The authors have motivated the problem well: there are significant data gaps in the time series of FY-3B SM daily products, particularly in the high northern latitudes. They are proposing to only use the good SM data, along with a spatiotemporal model for gap-filling, to create a complete daily SM time series for July 2011-August 2019. No ancillary data (precip, NDVI, etc) is used.

Re: Thank you very much for your positive comments on our paper. We have carefully considered your comments and revised the paper accordingly. Please see the response below.

Major comments:

- Line 202: How were the data composited? Mean, min, max?

Thank you for this comment. We have clarified the compositing method in the revised manuscript. Specifically, when both ascending and descending orbits provide valid observations, their mean value is used; when only one orbit provides a valid observation, that observation is retained; and when neither orbit provides a valid observation, the pixel is assigned an invalid value. The revised description has been added at Page 6, Lines 211–213.

Page 6, Lines 211–213

"The data from ascending and descending orbits were composited for each day. When both orbits provided valid observations, their mean value was used; when only one orbit provided a valid observation, that observation was retained; and when neither orbit provided a valid observation, the pixel was assigned an invalid value."

- Line 231: How are pixels determined to be "invalid"? Is there a flag in the original data?

Thank you for this valuable comment. We have clarified how invalid pixels are identified in the revised manuscript. Specifically, among non-background pixels, a pixel is considered invalid when neither the ascending nor descending orbit provides SM information. We have added this information to the revised manuscript at Page 6, Lines 241–243 to clarify the criteria used to identify invalid pixels. Additionally, we have added a source flag dataset, where 0 denotes background values, 1 denotes original observations, and 2 denotes reconstructed values. Only the invalid pixels in the original data will be reconstructed. The final dataset has been updated accordingly, and the corresponding description has also been added to the Readme.txt.

Page 6, Lines 241–243

"The mask is used to distinguish valid and invalid pixels based on the availability of original FY-3B SM observations, with 1 denoting a pixel with FY-3B SM observations and 0 denoting a pixel without observations."

Readme.txt

"## Source flag
A source flag is provided to distinguish background, original, and reconstructed pixels.
0: Background pixels
1: Original FY-3B SM observations
2: Reconstructed FY-3B SM values
Users are advised to use the source flag to distinguish original observations from reconstructed

estimates when using the dataset."

- Lines 226-244: The description of the training approach is confusing and incomplete. How many patches are being used? Are there any patches for the regions with such low coverage as in the high latitudes of Fig 1? How are masks created for the patch for other time steps (not T)?

  Thank you for this valuable comment. We have substantially rewritten the description of the construction of training dataset to make the sample construction procedure clearer and more complete (Page 6, Lines 236-256). The training dataset is constructed through two steps: Step 1, patch selection, in which only spatially complete 40 × 40 patches at the target time T are selected; and Step 2, simulated mask generation, in which the spatially complete patches at T are artificially masked using missing patterns derived from real FY-3B SM masks. Notably, for the 9 × 40 × 40 spatiotemporal patch group, it is important to clarify that the simulated masks extracted from real FY-3B SM missing patterns are applied only to the target time T. No additional artificial masking is applied to the other eight time steps (*from* $T-4$ *to* $T+4$, excluding T); their masks are directly derived from the original FY-3B SM observations and retain their original spatial availability. Based on these criteria, a total of 9937 patches are selected for model training. Because the patch selection step requires spatial completeness at T, the training patches are mainly distributed in regions with relatively higher FY-3B SM spatial coverage. Regions at higher latitudes with relatively low spatial coverage may therefore be less represented in the training dataset. Nevertheless, the model's spatial generalization in such regions is further assessed using independent in-situ observations, including stations located at higher latitudes.

  Page 6, Lines 236-256

"The training and prediction workflow based on the proposed GSP model is shown in Fig. 2. Training samples are created through patch selection and simulated mask generation. During patch selection, the geographical locations with spatially complete patches of 40 × 40 at T are identified. For each selected location, the FY-3B SM spatiotemporal information and corresponding masks over nine consecutive days (from $T-4$ to $T+4$) are extracted from the same spatial extent, forming a 40 × 40 × 9 spatiotemporal patch group. The mask is used to distinguish valid and invalid pixels based on the availability of original FY-3B SM observations, with 1 denoting a pixel with FY-3B SM observations and 0 denoting a pixel without observations. During simulated mask generation, the simulated mask patched (extracted from real orbital gaps in FY-3B SM) are only applied to the spatially complete patch at T within the 40 × 40 × 9 spatiotemporal patch group to reproduce the missing patterns of FY-3B SM as closely as possible. Specifically, 40 × 40 simulated mask patches are extracted from the original FY-3B SM masks over the training period from 16 July 2011 to 31 December 2016, and those with missing rates between 30% and 70% are retained. The spatiotemporal patch groups (from $T-4$ to $T+4$) with simulated gaps at T are used as input features. The originally known data under the mask at T are used as the corresponding training labels. The samples collected from all dates from 16 July 2011 to 31 December 2016 were employed as the training dataset. Because spatial completeness at T is required for constructing reliable training labels, the spatial distribution of the training patches is determined by the availability of complete FY-3B SM observations. A total of 9937 patches were selected for model training, which

were mainly distributed in areas with relatively higher spatial coverage to ensure complete patches to be obtained.

"

- Lines 248-252: Does "masked" mean it gets included? Or not included? Equations 1-2 suggests included, but the term could mean either depending on interpretation.

  Re: Thank you for this valuable comment. We corrected Equations (1)–(2) on Page 7, Lines 261-266, which were incorrectly formulated in the original manuscript and could imply that the masked observations were included in the input. In the revised manuscript, "masked" explicitly refers to excluding the original valid observations from the input patch, that is, masking the valid observations to artificially generate spatially incomplete patches at time t. The masked patches are then fed into the GSP to predict the FY-3B SM at the masked locations, and the predicted values are compared with the corresponding original valid observations to calculate the loss.

  Page 7, Lines 261-266

"

$$M_n = \begin{cases} 0, & if\ X_n\ is\ masked \\ 1, & otherwise \end{cases} \quad (1)$$

$$Loss(X, \hat{X}, M) = \frac{\sum_{n=1}^{N}\left(\left(\hat{X}_n - X_n\right) \odot (1 - M_n)\right)^2}{\sum_{n=1}^{N}(1 - M_n)} \quad (2)$$

where M represents the simulated mask matrix, with 0 representing masked and 1 not. M_n indicates the value of M at location n. X represents the original spatially complete FY-3B SM patch, and $\hat{X}$ represents the reconstructed spatially complete FY-3B SM patch. X_n and $\hat{X}_n$ represent the original and reconstructed values of FY-3B SM at location n, respectively.

"

- Line 259: How are patches stitched together? How many patches are there?

Thank you for this comment. We have revised this part to clarify both the number of patches used for daily prediction and the procedure for assembling the predicted patches into a complete FY-3B SM map. Specifically, we have clarified that the daily prediction is based on a fixed set of 525 non-overlapping 40 × 40 patches covering the entire study region. The revised description also explains how the individual prediction results are arranged and combined to obtain the final daily seamless map. The corresponding text has been added at Page 7, Lines 269–273.

Page 7, Lines 269–273

"A fixed set of 525 non-overlapping patches covering the globe is generated for each day and sequentially input into the trained GSP model to produce the corresponding seamless patches. The predicted patches are then stitched together according to their predefined spatial order to reconstruct the spatially complete FY-3B SM map for each date from 12 July 2011 to 19 August 2019."

- Line 277: Only the initial input is a four-month temporal average? Then more granular data is used?

Thank you for this valuable comment. We have clarified the roles of the different inputs in Section 3.2 (Page 8, Lines 282–290) to avoid ambiguity regarding the model inputs. The 9 × 40 × 40 spatiotemporal patch group and its corresponding masks constitute the initial inputs to the GSP model. The previous description of the four-month temporal average as an "initial input" has been corrected. As further clarified in Section 3.2.3 (Page 11, Lines 365–373), it is used to construct a spatially complete background reference, which is incorporated into the feature fusion process to alleviate the influence of large-scale spatial gaps in the original FY-3B SM data.

Page 8, Lines 282–290

"As shown in Fig. 3, the model consists of three main components, including the multi-scale gated convolution block, the residual Swin Transformer block, and the feature fusion block. The multi-scale gated convolution block and residual Swin Transformer block are organized in a parallel architecture to extract local and global spatiotemporal features, respectively, from the input FY-3B SM spatiotemporal patch group and corresponding masks. The extracted features are then concatenated and fed into the feature fusion block to fully exploit both types of features. To provide a spatially complete SM background reference, the four-month temporal average of FY-3B SM is also introduced into the feature fusion block, where it is fused with the extracted spatiotemporal features to alleviate the influence of large-scale spatial gaps in original FY-3B SM."

Page 11, Lines 365–373

"The attention-refined features are then fed into six convolutional layers for further feature extraction and fusion. Local residual connections are introduced by directly adding shallow features to deeper features, facilitating feature propagation across convolutional layers and mitigating gradient vanishing. In addition, the missing values in the original FY-3B SM spatiotemporal input are filled using the corresponding four-month temporal averages of FY-3B SM at each location to construct a spatially complete background, which is then introduced as a global residual to provide spatially continuous information and reduce the influence of large-scale spatial gaps in the original FY-3B SM data."

- Section 3.3: Figure 1 demonstrates that there are some places without data almost entirely for the whole period. Are there in situ observations available in those regions to validate the algorithm for extremely data-poor pixels? Suggest creating a figure that easily visualizes the temporal coverage proportion (as in Fig 1) for each of the in situ stations to try to answer the question of if the stations used for validation represent different temporal coverage proportions.

Thank you very much for this valuable suggestion. We added Fig. 6 to explicitly visualize the temporal coverage proportion of each in-situ station used for validation. The figure shows that the temporal availability of the ground observations varies among stations, with different stations exhibiting different temporal coverage proportions over the study period. This provides a more direct characterization of the temporal representativeness of the in-situ observations used for validating the reconstructed FY-3B SM. We further examined the availability and temporal coverage of the in-situ observations in regions where FY-3B SM has relatively low data coverage and where training samples are difficult to obtain. On the Tibetan Plateau, three in-situ networks, including MAQU, NAQU, and CTP-SMTMN, are available, comprising approximately 90 stations in total. The temporal coverage of these stations is approximately 30%. Based on these stations, the reconstructed FY-3B SM achieves an *R* of approximately 0.65 with the in-situ observations. In the high-latitude regions, the available stations are mainly distributed in North America, including stations from the SNOTEL, SCAN, and USCAN networks. Nearly 50 stations are located in these regions, with an average temporal coverage of approximately 50% and a maximum of about 65%. These results indicate that independent in-situ observations are still available for accuracy assessment in regions with relatively low FY-3B SM coverage and limited representation in the training samples. The relatively great reconstruction performance obtained at these stations suggests that GSP retains a certain degree of spatial generalization capability in regions that are difficult to represent during model training. The corresponding explanation has also been added to the revised manuscript on Page 2, Lines 391–396 and Page 12, Lines 406–408.

Fig.6

“

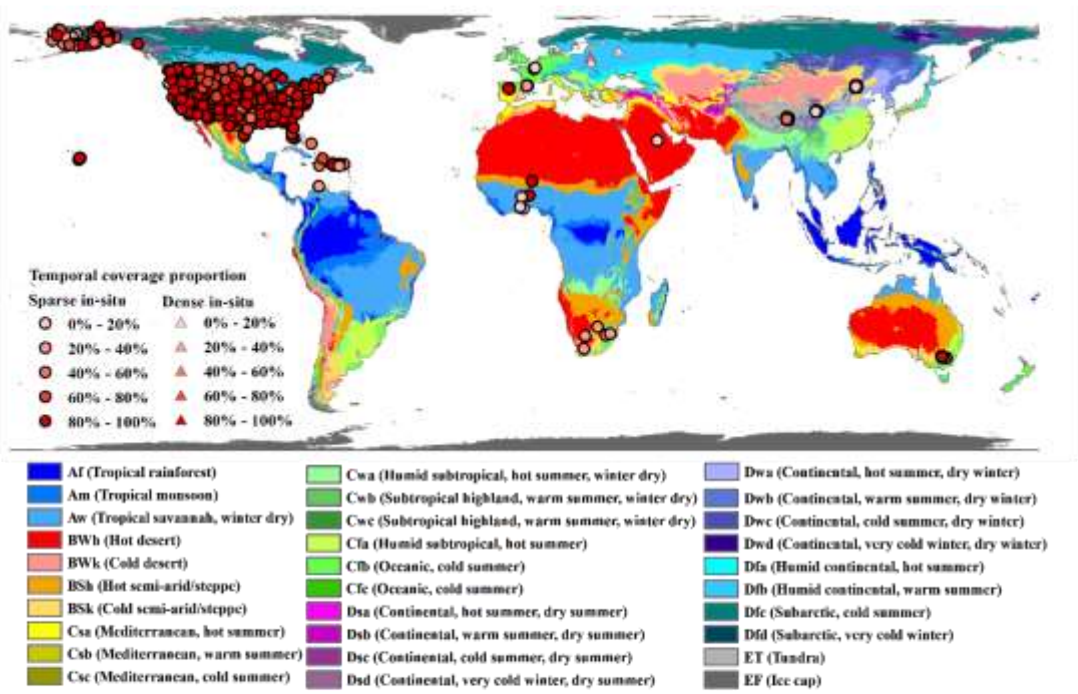


Fig. 6. *Spatial distribution and* *<u>temporal coverage proportion</u>* *of the in-situ data used for validation,* *<u>with the points and triangles</u>* *representing the in-situ stations in the sparse and dense networks, respectively (the Köppen-Geiger climate classification data were used as the base map).*

”

Page 2, Lines 391–396

<u>“This evaluation provides an independent assessment of the reconstructed FY-3B SM under real missing-data conditions by comparing the reconstructed FY-3B SM with ground observations at the corresponding locations. In particular, the in-situ observations enable the reconstruction performance of GSP to be evaluated in regions with relatively low FY-3B SM spatial coverage,</u>

where spatially complete patches are difficult to obtain and therefore are difficult to include in the simulated missing experiments."

Page 12, Lines 406–408

"Fig. 6 and Table 1 present the spatial distribution, temporal coverage proportion, and details of the two types of in-situ networks used in this study, including 17 sparse networks and five dense networks."

- Fig 5: Commentary on why the station data has consistently lower values?

Thank you for the valuable comment. Regarding the relatively lower in-situ observations during the trough periods in Fig. 5, we believe that this discrepancy is primarily associated with the different spatial representativeness of ground-based and satellite observations. First, in-situ measurements characterize SM at specific locations, while FY-3B SM provides a spatial estimate over a much larger pixel. Consequently, when substantial spatial heterogeneity exists within a pixel, the satellite-derived value may not necessarily correspond to the conditions at an individual station. For example, a pixel may include relatively wetter areas, such as low-lying terrain, wetlands, or densely vegetated regions, which could contribute to a higher pixel-averaged SM than that measured at a relatively dry station. Second, a similar discrepancy can also be observed between the original FY-3B SM and in-situ observations in Fig. 8, where the satellite estimates remain relatively higher during the trough periods. This indicates that the difference observed in Fig. 5 was not introduced by the GSP model. Instead, it mainly reflects the inherent differences in spatial representativeness between satellite-based and point-scale observations. We have added a corresponding explanation on Page 14, Lines 441-447.

Page 14, Lines 441-447

"The in-situ observations sometimes show lower values than FY-3B SM. This pattern is particularly evident in the AMMA-CATCH, SOILSCAPE, and XMS-CAT networks, likely due to differences in spatial representativeness. In-situ measurements represent point-scale conditions, whereas FY-3B SM represents spatially averaged conditions over a larger pixel. The pixel may contain relatively wetter areas, such as low-lying terrain, wetlands, or densely vegetated regions, which may increase the pixel-averaged SM relative to measurements at a relatively dry station. This effect can be more evident in areas with strong within-pixel spatial heterogeneity."

- Line 392: GSP seems to perform similarly to DCT-PLS. Have a mean R of 0.5387 is really not that great…

Thank you for this valuable comment. We agree that the mean R of 0.5387 indicates that there is still uncertainty in the reconstructed FY-3B SM when evaluated against in-situ observations. In the spatiotemporal reconstruction models, the original FY-3B SM provides the basis for learning the spatiotemporal characteristics and reconstructing missing values, while the model is trained to make the reconstructed FY-3B SM as close as possible to the original FY-3B SM labels. Therefore, the accuracy of the reconstructed results is constrained by the accuracy of the original FY-3B SM product. In addition, the reconstruction process inevitably introduces a certain degree of uncertainty, which may lead to some degradation in accuracy compared with the original FY-3B SM (Page 17, lines 477-480). As demonstrated by the validation under dense in-situ networks, the original FY-3B SM itself achieves a mean R of 0.5858 against the in-situ observations, indicating that the moderate correlation is not solely attributable to the reconstruction process. Although the reconstructed FY-3B SM may show a slight decrease in accuracy compared with the original FY-3B SM, the GSP-based reconstruction remains the closest to the original FY-3B SM among the compared reconstruction methods. In addition, the relatively moderate correlation is also affected by the substantial difference in spatial representativeness between satellite-derived FY-3B SM and point-scale in-situ measurements. Regarding the similar mean R between GSP and DCT-PLS, we acknowledge that their performance is relatively close when evaluated against in-situ observations. However, the simulated missing-data experiments show that GSP achieves substantially better reconstruction performance than DCT-PLS, particularly under large-scale and continuous spatial missing conditions. This result highlights the specific advantage of GSP in handling extensive spatial gaps. Therefore, although GSP and DCT-PLS show comparable performance in the in-situ validation, the simulated missing-data experiments demonstrate the distinct capability of GSP for reconstructing FY-3B SM under large-scale spatially continuous missing conditions.

Page 17, lines 477-480

"The reconstruction methods mainly rely on the valid spatiotemporal information of the original FY-3B SM to reproduce its characteristics. Since the reconstruction process inevitably introduces a certain degree of uncertainty, these methods may exhibit varying degrees of accuracy degradation compared with the original FY-3B SM."

- Fig 6: There is very little network data for 3/5 dense networks. Why are the authors including the FY-3B data here but not in the Fig 5 for the sparse networks?

Thank you for the valuable comment. First, the limited temporal coverage of the SWEX_POLAND, VAS, and Ru_CFR networks is mainly due to the small number of stations within these networks, while the individual stations also contain substantial periods of missing observations. Such data gaps may arise from instrument malfunctions, maintenance, data transmission problems, or quality-control procedures. The limited number of stations, combined with the extensive temporal gaps in the station observations, consequently results in relatively limited temporal coverage of the network-averaged in-situ time series. Nevertheless, during the available periods, the reconstructed FY-3B SM shows temporal variations consistent with those of the corresponding in-situ network observations. In addition, the reconstructed FY-3B SM maintains good temporal continuity with the original FY-3B SM, indicating that the reconstruction preserves the temporal characteristics of the original observations while providing continuous estimates during missing periods. Second, the original FY-3B SM was included in Fig. 6 because the stations within each dense network are distributed within a single FY-3B SM pixel. Therefore, the reconstructed FY-3B SM can be directly compared with the corresponding original FY-3B SM time series, providing a meaningful reference for evaluating the temporal consistency of the reconstruction. In contrast, stations in sparse networks may be distributed across multiple FY-3B SM pixels. The original FY-3B SM data may exhibit different missing patterns among these pixels. Consequently, when calculating the network-averaged original FY-3B SM, not all pixels contribute to the network mean at each time point. The resulting time series may therefore be less representative of the overall SM conditions of the network. For this reason, the original FY-3B SM was included in Fig. 6 for the dense networks but not in Fig. 5 for the sparse networks. The corresponding explanation has been added to the Page 17, Lines 469-472.

Page 17, Lines 469-472

"The original FY-3B SM was also included as a reference because the stations within each dense network are distributed within a single FY-3B SM pixel, allowing a direct temporal comparison between the reconstructed FY-3B SM time series and the corresponding original FY-3B SM time series."

- Table 3: So the original FY-3B data performs better than the GSP model when comparing to station data?

Thank you for the valuable comment. GSP is designed to reconstruct missing FY-3B SM using the available spatiotemporal neighborhood information from the original FY-3B observations, rather than to improve the retrieval accuracy of the original product. Meanwhile, the reconstruction process inevitably introduces a certain degree of uncertainty, which may result in a slight reduction in accuracy compared with the original FY-3B SM. As shown in Table 3, although the reconstructed FY-3B SM based on GSP shows a slight decrease in overall accuracy, the differences in RMSE, MAE, and ubRMSE remain small. This suggests that the uncertainty introduced by the reconstruction process is limited and that GSP can effectively preserve the spatiotemporal information contained in the original FY-3B SM. Therefore, the primary contribution of GSP is to extend the spatial coverage and continuity of FY-3B SM while maintaining comparable accuracy. We have provided detailed explanations on Page 17, lines 447-480.

Page 17, lines 447-480

"The reconstruction methods mainly rely on the valid spatiotemporal information of the original FY-3B SM to reproduce its characteristics. Since the reconstruction process inevitably introduces a certain degree of uncertainty, these methods may exhibit varying degrees of accuracy degradation compared with the original FY-3B SM."

- Section 5.3: I hesitate with the choice to consider the GLDAS-Noah as "truth" reference. If considering this the truth dataset, and it is spatially complete, what is even the advantage of the GSP dataset?

  Re: Thank you very much for this valuable comment. We agree that GLDAS-Noah should not be treated as a "truth" reference for validating the reconstructed FY-3B SM, because it is a land surface model product and contains its own uncertainties. We have therefore revised Section 5.3 to reposition GLDAS-Noah as a comparison dataset rather than a validation reference. Specifically, the revised analysis focuses on comparing the spatial characteristics of the TTP-reconstructed FY-3B SM, GSP-reconstructed FY-3B SM, and GLDAS-Noah SM. The previous validation based on GLDAS-Noah has been removed. Instead, spatial variation profiles and regional spatial patterns are used to examine how well the reconstructed datasets based on GSP preserve spatial heterogeneity. The revised results show that TTP tends to produce more spatially homogeneous distributions, whereas GSP retains more pronounced spatial variations and finer spatial structures. GLDAS-Noah generally captures the large-scale spatial pattern but appears smoother in some regions. We have also clarified that GSP is intended to reconstruct the missing observations of the FY-3B satellite SM product while preserving the spatial characteristics of the original satellite observations. Compared with TTP, GSP makes better use of surrounding spatiotemporal information and therefore shows a stronger ability to retain the spatial structure of FY-3B SM. Accordingly, the discussion of Section 5.3 has been revised on Page 24, Lines 593–619.

Page 24, Lines 593–619

"*5.3 Spatial comparison among the GSP- and TTP-reconstructed FY-3B SM datasets and GLDAS-Noah SM*

To further examine the spatial characteristics of the reconstructed FY-3B SM dataset, the 25 km GLDAS-Noah SM dataset, which provides spatially complete SM estimates, was introduced as a comparison dataset. Taking the GLDAS-Noah SM dataset and TTP- and GSP-reconstructed FY-3B SM on 16 July 2011 as an example, four lines and six patches were randomly selected for spatial comparison, as shown in Fig. 14. As shown in Fig. 14(b), the TTP-reconstructed FY-3B SM exhibits very limited spatial variation, with locally nearly invariant SM. In contrast, the GSP-reconstructed FY-3B SM shows more pronounced spatial variations. The variation pattern of the difference between GSP and TTP is also generally consistent with that of GSP, indicating that much of the spatial variation between the two reconstructed datasets originates from the spatial variability retained by GSP, whereas the TTP reconstruction tends to produce a more spatially homogeneous distribution. Moreover, the spatial variation patterns of GSP and GLDAS-Noah show broadly consistent fluctuations along several transects. A similar pattern can be observed from the spatial maps in Fig. 14(c). The TTP-reconstructed FY-3B SM exhibits relatively weak spatial texture, whereas the GSP-reconstructed results present more continuous and complete spatial structures while retaining finer spatial details. GLDAS-Noah also reproduces the major large-scale spatial variations, but appears relatively smooth in some regions. This is likely related to the nature of GLDAS-Noah as a land surface model product, for which the spatial distribution is influenced by meteorological forcing, model parameterization, and grid-scale averaging, thereby smoothing part of the local-scale spatial heterogeneity. These comparisons further indicate that, compared with TTP, which primarily considers temporal relationships, GSP can better preserve and reconstruct the spatial structure of the original FY-3B SM by incorporating information from the surrounding spatiotemporal neighborhood. Meanwhile, compared with the model-based GLDAS-Noah dataset, the GSP-reconstructed FY-3B SM based on satellite SM retains more localized spatial details in some regions, reflecting the spatial characteristics inherited from the original satellite observations.

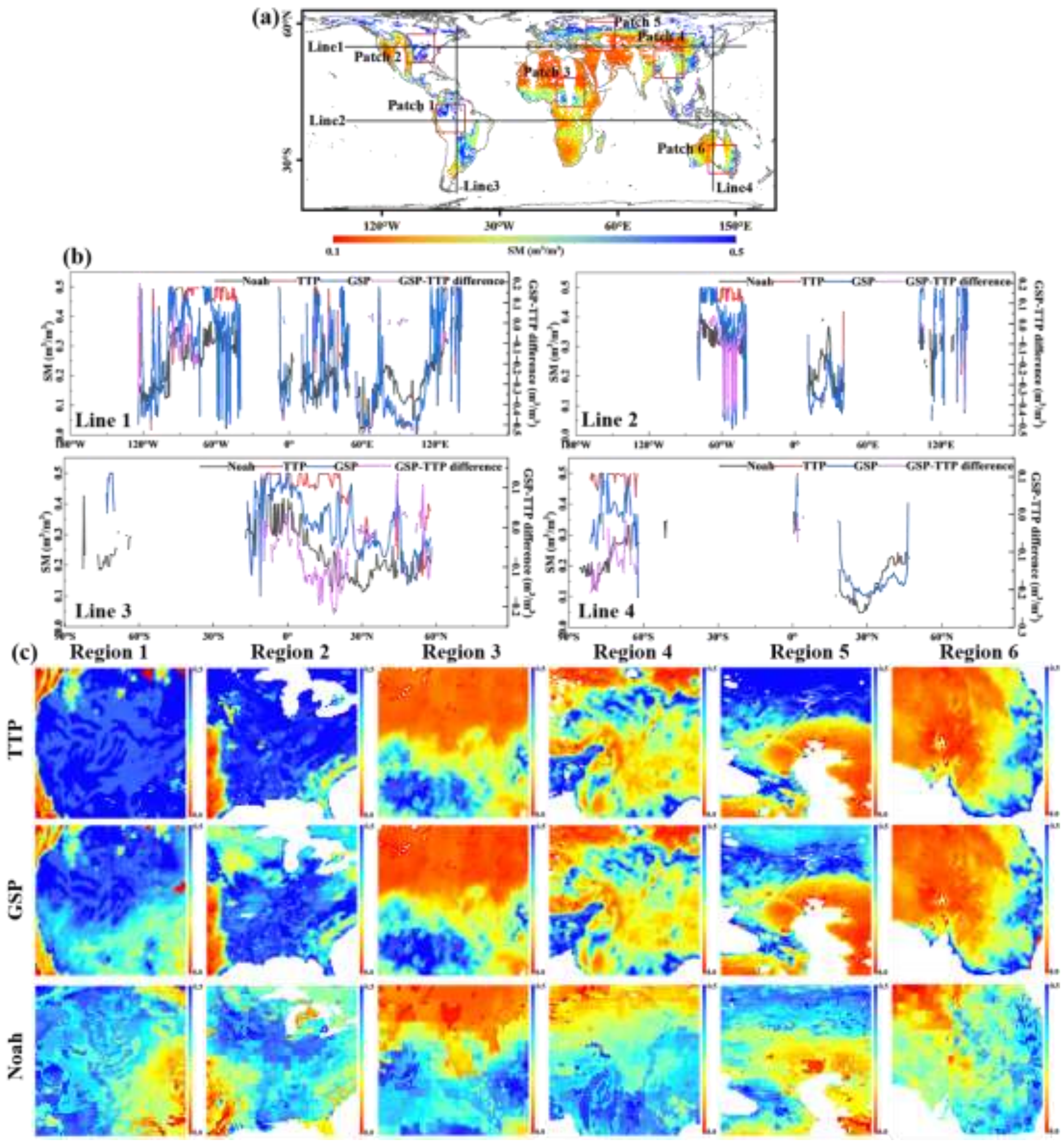


Fig. 14. Comparison of spatial heterogeneity among the TTP-reconstructed FY-3B SM, GSP-reconstructed FY-3B SM, and GLDAS-Noah SM. (a) Locations of six randomly selected patches and four lines at the global scale, where the four lines include two horizontal lines at 41.2485° N and 3.0274° S, and two longitudinal lines at 62.9935° W and 132.2343° E. The center coordinates of the six patches are Region 1 (4.0054° S, 67.4187° W), Region 2 (37.7069° N, 88.2429° W), Region 3 (11.6971° N, 26.2907° E), Region 4 (28.3807° N, 99.1757° E), Region 5 (48.4103° N, 47.1150° E), and Region 6 (28.6028° S, 140.8243° E). (b) Spatial variation profiles of SM along the four lines, comparing TTP-reconstructed FY-3B SM, GSP-reconstructed FY-3B SM, and GLDAS-Noah SM. (c) Visual comparison of the spatial SM patterns of the TTP-reconstructed FY-3B SM, GSP-reconstructed FY-3B SM, and GLDAS-Noah SM for the six selected patches.

"

Minor comments:

- Line 256-7: Incomplete sentence

Thank you for the comment. We have revised the sentence to ensure grammatical completeness on Page 7, Lines 267-269.

Page 7, Lines 267-269

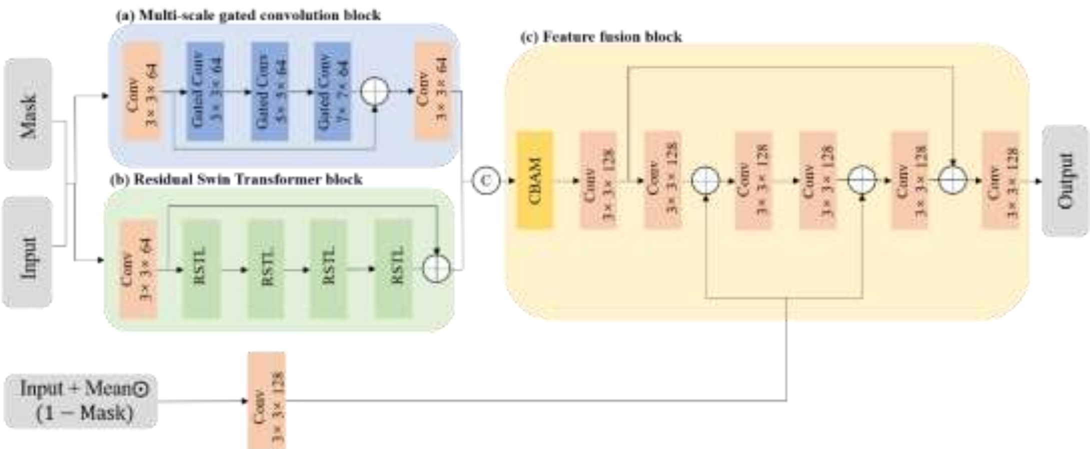
"During prediction, for each day, the spatiotemporal patches from nine temporally closest days (with the same size of 40×40×9 as the training data) <u>are extracted</u> for locations with real orbital gaps."

- Line 268: Does "feature fusion block" include CBAM and beyond?

Thank you for this comment. The feature fusion block includes the CBAM and multiple convolutional layers. To explicitly illustrate the components included in the feature fusion block, we have redrawn the overall GSP architecture (Fig. 3). Specifically, we have outlined the feature fusion block with a yellow rounded rectangle. In addition, we have reorganized Section 3.2.3 on Page 11, Lines 355-376 as "Feature Fusion Block" and provided a clearer description of its specific components and their roles in the feature fusion process.

Fig. 3

"



"

Page 11, Lines 355-376

<u>"3.2.3 Feature fusion block</u>

<u>The feature fusion block includes the CBAM and multiple convolutional layers (see (c) in Fig. 3). Residual connections are incorporated to transfer features from previous layers to subsequent layers, thereby mitigating the problem of gradient vanishing. Furthermore, to enhance the spatial coverage of the initial reference data and incorporate more reliable FY-3B SM information from temporally neighboring data, the four-month temporal average of FY-3B SM at each location is introduced into the feature fusion block as the initial input to mitigate the effect of large-scale data gaps on FY-3B SM reconstruction. As shown in Fig. 5, the CBAM is a lightweight attention</u>

mechanism that automatically focuses on important channel features and key spatial locations by sequentially combining channel and spatial attention mechanisms (Woo et al., 2018). The channel attention mechanism emphasizes informative channels by applying parallel max pooling and average pooling to compress the feature map along the spatial dimension, after which a shared MLP is used to compute the channel attention map. The spatial attention mechanism similarly compresses the feature map along the channel dimension using max pooling and average pooling, generating the spatial attention map through convolution (Zhang et al., 2024b). The input is weighted by the channel and spatial attention maps in sequence to produce more reliable feature representations.

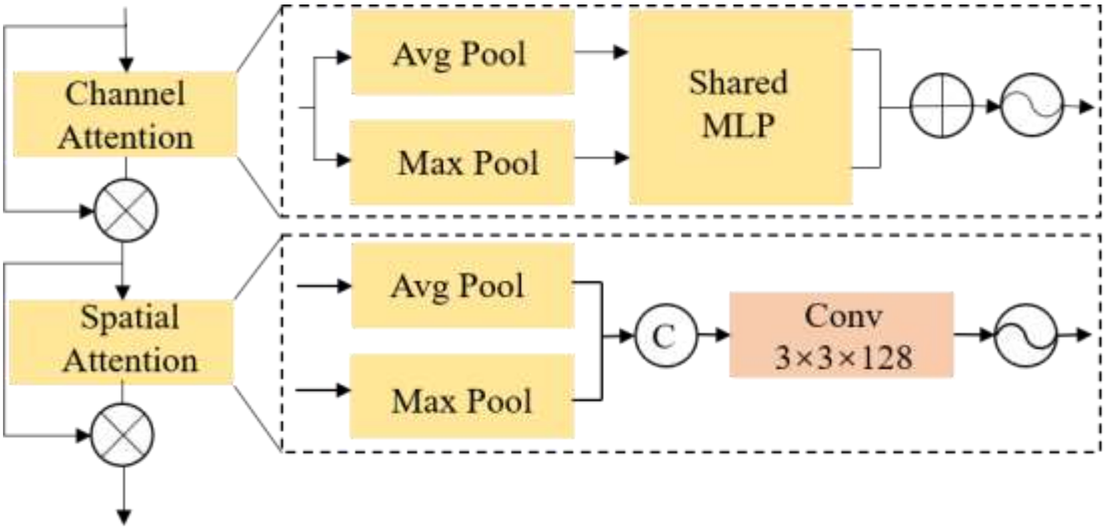


Fig. 5. *The architecture of the CBAM ("Avg Pool" represents the average pooling, "Max Pool" represents the max pooling, and "Shared MLP" represents the shared multilayer perceptron).*

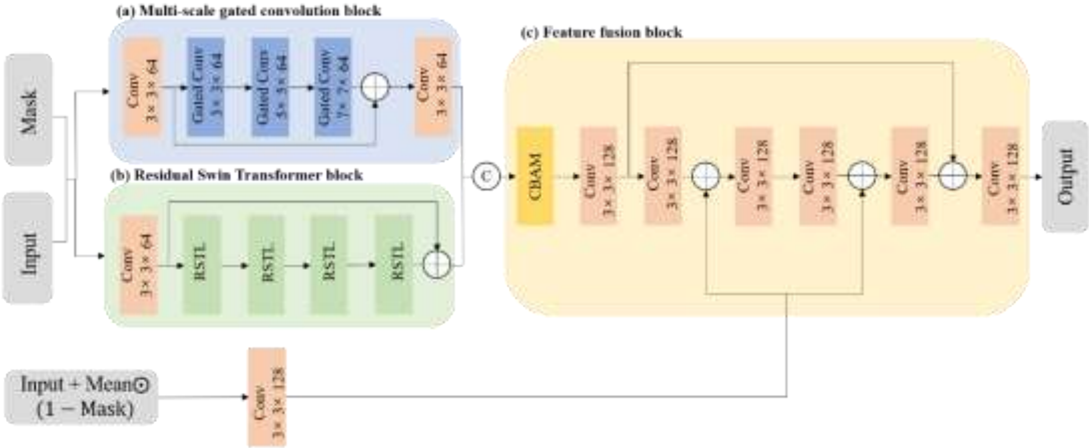
"

- Fig 3: Add definition of yellow box (blue and green are defined). Consider breaking out the RSTL, STB, and CBAM definitions into separate figures (or at least delineate them from the main architecture diagram somehow).

Thank you for this helpful suggestion. We have revised the figures to improve the clarity of the GSP architecture. In Fig. 3, the feature fusion block is highlighted with a yellow rounded rectangle, and its definition is now explicitly indicated. In addition, we have separated the detailed structures of RSTL and STB into Fig. 4 and the structure of CBAM into Fig. 5, respectively. These revisions allow the main architecture and the detailed structures of the individual modules to be presented more clearly.

Fig. 3

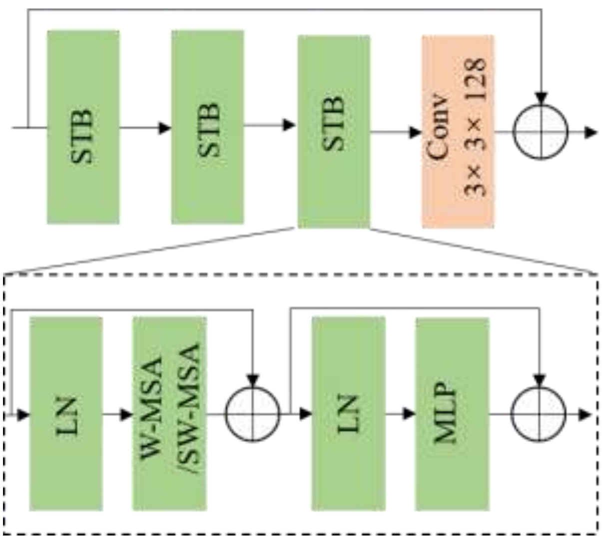
"



*[**Fig. 3.** The architecture of the GSP model. (a) is the multi-scale gated convolution block, (b) represents the residual Swin Transformer block, and (c) represents the feature fusion block (“Gated Conv” represents the gated convolution operation, “RSTL” represents the residual Swin Transformer layer, “Conv” represents the convolution operation).”]*

Fig. 4

“



*[**Fig. 4.** The architecture of RSTL (“STB” represents the Swin Transformer block. “LN” is the abbreviation for “Layer Normalization”, “W-MSA” and “SW-MSA” represent window-based multi-head self-attention and shifted window multi-head self-attention, respectively, and “MLP” represents the multilayer perceptron).”]*

Fig. 5

“

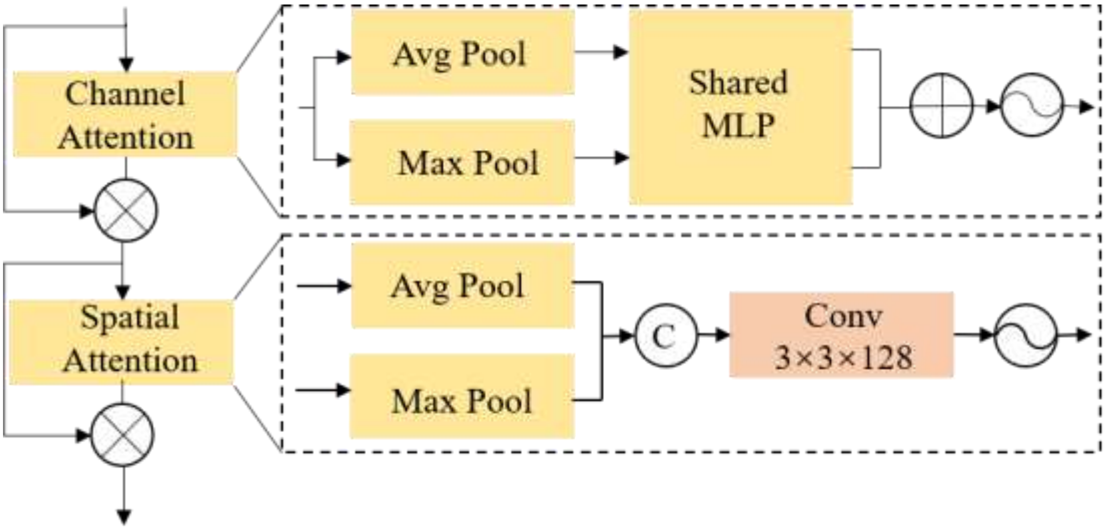


Fig. 5. *The architecture of the CBAM ("Avg Pool" represents the average pooling, "Max Pool" represents the max pooling, and "Shared MLP" represents the shared multilayer perceptron).*
"

- Table 1: How can SD_DEM be a dense network with only 1 station?

  Thank you for the comment. We have clarified the definition of a dense network on Page 12, Lines 402-406. Under this definition, a dense network can contain either one or multiple stations, as long as all stations within the network are located in a single FY-3B SM pixel. For SD_DEM, the network contains only one station in our network classification, and this station is located within a single FY-3B SM pixel. Therefore, SD_DEM is classified as a dense network according to the above definition.

  Page 12, Lines 402-406
  "Meanwhile, to minimize the effects of differences in spatial locations of stations, all networks were divided into two types according to the distribution of stations, including sparse (i.e., stations are distributed across multiple FY-3B SM pixels) and dense (i.e., all stations in a network, which may contain one or more stations, are distributed within a single FY-3B SM pixel) in-situ data."

- Tables 2,3: Consider adding horizontal lines to separate each network.

Thank you for this suggestion. We have added horizontal lines to Table 2 and Table 3 to clearly separate the different networks, as recommended.

Table 2

"**Table 2**
Accuracy evaluation of the reconstructed FY-3B SM dataset at the 17 sparse networks (bolded and underlines values indicate the greatest and second greatest accuracy).

		R	RMSE (m^3/m^3)	MAE (m^3/m^3)	ubRMSE (m^3/m^3)			*R*	RMSE (m^3/m^3)	MAE (m^3/m^3)	ubRMSE (m^3/m^3)
AMMA-CATCH	GSP	0.8000	0.0594	0.0529	**0.0325**	ARM	GSP	0.4969	0.0891	0.0727	0.0746
	U-TILISE	0.8017	**0.0585**	**0.0520**	**0.0325**		U-TILISE	0.4378	0.0958	0.0775	0.0839
	STpconv	0.8019	0.0588	0.0524	**0.0325**		STpconv	0.4779	0.0950	0.0772	0.0813
	DCT-PLS	**0.8063**	0.0591	0.0526	**0.0325**		DCT-PLS	**0.5150**	**0.0875**	**0.0714**	**0.0733**
CTP_SMTMN	GSP	0.7433	0.0969	0.0767	0.0791	MAQU	GSP	**0.5916**	**0.1246**	**0.0999**	**0.0968**
	U-TILISE	0.6711	0.1149	0.0914	0.0972		U-TILISE	0.4434	0.1596	0.1294	0.1239
	STpconv	0.7379	0.1009	0.0794	0.0841		STpconv	0.5068	0.1344	0.1064	0.1101
	DCT-PLS	**0.7553**	**0.0946**	**0.0740**	**0.0786**		DCT-PLS	0.5509	0.1407	0.1116	0.1149
NAQU	GSP	0.6707	**0.0924**	0.0732	**0.0673**	ORACLE	GSP	0.3048	**0.0958**	**0.0769**	**0.0955**
	U-TILISE	0.6017	0.0963	0.0740	0.0841		U-TILISE	0.1443	0.1161	0.0881	0.1147
	STpconv	0.6534	0.1006	0.0757	0.0788		STpconv	0.1984	0.1135	0.0886	0.1131
	DCT-PLS	**0.6811**	0.0928	**0.0708**	0.0705		DCT-PLS	**0.3080**	0.0996	0.0808	0.0995
OZNET	GSP	0.4727	0.1753	0.1420	**0.1172**	PBO_H2O	GSP	**0.4689**	0.0991	0.0843	0.0649
	U-TILISE	0.3535	**0.1729**	**0.1365**	0.1355		U-TILISE	0.3201	**0.0949**	**0.0769**	0.0748
	STpconv	0.4385	0.1836	0.1455	0.1307		STpconv	0.4430	0.1040	0.0879	0.0684
	DCT-PLS	**0.4829**	0.1923	0.1541	0.1271		DCT-PLS	0.4342	0.0984	0.0827	0.0672
REMEDHUS	GSP	0.5193	**0.0919**	**0.0772**	**0.0579**	SCAN	GSP	**0.4034**	0.0866	0.0725	0.0627
	U-TILISE	0.4180	0.0948	0.0784	0.0634		U-TILISE	0.3103	0.0887	0.0730	0.0693
	STpconv	0.5131	0.0942	0.0785	0.0610		STpconv	0.3130	0.0964	0.0805	0.0712
	DCT-PLS	**0.5871**	0.0944	0.0792	0.0593		DCT-PLS	0.3981	**0.0834**	**0.0700**	**0.0617**
SMN-SDR	GSP	0.5945	**0.0996**	**0.0778**	**0.0720**	SNOTEL	GSP	**0.3988**	**0.1271**	**0.1063**	**0.0939**
	U-TILISE	**0.7085**	0.1063	0.0846	0.0752		U-TILISE	0.2958	0.1319	0.1088	0.1048
	STpconv	0.1660	0.1379	0.1018	0.1147		STpconv	0.2759	0.1315	0.1083	0.1037
	DCT-PLS	0.5935	0.1018	0.0801	0.0731		DCT-PLS	0.3452	0.1300	0.1082	0.0975
SOILSCAPE	GSP	0.5864	0.1707	0.1561	0.0787	TAHMO	GSP	0.6610	**0.0448**	0.0371	0.0346
	U-TILISE	0.2870	**0.1664**	**0.1468**	0.1157		U-TILISE	0.6505	0.0463	0.0377	0.0356
	STpconv	**0.5925**	0.1814	0.1651	0.0832		STpconv	**0.6640**	0.0449	**0.0368**	**0.0345**
	DCT-PLS	0.5728	0.1681	0.1537	**0.0771**		DCT-PLS	0.6560	0.0460	0.0378	0.0349
TxSON	GSP	**0.4928**	**0.0798**	**0.0682**	**0.0536**	USCAN	GSP	**0.3940**	0.0859	0.0749	**0.0525**
	U-TILISE	0.4624	0.0810	0.0687	0.0561		U-TILISE	0.2904	**0.0835**	**0.0708**	0.0609
	STpconv	0.4343	0.0835	0.0701	0.0597		STpconv	0.3764	0.0879	0.0759	0.0554
	DCT-PLS	0.4882	0.0804	0.0683	0.0548		DCT-PLS	0.3719	0.0851	0.0733	0.0540
XMS-CAT	GSP	**0.5583**	**0.1234**	**0.1114**	0.0619	Mean	GSP	**0.5387**	**0.1025**	**0.0859**	**0.0703**
	U-TILISE	0.2616	0.1344	0.1172	0.0893		U-TILISE	0.4387	0.1084	0.0889	0.0834

STpconv	0.4879	0.1335	0.1189	0.0701	STpconv	0.4753	0.1107	0.0911	0.0795
DCT-PLS	<u>0.5493</u>	<u>0.1254</u>	<u>0.1134</u>	**0.0618**	DCT-PLS	<u>0.5350</u>	<u>0.1047</u>	<u>0.0872</u>	<u>0.0728</u>

"

Table 3

"**Table 3**

Accuracy evaluation of the reconstructed FY-3B SM dataset at the five dense networks (bolded and underlined values indicate the greatest and second greatest accuracy).

		R	**RMSE (m^3/m^3)**	**MAE (m^3/m^3)**	**ubRMSE (m^3/m^3)**			***R***	**RMSE (m^3/m^3)**	**MAE (m^3/m^3)**	**ubRMSE (m^3/m^3)**
BIEBRZA_S-1	FY-3B	0.4304	**0.1838**	**0.1677**	**0.0782**	**Ru_CFR**	FY-3B	0.5676	0.0874	0.0709	0.0871
	GSP	<u>0.5039</u>	<u>0.2017</u>	<u>0.1866</u>	<u>0.0800</u>		GSP	**0.5944**	**0.0854**	<u>0.0690</u>	**0.0851**
	U-TILISE	**0.5145**	0.2208	0.1988	0.0972		U-TILISE	0.5777	**0.0854**	<u>0.0690</u>	**0.0851**
	STpconv	0.4770	0.2093	0.1907	0.0874		STpconv	0.5876	<u>0.0860</u>	0.0700	<u>0.0856</u>
	DCT-PLS	0.4719	0.2045	0.1879	0.0826		DCT-PLS	<u>0.5939</u>	**0.0854**	**0.0688**	**0.0851**
SD_DEM	FY-3B	<u>0.3824</u>	<u>0.0545</u>	<u>0.0508</u>	0.0299	**SWEX_POLAND**	FY-3B	**0.7869**	<u>0.1025</u>	<u>0.0819</u>	**0.0740**
	GSP	0.3783	0.0557	0.0521	0.0298		GSP	0.7809	0.1032	0.0828	0.0744
	U-TILISE	**0.3968**	**0.0540**	**0.0505**	**0.0294**		U-TILISE	**0.7869**	<u>0.1025</u>	<u>0.0819</u>	**0.0740**
	STpconv	0.3810	0.0549	0.0513	0.0299		STpconv	<u>0.7837</u>	0.1026	0.0824	0.0742
	DCT-PLS	0.4167	0.0559	0.0521	<u>0.0296</u>		DCT-PLS	0.7831	**0.1012**	**0.0810**	<u>0.0741</u>
VAS	FY-3B	0.7614	**0.0647**	**0.0534**	**0.0445**	**Mean**	FY-3B	0.5858	**0.0986**	**0.0850**	**0.0627**
	GSP	**0.7785**	0.0662	0.0545	<u>0.0462</u>		GSP	<u>0.6072</u>	<u>0.1024</u>	<u>0.0890</u>	<u>0.0631</u>
	U-TILISE	0.7606	<u>0.0657</u>	<u>0.0536</u>	0.0471		U-TILISE	**0.6073**	0.1057	0.0908	0.0666
	STpconv	0.7178	0.0698	0.0574	0.0486		STpconv	0.5894	0.1045	0.0904	0.0652
	DCT-PLS	<u>0.7670</u>	0.0665	0.0549	0.0471		DCT-PLS	0.6065	0.1027	0.0890	0.0637

"

- Figure 12: Although I understand the premise for this figure, it is very hard to see any detail with so many panels. Consider another way to demonstrate the stability of spatial patterns.

  Thank you for the valuable comment. The 30 subplots in the original Fig. 12 were too small in size, making it difficult for readers to distinguish meaningful spatial details. To address this issue and more effectively demonstrate the spatial pattern of the reconstructed FY-3B SM, we first selected four representative dates at approximately 10-day intervals (Fig. 14 (a)–(d)). This significantly clarifies the spatial details while still demonstrating the model's capability to capture continuous daily variations. Furthermore, we calculated the monthly mean and standard deviation of the reconstructed FY-3B SM and precipitation (Fig. 14 (e)–(h)). The standard deviation of the reconstructed FY-3B SM falls within a reasonable range (0 to 0.16 m^3/m^3), quantitatively indicating that the overall spatial structure remained stable throughout the month. The spatial coincidence between regions of high SM fluctuation and active precipitation zones confirms that the localized variations captured by GSP are physically sound and driven by actual meteorological forcing mechanisms. Specific revisions can be found on Page 22, Lines 546-557 and Fig. 12 of the revised version.

Page 22, Lines 546-557

“Fig. 12 illustrates the spatial patterns of the reconstructed FY-3B SM and precipitation. (a)–(d) demonstrate that the reconstructed results maintain a consistent spatial structure while exhibiting certain variations, indicating the model's capability to capture continuous daily evolutions. Furthermore, the monthly mean of reconstructed FY-3B SM (Fig. 12 (e)) falls within a physically reasonable range of 0 to 0.5 m^3/m^3, and the standard deviation (Fig. 12 (f)) is constrained within 0 to 0.16 m^3/m^3. These quantitative results demonstrate that the overall spatial configuration remains stable, alongside the presence of localized daily fluctuations. To validate the physical rationale behind the reconstructed FY-3B SM dynamics, we introduced the monthly mean and standard deviation of precipitation for the same period (Fig. 12 (g)–(h)). It is evident that regions with a higher FY-3B SM standard deviation spatially consistent well with active precipitation zones. This consistency confirms that the GSP accurately captures the dynamic hydrological responses.”

Fig. 12

“

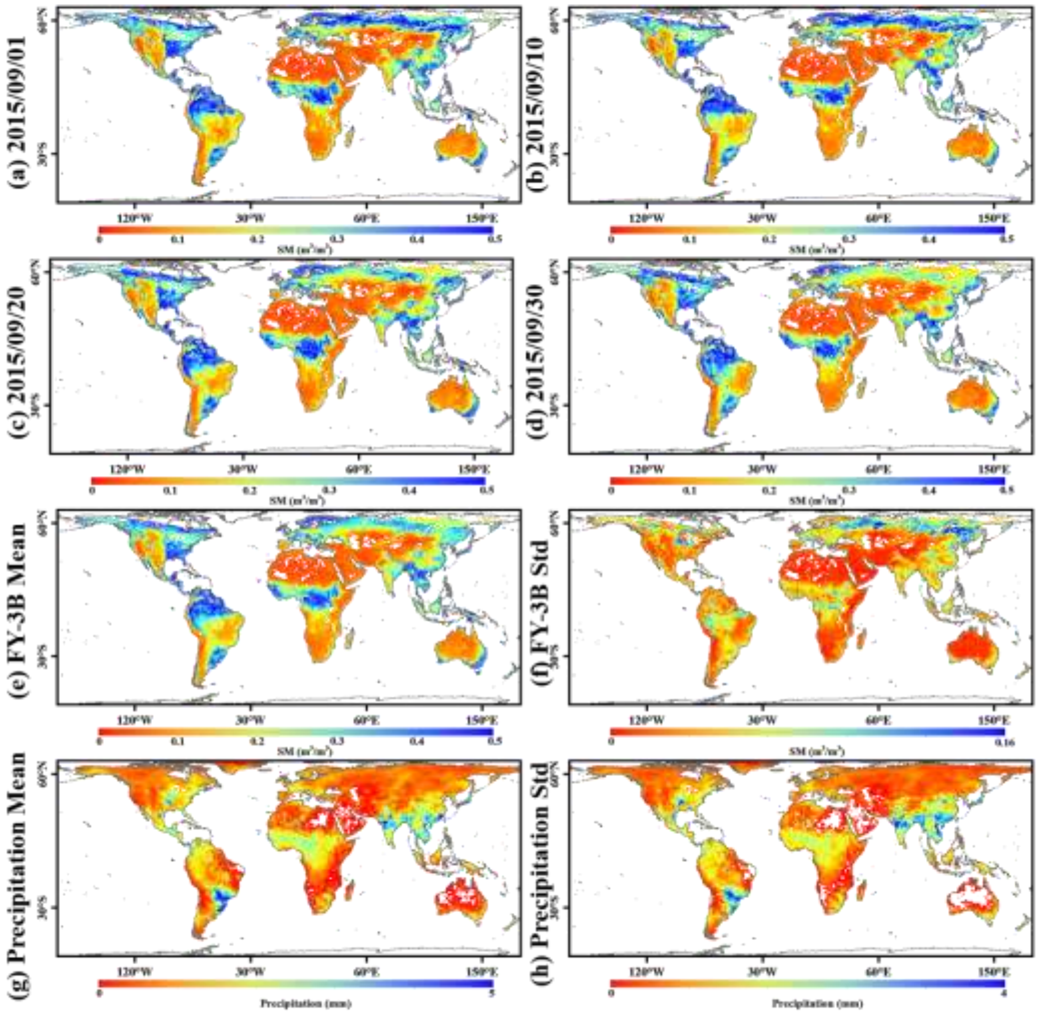


***Fig. 12.** Spatial distributions of the reconstructed FY-3B SM and precipitation in September 2015. (a)–(d) present daily reconstructed FY-3B SM on 1 September 2015, 10 September 2015, 20 September 2015, and 30 September 2015, respectively. (e)–(f) present the monthly mean and temporal standard deviation of FY-3B SM for the September 2015. (g)–(h) present the monthly mean and STD of precipitation for the same period.*

"

- Table 7: Consider doing the bold and underlined for the climate types with the best R, RMSE, MAE and ubRMSE values (like in previous tables)

Thank you for this comment. We have revised Table 6 (in the revised version) accordingly by bolding and underlining the best R, RMSE, MAE, and ubRMSE for each climate type.

"**Table 6**

Accuracy of the GSP-reconstructed FY-3B SM under different climate types (bolded and underlined values indicate the greatest and second greatest accuracy of all climate types).

	R	**RMSE**	**MAE**	**ubRMSE**
Af	0.9259	0.0419	0.0300	0.0416
Am	0.9384	0.0408	0.0291	0.0407
Aw	0.9643	0.0357	0.0248	0.0355
BWh	0.9333	**0.0154**	**0.0117**	**0.0118**
BWk	0.9395	0.0194	0.0141	0.0164
BSh	**0.9691**	0.0259	0.0175	0.0251
BSk	0.9558	0.0251	0.0169	0.0240
Csa	0.9480	0.0379	0.0251	0.0369
Csb	0.9595	0.0381	0.0271	0.0373
Csc	0.9140	0.0355	0.0267	0.0350
Cwa	0.9580	0.0361	0.0251	0.0361
Cwb	0.9549	0.0347	0.0246	0.0345
Cwc	0.9233	0.0318	0.0240	0.0298
Cfa	0.9512	0.0390	0.0279	0.0390
Cfb	0.9010	0.0543	0.0364	0.0542
Cfc	0.8910	0.0620	0.0472	0.0609
Dsa	0.9267	0.0216	0.0150	0.0202
Dsb	0.9658	0.0293	0.0196	0.0287
Dsc	0.9674	0.0396	0.0257	0.0393
Dsd	0.8812	0.0638	0.0443	0.0638
Dwa	0.9507	0.0404	0.0271	0.0402
Dwb	0.9562	0.0371	0.0251	0.0369
Dwc	0.9630	0.0388	0.0261	0.0388
Dwd	0.9037	0.0614	0.0427	0.0605

Dfa	0.9389	0.0432	0.0302	0.0432
Dfb	0.9447	0.0427	0.0285	0.0427
Dfc	0.9538	0.0438	0.0283	0.0438
Dfd	0.9288	0.0517	0.0347	0.0517
Mean	0.9496	0.0389	0.0272	0.0381

”

- Line 577: Remove “(i.e.,” and just have a “:”

Thank you for this comment. We have revised the original sentence accordingly on Page 27, Lines 654–659.

Page 27, Lines 654–659
“To validate the effectiveness of each block and parallel architecture in the GSP model, four ablation experiments were carried out: (1) with only the multi-scale gated convolution block, (2) with only the residual Swin Transformer block, (3) with the multi-scale gated convolution block and the residual Swin Transformer block in parallel without the CBAM block; and (4) with the multi-scale gated convolution block and the residual Swin Transformer block in sequential connection.”