaluation of reanalysis soil moisture products using cosmic ray neutron sensor observations across the globe, *Hydrol. Earth Syst. Sci.*, 28, 1999–2022, <https://doi.org/10.5194/hess-28-1999-2024>, 2024.

Specific comments:

L61: This may also be related to the uneven spatial distribution of the input meteorological data, particularly the limited observational coverage in tropical regions.

Thank you for the helpful suggestion. We have revised the sentence as follows:

“Yet, these SM products may suffer from large uncertainties arising from defective forcing data, imperfect model parameterization, and the uneven spatial distribution of input meteorological observations, particularly the limited observational coverage in tropical regions.”

L64: also [10.1016/j.rse.2022.112921](https://doi.org/10.1016/j.rse.2022.112921); [10.1016/j.rse.2022.113272](https://doi.org/10.1016/j.rse.2022.113272)

We have added the recommended references to the manuscript.

2 Datasets: Predictor variables lack of vegetation-related variables

We tested several AVHRR-derived NDVI and LAI products as additional predictor variables; however, they did not improve the model performance and in some cases even slightly degraded it. This may be due to the relatively high level of noise and uncertainty in the long-term AVHRR vegetation products, particularly over regions with sparse vegetation or frequent cloud cover.

Table 2: As noted above, it is inappropriate to use ERA5-Land-derived soil moisture as an input when the objective is to analyze or generate a new soil moisture product.

As mentioned in our response to the first general comment, we have explained the reason for choosing the ERA5-Land SM product. In addition, we have clarified the main objective of this study at the end of the Introduction section, which is:

“(1) To develop a DL-based global SM estimation model by integrating multi-source datasets and leveraging their complementary strengths, so as to derive a seamless and reliable long-term global SM product.”

L146: What kind of method?

We applied linear interpolation and have emphasized this in the manuscript.

L156: see 10.1016/j.rse.2023.113721

Thank you for providing this reference. We have included it in the manuscript.

L175: This also means that validation with ISMN is not independent.

For Period I (2000–2018), the ISMN stations used for validation are not completely independent, and we have clarified this point in the text. However, the 45 ISMN stations for Period II (1982–1999) are fully independent.

L196: The meaning of the label "G" should be clearly defined.

We revised the sentence to “considering only those flagged as “G” for good quality.”

L201: How do the authors justify the appropriateness of using COSMOS observations, which primarily sense deeper soil moisture, for evaluating surface soil moisture products?

We acknowledge that COSMOS sensors primarily measure SM at deeper layers, with the median sensing depths of the three COSMOS networks used for validation in this study being 21 cm, 30 cm, and 22 cm, respectively (Fig. A2). Nevertheless, several studies, such as Montzka et al. (2017) and Peng et al. (2021), have also utilized the COSMOS SM datasets to validate microwave and modelled surface SM products, showing good correlations between COSMOS observations and these products. We have added this point into the manuscript (Page 8).

Moreover, in our evaluation, we used COSMOS data cautiously by analyzing them separately from other datasets and clarifying their sensing depth characteristics in the manuscript. This ensures that the use of COSMOS data provides complementary information without misleading the evaluation of surface SM products.

Montzka, C., Bogena, H. R., Zreda, M., Monerris, A., Morrison, R., Muddu, S., and Vereecken, H.: Validation of Spaceborne and Modelled Surface Soil Moisture Products with Cosmic-Ray Neutron Probes, *Remote Sens.*, 9, <https://doi.org/10.3390/rs9020103>, 2017.

Peng, J., Tanguy, M., Robinson, E. L., Pinnington, E., Evans, J., Ellis, R., Cooper, E., Hannaford, J., Blyth, E., and Dadson, S.: Estimation and evaluation of high-resolution soil moisture from merged model and Earth observation data in the Great Britain, *Remote Sens. Environ.*, 264, 112610,

<https://doi.org/https://doi.org/10.1016/j.rse.2021.112610>, 2021.

Table 3: Are these COSMOS networks intercalibrated?

Currently, we have not found any literature or reports indicating that these COSMOS networks were intercalibrated. However, we conducted independent validations for data from different COSMOS networks separately without merging the datasets, so this does not affect the main findings of the validation.

Sections 3.3 and 3.4 can be simplified

Thank you for the suggestion. We have simplified Sections 3.3 and 3.4 by removing some less important sentences to improve clarity and conciseness.

L377: The term "underestimate" should be used with caution, as all reference datasets represent proxy observations rather than true ground-truth measurements at the grid scale.

We agree with your comment and acknowledge that the term "underestimate" should be used with caution. Accordingly, we have revised the term "underestimate" in this section to "lower estimates" to avoid potential misinterpretation.

Figure 3: Lack of units

Thank you for pointing this out. The units have now been added to all RMSE values in Figure 3.

L506-507: Why doesn't ERA5-Land have this problem?

Both the GLASS-AVHRR and ERA5-Land SM products are affected by the differences in spatial scale, vertical sensing depth, and local soil characteristics when validated against the COSMOS networks, especially in terms of bias and ubRMSE metrics. Notably, both products exhibit the highest ubRMSE values on the COSMOS-UK network. The biases in the ERA5-Land product, however, stem from a variety of factors and can vary considerably across different regions, making it difficult to determine the major cause of the bias magnitude.

Figure 6: These three datasets, including ERA5-Land, GLASS-MODIS and GLDAS-AVHRR, have different spatial resolutions. Please clarify how the spatial mismatch was addressed when using different in-situ observations.

We understand the concern that spatial mismatches between gridded SM products and in-situ observations may affect the accuracy assessment. However, most of the in-situ SM measurements from the ISMN dataset are collected from sparse sites, where typically only a single station is located within each product's grid cell. In such cases,

spatial averaging is not feasible, and the point-scale measurements are directly used to represent the grid cell for validation. To help mitigate the impact of spatial mismatch, we also conducted accuracy assessments at representative ISMN stations. As shown in Figure 6, all SM products exhibit improved accuracy at these representative sites.

Compared to ISMN stations, COSMOS sensors are capable of providing area-averaged SM measurements over a much larger footprint, typically with a radius of 130 to 240 meters. This characteristic makes them more compatible with the spatial resolution of gridded products, thereby reducing the potential impact of spatial mismatch during the validation process.

We have added the following statement in the manuscript to clarify the potential effects of spatial mismatch on the validation metrics: “This variability may be attributed to the different footprint radii of COSMOS sensors, which result in varying degree of spatial representativeness and spatial mismatches with gridded SM products. These factors can introduce uncertainty into the validation results, particularly affecting the bias and ubRMSE metrics.”

L529: Why doesn't ERA5-Land have this problem?

Due to the limited number and uneven distribution of in situ stations within each CVS, the upscaled 9 km CVS SM dataset may contain considerable uncertainty at sites with complex land surface conditions, such as the Little River. Meanwhile, the 2×2 window averaging used to align the 5 km GLASS-AVHRR product with the CVS grid at such site could introduce additional errors due to subpixel heterogeneity. Therefore, using this site for validation is less reliable. Although the ERA5-Land product yields a higher R value at the Little River site, it also exhibits the largest bias ($0.22 \text{ m}^3 \text{ m}^{-3}$), which is also evident from the time series plot.

We have revised the corresponding sentence to: “This is probably because the land cover type at this site is ‘Cropland or Natural mosaic’, making the upscaled in situ SM measurements less representative and the validation results at this site less reliable.”

Figure 10: All products should be masked out over lake areas where soil moisture retrievals are not valid.

We have revised Figure 10 as suggested. In the updated figure, permanent snow and ice as well as water bodies have been masked out in all products using the MODIS land cover type product (MCD12C1).

RC4: 'Comment on essd-2024-553', Noemi Vergopolan

This paper introduces a dataset and approaches integrating AVHRR albedo and land surface temperature using an LSTM-based deep learning model to reconstruct soil moisture from 1982 to the present from GLASS-MODIS. The paper is well written, and the methodology is clear and well-detailed, but the paper does not highlight the novelty of the approach and how it differentiates from other efforts in the field. One major concern is that GLASS-MODIS already uses ERA5-Land, MERIT-DEM, and Soil-Grids as inputs; as such the proposed DL models are simply acting as surrogate models to GLASS-MODIS. The authors should demonstrate that AVHRR and the proposed DL model substantially add value to the predictions of soil moisture. As it reads, the added value of the approach is limited to extending the soil moisture time series to the AVHRR time series, which is helpful to evaluate long-term trends, but I wonder how it adds value beyond a CDF matching bias correction, for example.

Thank you very much for your insightful comments and the time you spent reviewing our manuscript, which have greatly helped improve its clarity. The primary motivation for this study arises from the scarcity of long-term global SM products characterized by both seamless coverage and reliable accuracy. We believe that developing such a dataset is crucial for supporting research on climate change, hydrological modeling, and land–atmosphere interactions over multi-decadal scales.

While the input features in this study are similar to those used in the GLASS-MODIS SM product, our work is not a simple extension. Considerable effort was devoted to constructing a large-scale training dataset, exploring the state-of-the-art DL models, and validating the final product to ensure its accuracy and robustness. Compared to the XGBoost model used in developing the GLASS-MODIS SM product, the AtLSTM model employed in this study demonstrates significant improvements in estimation accuracy, particularly at high SM levels ($> 0.4 \text{ m}^3 \text{ m}^{-3}$), by leveraging bidirectional temporal information and an attention module.

Typically, CDF matching bias correction adjusts statistical distributions locally, which limits its spatial generalization capability, especially in regions lacking in situ SM data. Furthermore, as a distribution-based correction method, it overlooks the temporal dependencies and nonlinear dynamics inherent in SM time series. In contrast, our approach integrates multi-source datasets using DL models and leverages their complementary strengths to effectively capture nonlinear temporal dependencies and spatial patterns, thus enhancing the generalizability of SM estimates, as demonstrated by the validation results in Section 4.2. Moreover, our product contains much finer spatial details, as evidenced by the comparison over the Tibetan Plateau (Fig. 10).

In summary, both the proposed DL-based SM estimation framework and the derived long-term global SM product present clear value, and we believe that this work aligns well with the journal's standards for innovation and high-quality datasets. To enhance the clarity of our contributions, we have added a comparison between our approach and the CDF matching bias correction method in the discussion section (Page 30).

Major:

GLASS-MODIS already uses ERA5-Land, MERIT-DEM, and Soil-Grids, which are also input data for your approach in this paper. How do you ensure your DL model is actually learning anything from GLASS-AVHRR and not just a surrogate for GLASS-MODIS learning directly the relationship from ERA5-Land, MERIT-DEM and SoilGrids. In other words, if the target data is not independent from the input data, your DL model is just a surrogate for the existing GLASS-MODIS model. Two ways to quantitatively test for that is to 1) check how statistically significant is the difference in the accuracy of your trained DL models with respect to GLASS-MODIS when validating the approach with in-situ observations, and 2) by testing how different is your DL accuracy if trained only on ERA5-Land, MERIT-DEM and Soil-Grids for the past period (1982-1999). These experiments would help elucidate how the proposed model is actually learning new things from the AVHRR data or just mimicking GLASS-MODIS from ERA5-Land, MERIT-DEM, and Soil-Grids.

Thank you for raising this insightful question. In this study, we aimed to integrate multi-source datasets and leverage their complementary strengths to develop a long-term, seamless global SM product. As MODIS satellite observations have only been available since 2000, we adopted the long-archive AVHRR albedo and LST products as the primary remote sensing inputs. Due to the limited number of ISMN in situ SM stations, it is not feasible to construct sufficient time-series training samples solely based on them for training a DL model. Therefore, we chose to use the GLASS-MODIS SM product as the training target for our models.

To ensure consistency in the feature space, our DL models were trained using input datasets similar to those employed to generate the GLASS-MODIS SM product, including ERA5-Land SM, MERIT DEM, SoilGrids, and the GLASS-AVHRR albedo and LST products. This setup enables the models to generalize effectively across different time periods, extending SM estimations to the pre-MODIS era (e.g., 1982–1999). Validation results indicate that although the inputs are derived from different satellite sensors, the GLASS-AVHRR SM product shows high consistency with the GLASS-MODIS SM after 2000, and achieves satisfactory accuracy against in situ observations prior to 2000.

We have further checked the differences in R between the GLASS-AVHRR and GLASS-MODIS SM products when evaluated against in situ SM datasets at different spatial scales (ISMN & COSMOS). Using both the paired t-test and the Wilcoxon signed-rank test, we found that the differences are statistically significant. It should be noted that these differences can be partially attributed to the inconsistency in spatial resolution between the two GLASS SM products (5 km vs. 1 km).

As also illustrated in the feature importance ranking plot (Fig. A3), both the GLASS-AVHRR albedo and LST play significant roles in the AtLSTM model's prediction. In fact, when these AVHRR-based variables were excluded from the inputs, the model performance on the test set drops notably (with R^2 decreasing from 0.987 to 0.968, RMSE increasing from $0.011 \text{ m}^3 \text{ m}^{-3}$ to $0.018 \text{ m}^3 \text{ m}^{-3}$). Since the GLASS-MODIS SM product is unavailable during the 1982–1999 period, this experiment was conducted using samples from 2000 to 2018.

In summary, these results suggest that the proposed model is not merely mimicking the GLASS-MODIS from ERA5-Land SM, MERIT DEM, and SoilGrids, but is indeed learning useful information from the AVHRR-based observations to enable robust, long-term SM estimation.

Moderate:

R^2 0.987 indicates that probably the DL is overfitting or there is contamination between the training and testing datasets used. In fact, when using the model to produce the GLASS-AVHRR dataset accuracy drops from R 0.63 to 0.79

We acknowledge your concern, but would like to clarify that the DL models do not exhibit overfitting, as their performance on the test set is very close to that on the training set. In addition, spatial independence was carefully ensured during the dataset splitting process to avoid any data leakage between the training and test sets.

The high accuracy ($R^2 = 0.987$) achieved here results from the model being trained using the GLASS-MODIS SM product, rather than in situ observations. Thus, this metric reflects the consistency between the GLASS-AVHRR and GLASS-MODIS products, rather than the actual accuracy of the GLASS-AVHRR SM product itself. We have removed model accuracy from the abstract to avoid misunderstanding.

The actual accuracy of the GLASS-AVHRR SM product was assessed using in situ SM datasets at different spatial scales. As discussed in Section 4.2, such validation shows that the product maintains competitive performance even when applied to AVHRR inputs prior to 2000, which supports its robustness and generalizability.

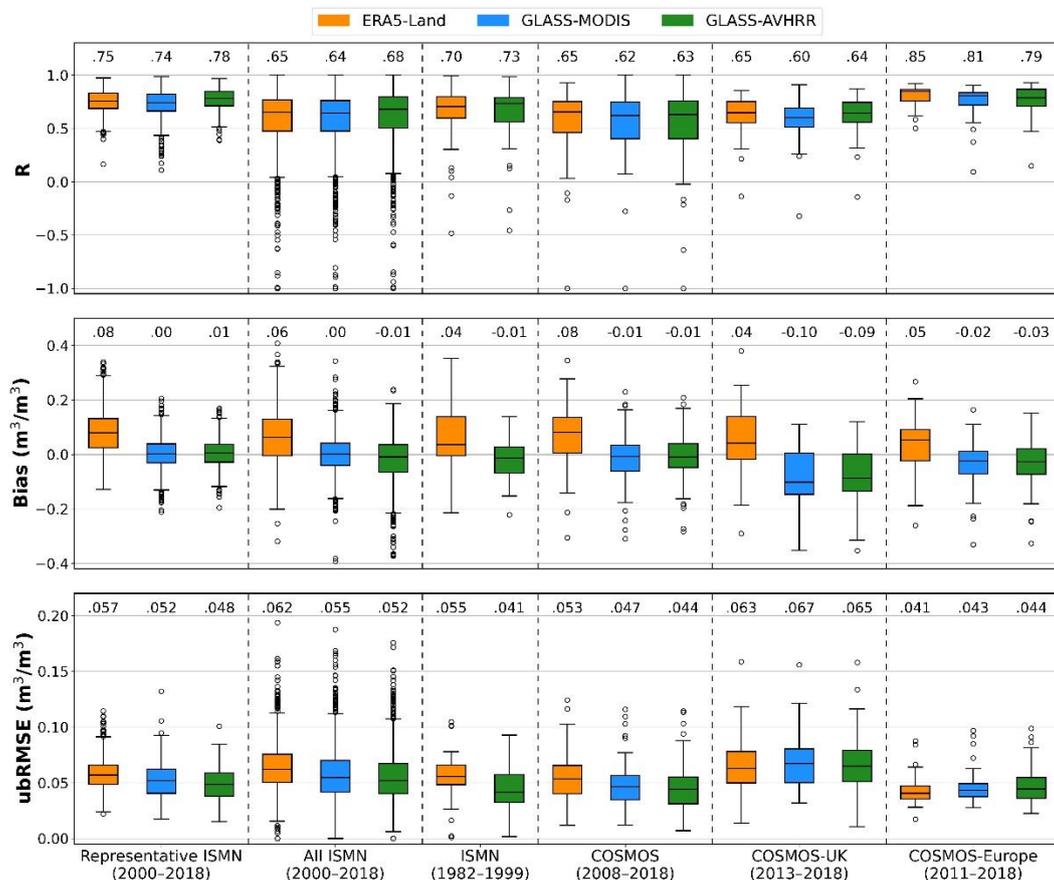
Minor:

I appreciate how the authors ensured that testing grids are within 25km distance from training grids. This prevents the testing sample is contaminated with inputs from the training datasets (since the resolution of precipitation input data from ERA5 into ERA5-Land runs is about 25km resolution).

Thank you for your positive feedback. As noted, ensuring a minimum distance of 25 km between any two samples helps prevent data leakage, especially given the spatial resolution of the ERA5 precipitation inputs.

Please add GLASS-MODIS to Figure 6 for the COSMOS comparisons. Please include marks indicating whether the difference between the accuracy of the SM data products is statistically significant.

We have included the validation results of the GLASS-MODIS SM product at the three COSMOS networks in the updated figure. The two GLASS SM products show comparable overall accuracy across these networks, although some site-specific discrepancies are observed, which are likely attributable to differences in the satellite remote sensing inputs and spatial resolution (Page 20). As the accuracy differences between the two GLASS SM products on all ISMN and COSMOS networks have been demonstrated to be statistically significant, we have not added additional significance markers in the figure.



Please add GLASS-MODIS to Table 6, which will be best displayed in landscape format. Also, please mark in bold the best statistically significant data product/metric for each site.

We used the upscaled 9 km SMAP CVS SM dataset for site-specific validation because its spatial resolution is closer to that of the 5 km GLASS-AVHRR SM product. Using these sites to validate the 1 km GLASS-MODIS SM product could introduce more uncertainty due to the larger scale differences. Furthermore, the validation results based on the ISMN and COSMOS networks indicate that the two GLASS SM products exhibit similar overall accuracy. Therefore, we consider that including GLASS-MODIS in Table 6 may not provide additional information. Even so, following your suggestion, we have marked in bold the best-performing metrics for each site.

Please add GLASS-MODIS to Figure 7.

Thank you for the suggestion. Considering the reasons mentioned above and to avoid excessive overlap of scatter points that could reduce the figure's clarity, we prefer not to include the GLASS-MODIS SM product in this figure.