Supplement of Atmos. Chem. Phys., 26, 5763–5780, 2026
https://doi.org/10.5194/acp-26-5763-2026-supplement

Atmospheric Chemistry and Physics
Open Access
EGU

Supplement of

Seasonality of the Quasi-biennial Oscillation signal in water vapor in the tropical stratosphere

Qian Lu et al.

Correspondence to: Jian Rao (raojian@nuist.edu.cn)

GDCM Data Completion Model

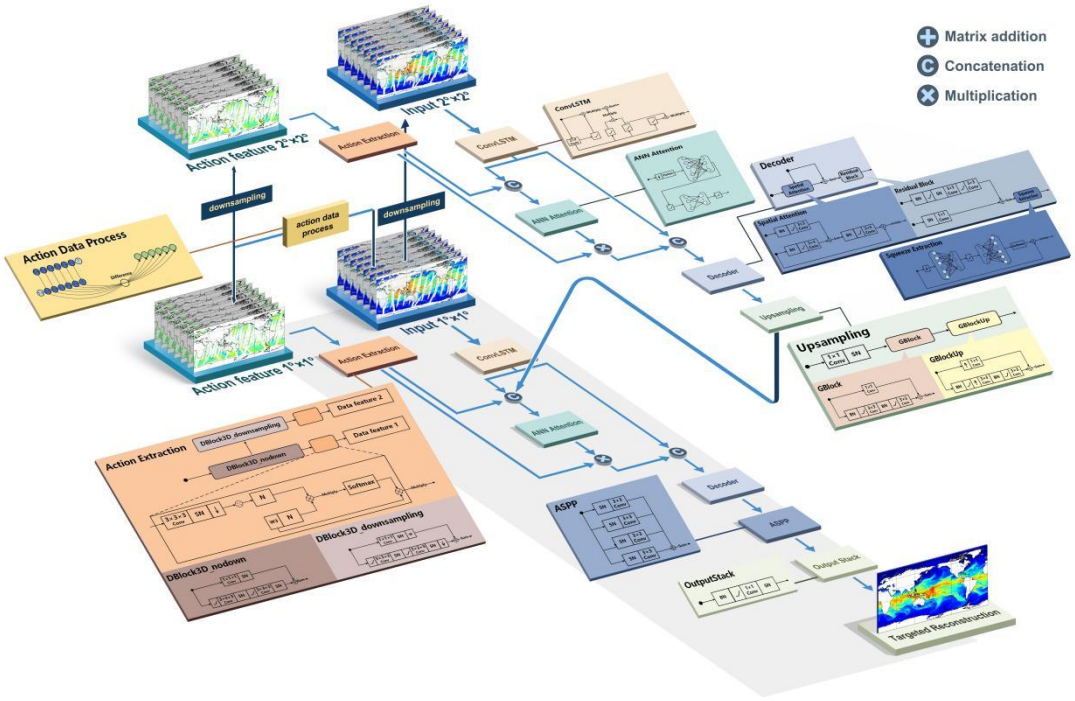


Figure S1. Architecture of the GDCM model. The model adopts an encoder-decoder structure and consists of spatiotemporal feature extraction blocks, spatiotemporal motion extraction blocks, spatiotemporal multi-source feature attention selection blocks, and an ASPP block.

The GDCM model adopts a dual-path encoding and multi-scale decoding architecture, which is suitable for the task of fine-grained reconstruction of spatio-temporal data. Features of different dimensions are processed separately in the encoding stage: the spatio-temporal state branch extracts data regularities layer by layer using ConvLSTM as well as downsampling; the dynamic behaviour branch captures motion features using a 3D convolution kernel and a step-size compression strategy. The outputs of the two branches enter the spatio-temporal multi-source feature attention selection block, and the features are filtered using the attention weight matrix.

The decoder architecture of the model firstly introduces the ASPP module, which extracts the contextual information through the convolutional kernels with different expansion rates in parallel; secondly, it adopts the residual attention mechanism, which calibrates the weights of the channels through the gating unit during the feature splicing process; and finally, it combines the cross-layer jump connection to ensure that the network still retains the details of the original data in the process of the

global feature integration. The design of the GDCM model is capable of capturing the wide range of spatial-temporal The GDCM model is designed to capture a wide range of temporal and spatial correlation patterns, capturing the subtle changes in key areas of the data, which significantly improves the accuracy of data complementation and the model generalisation ability in different complex environments. The following is a detailed description of each module of the GDCM model:

Spatiotemporal feature extraction blocks

In the spatio-temporal feature extraction module, the model uses a multi-level convolutional loop architecture to parse the input information. The innovation of this module lies in the fusion of traditional sequence modelling with spatial convolution, by replacing the fully-connected weights in the long and short-term memory network with a two-dimensional convolution kernel to form a ConvLSTM unit that can capture spatial features. The gating mechanism is implemented using a spatial convolution to replace the linear transformation, where the forgetting gate is implemented through a 2 $\times$ 2 convolution kernel that calculates the local spatial correlation between the current input and the historical state, and selects important features using a spatial weight matrix. The data input is gradually spatially dimensioned to form a multi-resolution feature pyramid, with each level corresponding to an independent ConvLSTM processing unit that incrementally extracts data features.

Spatiotemporal motion extraction blocks

The spatio-temporal action extraction module focuses on extracting the motion evolution features of the data, firstly by calculating the data differences between adjacent time steps to capture the more active regions of data changes, in this process, part of the spatio-temporal receptive field is constructed using cascaded $3 \times 3 \times 3$ convolution kernels to capture the local features of the short-term actions; the other part of the feature reorganisation is carried out through $1 \times 1 \times 1$ convolution to form the global motion basis vectors. After linear superposition of the two parts of the output, progressive resolution compression is used to halve the feature map size layer by layer by stepwise convolution, and convolution units are set up at each downsampling level. The outputs form multi-level features, with the shallow network being responsible for capturing the fast transform signals, and the deeper network extracting the motion variations. The model ultimately outputs spatio-temporal action features at two

different scales after normalisation.

Spatio-temporal multi-source feature attention selection blocks

The spatio-temporal multi-source feature attention selection module realizes the optimal selection of model features through weight allocation. The processing flow is divided into three steps: firstly, the input features are compressed in the channel dimension, and the two-dimensional spatial information of each feature channel is aggregated into single-value representations, which reduces the amount of data while retaining the core features of the data. Subsequently, a weight generation network is established. The network adopts the symmetric structure of the fully connected layer to analyse the compressed features and output the importance coefficients of each channel. These coefficients are normalised and multiplied channel by channel with the original features as a dynamic adjustment parameter to realise the reinforcement of the key information. The final output features focus on the spatio-temporal patterns that have the highest degree of contribution to the data reconstruction, which significantly improves the discriminative value of the feature expression while guaranteeing the computational efficiency. significantly improve the discriminative nature of feature expression.

ASPP block

The ASPP module adopts the adaptive spacing convolution strategy to enhance the ability of the model to extract features. Its core achieves multi-level feature capture by adjusting the sampling spacing of the convolution kernel, with small spacing convolution being able to focus on the local details of the texture information, and large spacing convolution covering a wider range of contextual correlations. The module is set up with four parallel convolution branches, which extract cross-scale features from the data by adopting different spacing parameters. The output of each branch is spliced and coupled with cross-scale features by single-point convolution. This design extends the network's understanding of spatial structure while keeping the number of parameters constant. It is suitable for the task of dealing with the needs of nested multi-scale physical processes in meteorological data.

GDCM model training strategy

In the GDCM training phase, the training input data include samples extracted from the reanalysed

dataset, binary matrices of missing data distributions corresponding to the remotely sensed data, and randomly generated strip data. We randomly select these three types of data for each training, choose the same batch size for each type, and then perform element-level multiplication operations to generate the final training samples. These samples are then fed into the model for training.

To enable the model to fully learn the spatio-temporal relationship of the input data, a progressive learning strategy is designed for the GDCM model. In the early stage of training (epoch=0), the input of the model is the complete reanalysed data for 7 consecutive days. As the training proceeds, the missing regions are introduced gradually, to make the proportion of the null values in the data increase gradually, and finally reach a match with the missing situation of the real satellite data. Throughout the process, the rate of increasing missing regions was set to 0.002% per epoch. This strategy allows the model to adapt to the complete data while gradually learning to deal with the missing data situation, thus improving its robustness, stability and generalisation ability, and making it more applicable to real data missing situations.

MPFNet inversion model

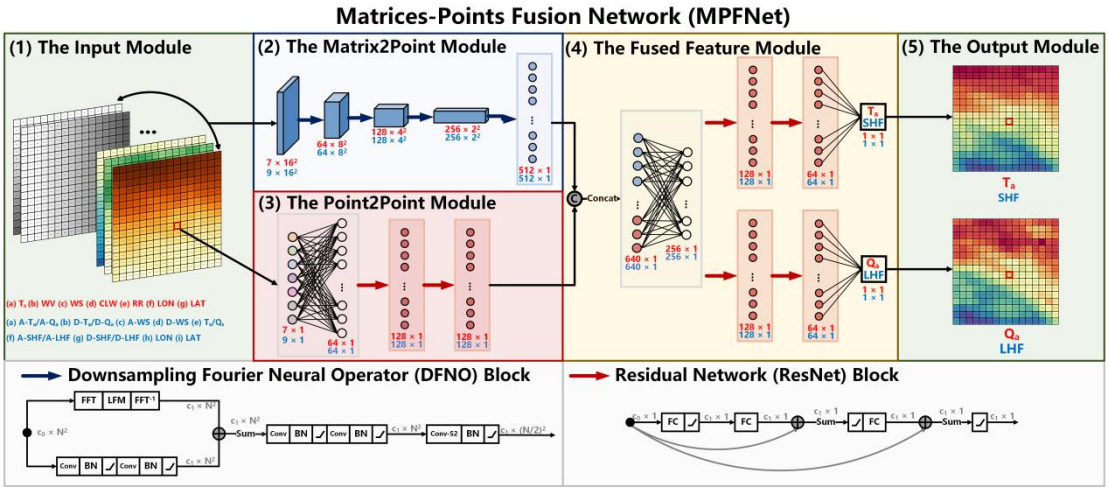


Figure S2. MPFNet model architecture. The red-colored components indicate the first-stage model used to retrieve T_a and Q_a. The blue-colored components represent the second-stage model, which uses the first-stage outputs along with other variables (A-T_a/A-Q_a, D-T_a/D-Q_a, A-SSW, D-SSW, T_s/Q_s, LON, and LAT) to compute the final SHF and LHF.

To enhance the retrieval accuracy of air-sea heat fluxes, we design a two-stage model architecture, where each stage targets a specific physical quantity. As illustrated in Fig. S2, the first-stage model

(highlighted in red) is dedicated to retrieving T_a and Q_a by integrating spatial and point-specific satellite information. The second-stage model (highlighted in blue) takes the outputs from the first stage—along with additional variables such as T_a, Q_a, SSW, T_s, Q_s, LON, and LAT—as inputs to compute the SHF and LHF based on bulk aerodynamic formulas.

The Input Block

The input data for the MPFNet model are derived from SSM/I observations filled and processed by the GDCM and contain five important air-sea variables: SSW, CLW, WV, RR, and SST. These data are reconstructed into 16 × 16 grid cells, where the grid centre at position (8, 8) coordinates is used as a geographic anchor for the point inputs. To enhance the model's ability to extract spatial features, the input feature set is extended with two physical dimension channels: the radial image resolution feature (f) characterises the geometrical properties of the observations, and the vertical image scale feature (g) describes the degree of discretisation of the atmospheric column structure. The stacked processing is constructed to form a 7 × 16 × 16 (9×16×16 in the second stage) array and fed into the Matrix2Point and Point2Point modules for feature extraction.

The Matrix2Point Block

The feature extraction core uses four downsampling Fourier neural operators (DFNO), and the input data are arranged in a 7×16×16 (9×16×16 in the second stage) gridded structure. The DFNO module consists of a frequency domain operation unit working in concert with a downsampling unit. The frequency domain operation unit contains three key steps: firstly, mapping the spatial data to the frequency dimension through fast frequency domain transformation, feature screening and linear transformation in the low frequency region, and finally Finally, the inverse transformation is restored to the spatial dimension, and this process can effectively extract a wide range of spatial correlation patterns. After the frequency domain operation, the module accesses a three-layer convolutional network to strengthen the local features, and each layer uses a 3 × 3 convolutional kernel with normalisation and activation function to gradually extract information. Feature dimensionality reduction uses a pooling layer with a step size of 2 to compress the feature map size so that the model focuses on the key spatial patterns. Through the DFNO module, the original input data is gradually

transformed into feature representations with high information density.

The Point2Point Block

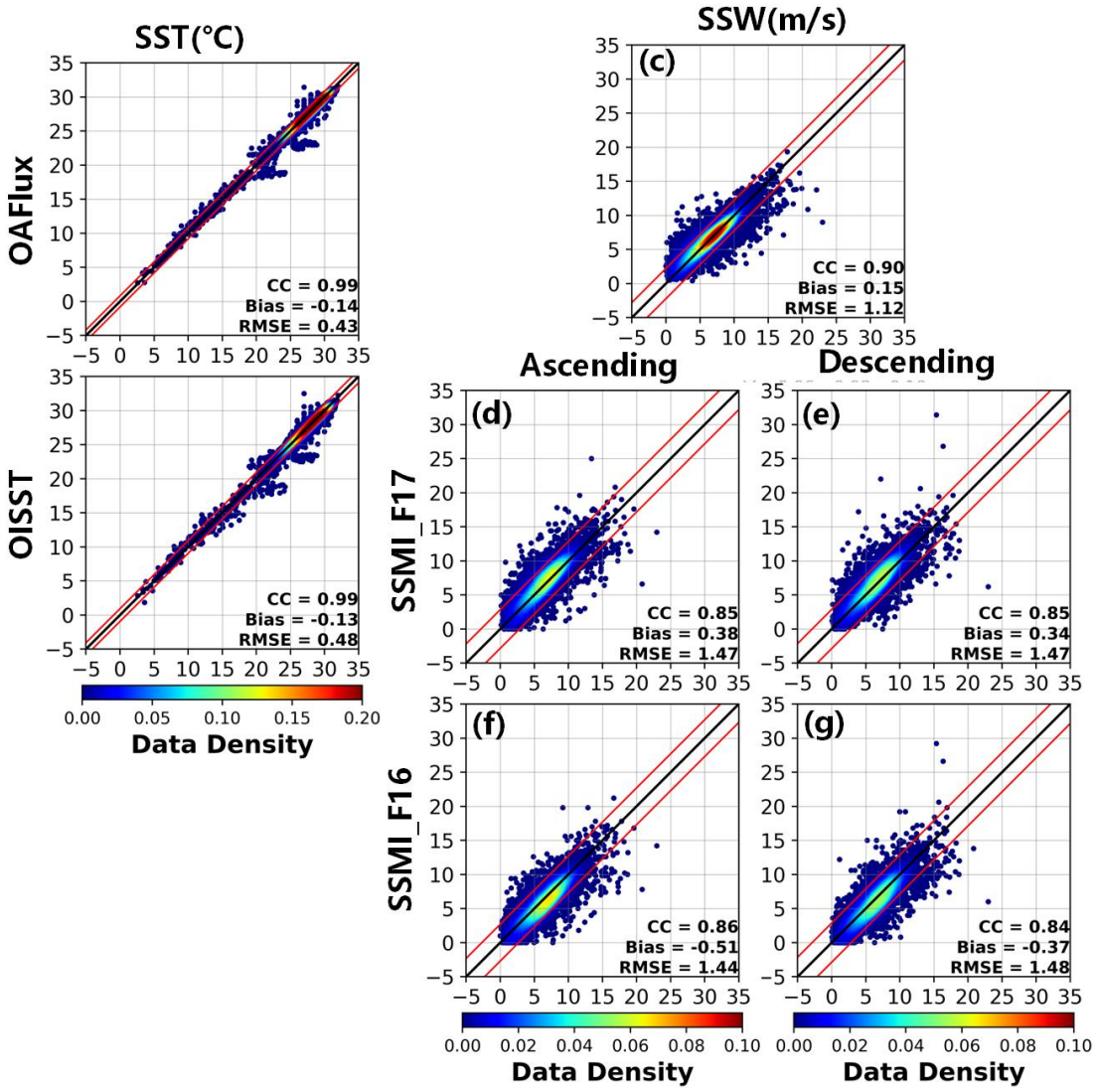
The Point2Point module is used to process 7-dimensional (9-dimensional in the second stage) feature data at a single point, where the input features are transformed into a base transformation using a fully connected layer and ReLU activation functions, and the features are mined by two Residual Network (ResNet) modules. The core of ResNet is the 'jump-junction' design - instead of learning the target output directly, the network learns the difference between the current output and the input (i.e., the residual). This structure makes deep network training more stable and avoids the problem of vanishing gradients. By superimposing multiple residual blocks, the model can flexibly capture complex patterns in the data, and is particularly good at dealing with highly non-linear sea-air relationships such as those found in satellite remote sensing data. This design significantly improves the accuracy of the model's estimation of parameters such as ocean temperature and humidity, allowing it to remain robust in complex scenarios.

The Fused Feature Block

The fusion feature module achieves efficient prediction by merging the information of two features, firstly splicing the spatial grid features (from Matrix2Point) with the point features (from Point2Point) to form a fused feature containing both global and local details, which is assigned to two independent branches: one branch specialises in predicting temperature (SHF in the second stage) and the other in predicting humidity (LHF in the second stage), and each branch consists of two segments The core structure consists of two segments - the front segment uses a neural network module with jump connections (ResNet) to extract deep regularities. The back segment maps the features to specific values through a fully connected layer. The sharing of fused features between the two branches enables the model to automatically learn the physical correlation between temperature and humidity (SHF and LHF in the second stage). This synergistic mechanism significantly improves the overall accuracy of parameter inversion.

The Output Block

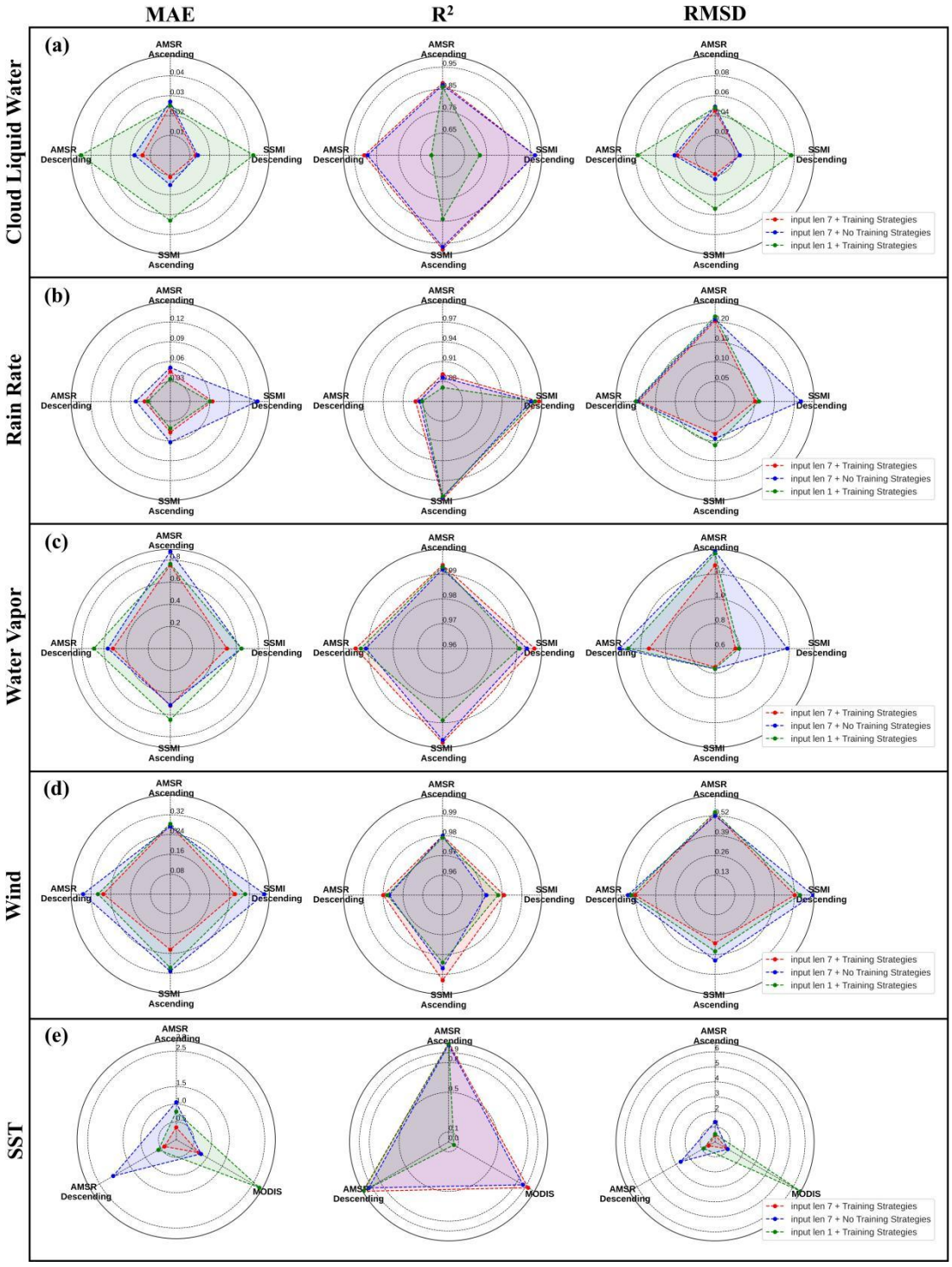
The proposed model architecture is applied in two stages. In the first stage, the model integrates spatial and point-specific information to retrieve global T_a and Q_a. In the second stage, these retrievals—along with other relevant variables including (a) A-T_a/A-Q_a (b) D-T_a/D-Q_a (c) A-SSW (d) D-SSW (e) T_s / Q_s (f) A-SHF/A-LHF (g) D-SHF/D-LHF (h) LON (i) LAT —are re-input into the model to correct for compounded uncertainties. These corrections are significant for SSW inputs, where significant discrepancies are observed in SSM/I retrievals compared to in situ measurements, as demonstrated in Fig. S3d-g. Based on the corrected inputs, the model then computes globally high-accuracy SHF and LHF using the bulk aerodynamic formulas.



155

Figure S3. Scatter plots comparing (a-b) SST and (c-g) SSW products with in situ measurements. Panels (a) OAFlux SST versus observations and (b) OISST versus observations. SSW comparisons include (c) OAFlux SSW, and (d-g) SSM/I F16-17 retrievals from both ascending and descending orbits versus in situ SSW. Black lines indicate 1:1 relationships, with CC, bias, and RMSE annotated for each comparison. The red

160 **line represents the symmetrical linear fit, indicating two standard deviations of the differences between the predicted and observed values, encompassing approximately 95% of the data points.**



165 **Figure S4. Comparison of evaluation indexes of five variables of AMSR2, SSMI, and MODIS under different training strategies.**

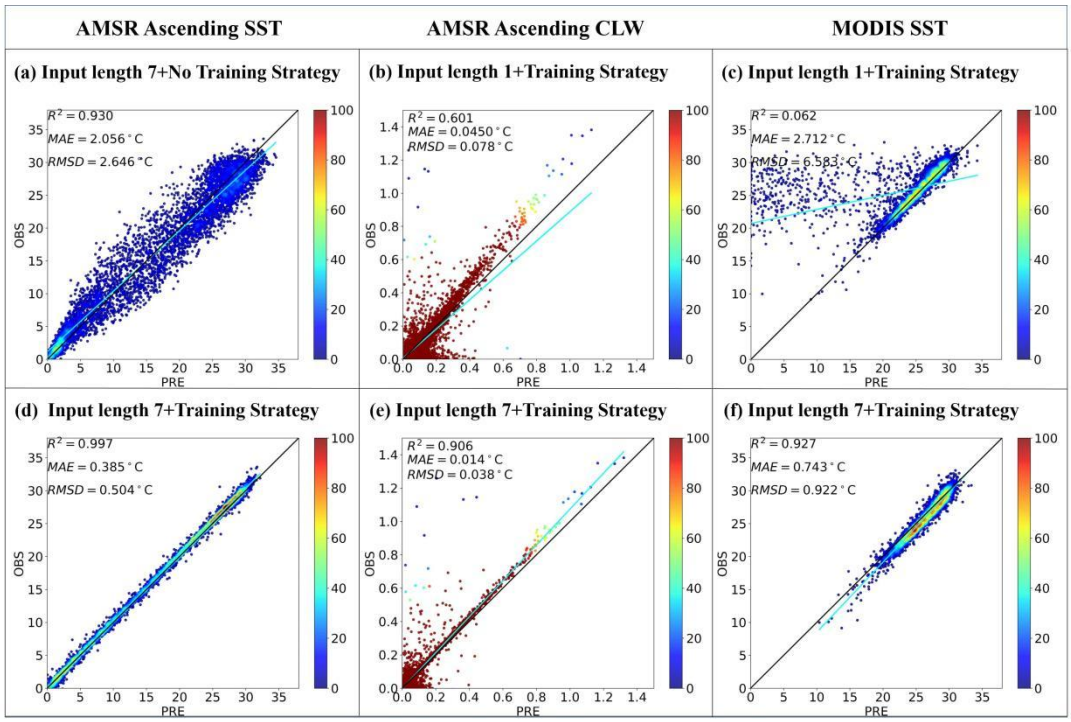
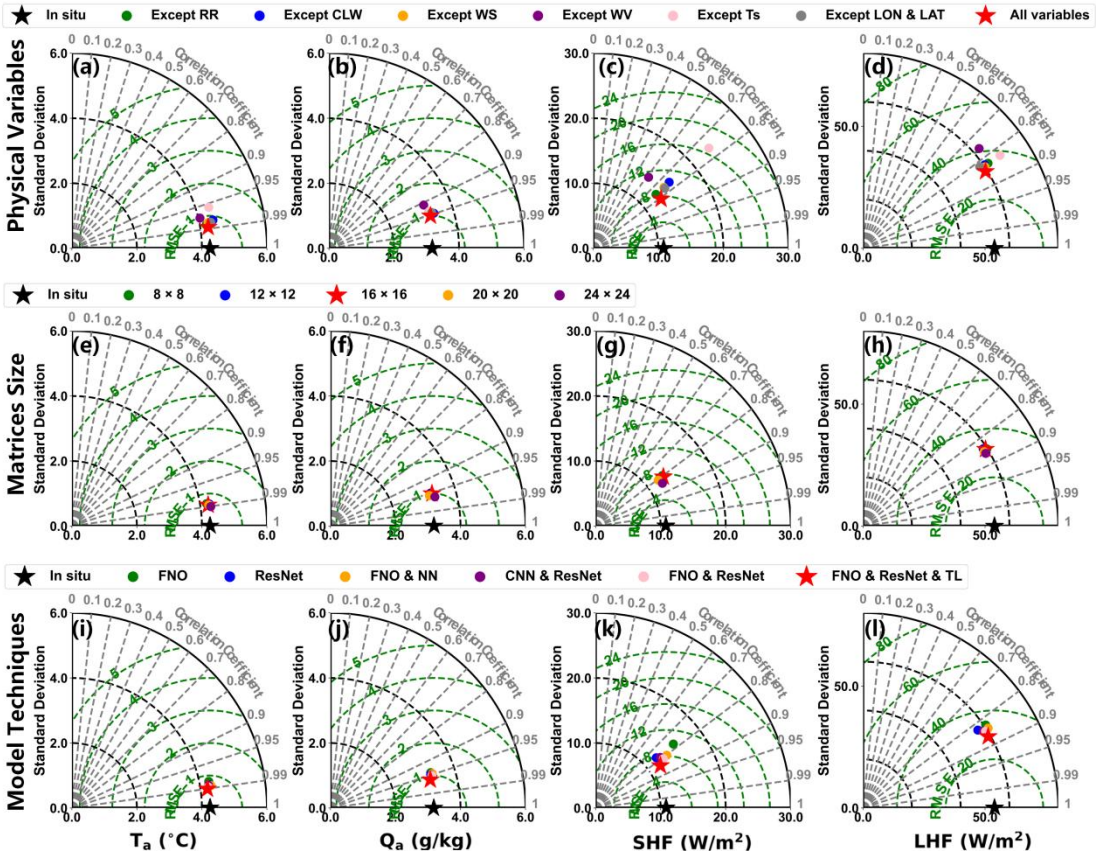


Figure S5. Scatter comparison plots of AMSR2, MODIS SST, and Cloud liquid water Descending data after GDCM model data complements.

170



FigureS6. Taylor Diagrams for three ablation studies on MPFNet are plotted using polar coordinates, with the radial axis representing STD and the angular axis representing CC. Green contours indicate the RMSE. Intercomparisons between in situ measurements (black pentagon) and variations in (a) – (d) variable selection, (e) – (h) matrix size, and (i) – (l) model techniques across test sets for , Q_a, SHF, and LHF. The red pentagons represent the final MPFNet model configuration determined by optimal variable selection, matrix size, and model techniques.

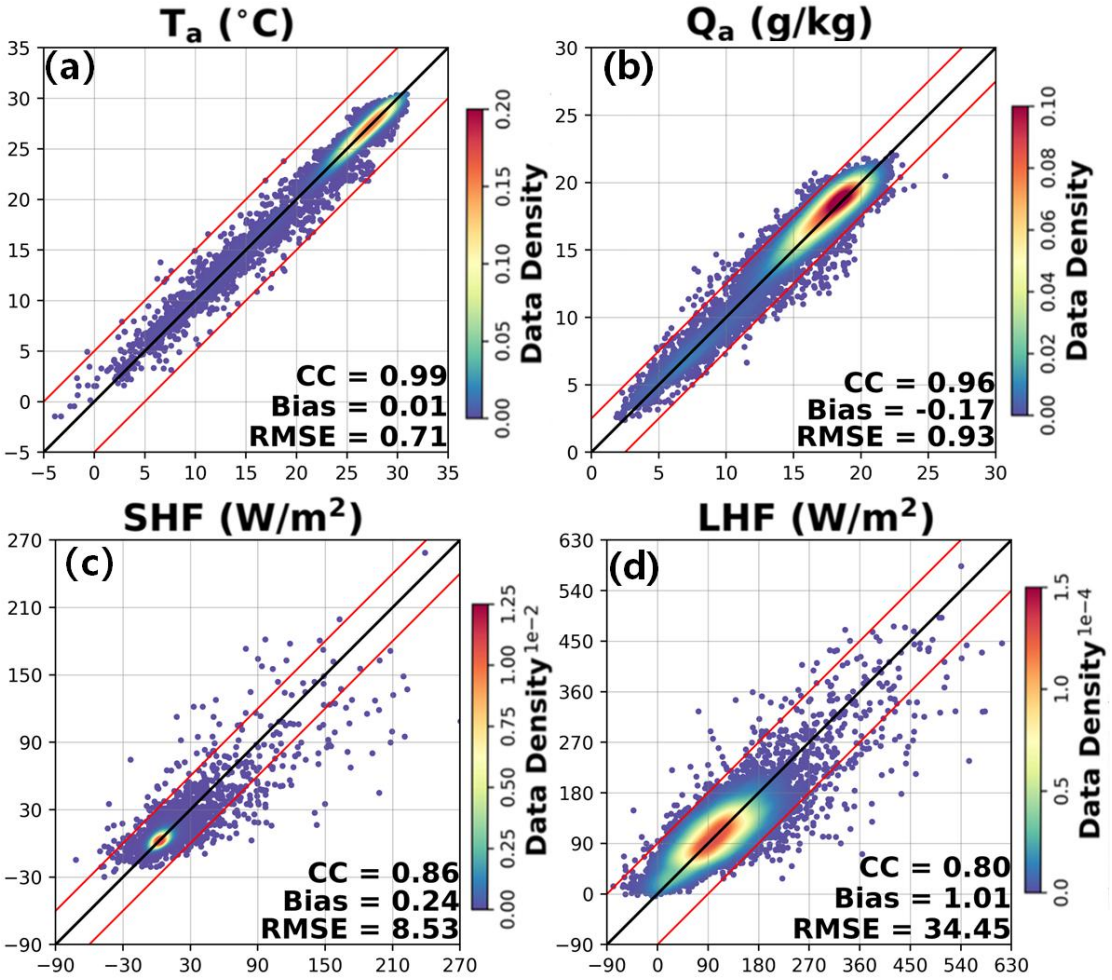


Figure S7. Scatterplots comparing retrieved values from the SSM/I-based fine-tuned MPFNet against in situ measurements for the test set, including (a) T_a, (b) Q_a, (c) SHF, and (d) LHF. The black line in scatterplots denotes the reference line with a slope of 1, indicating perfect agreement between the retrieved and observed values. The red line represents the symmetrical linear fit, indicating two STDs of the differences between the predicted and observed values, encompassing approximately 95% of the data points.

Table S1 Biases of the whole test area during 11 years (2012 to 2022) in different SST ranges for the GDCM completed AMSR2 SST data in the data-missing and non-data-missing parts and the AMSR2 SST data(ascending and descending orbit data).

Error (Bias, ℃)	SST range (℃)		
	All	10 to 20	> 20
Ascending-completed data in the data-missing part	1.05	0.32	0.22
Ascending-completed data in the non-data-missing part	0.21	0.22	0.29
AMSR2 ascending SST data	-0.09	-0.00	-0.17
Descending-completed data in the data-missing part	0.87	0.13	-0.03
Descending-completed data in the non-data-missing part	0.12	0.21	0.09
AMSR2 descending SST data	0.08	0.12	0.09

Table S2 Model effects of different training strategies

Variable Name	MAD	RMSD	R^2
AMSR2 SST Ascending + Training Strategies + input length 7	0.347	0.454	0.998
AMSR2 SST Descending + Training Strategies + input length 7	0.385	0.504	0.997
AMSR2 Water vapor Ascending + Training Strategies + input length 7	0.758	1.272	0.994
AMSR2 Water vapor Descending + Training Strategies + input length 7	0.520	1.134	0.995
AMSR2 Cloud liquid Ascending + Training Strategies + input length 7	0.023	0.045	0.878
AMSR2 Cloud liquid Descending + Training Strategies + input length 7	0.014	0.038	0.906
AMSR2 Rain rate Ascending + Training Strategies + input length 7	0.045	0.203	0.891
AMSR2 Rain rate Descending + Training Strategies + input length 7	0.039	0.192	0.891
AMSR2 Wind speed Ascending + Training Strategies + input length 7	0.278	0.533	0.980
AMSR2 Wind speed Descending + Training Strategies + input length 7	0.270	0.524	0.980
SSMI Water vapor Ascending + Training Strategies + input length 7	0.519	0.751	0.998
SSMI Water vapor Descending + Training Strategies + input length 7	0.533	1.134	0.995
SSMI Cloud liquid Ascending + Training Strategies + input length 7	0.011	0.019	0.978
SSMI Cloud liquid Descending + Training Strategies + input length 7	0.013	0.025	0.970

SSMI Rain rate Ascending + Training Strategies + input length 7	0.047	0.082	0.998
SSMI Rain rate Descending + Training Strategies + input length 7	0.064	0.101	0.996
SSMI Wind speed Ascending + Training Strategies + input length 7	0.225	0.316	0.993
SSMI Wind speed Descending + Training Strategies + input length 7	0.261	0.523	0.981
AMSR2 SST Ascending + input length 7	1.060	1.337	0.981
AMSR2 SST Descending + input length 7	2.056	2.646	0.930
AMSR2 Water vapor Ascending + input length 7	0.880	1.387	0.992
AMSR2 Water vapor Descending + input length 7	0.566	1.371	0.991
AMSR2 Cloud liquid Ascending + input length 7	0.027	0.049	0.869
AMSR2 Cloud liquid Descending + input length 7	0.018	0.041	0.891
AMSR2 Rain rate Ascending + input length 7	0.051	0.208	0.886
AMSR2 Rain rate Descending + input length 7	0.052	0.197	0.885
AMSR2 Wind speed Ascending + input length 7	0.273	0.520	0.980
AMSR2 Wind speed Descending + input length 7	0.352	0.572	0.977
SSMI Water vapor Ascending + input length 7	0.514	0.765	0.997
SSMI Water vapor Descending + input length 7	0.646	1.183	0.994
SSMI Cloud liquid Ascending + input length 7	0.015	0.024	0.967

SSMI Cloud liquid Descending + input length 7	0.014	0.025	0.969
SSMI Rain rate Ascending + input length 7	0.062	0.094	0.996
SSMI Rain rate Descending + input length 7	0.132	0.217	0.984
SSMI Wind speed Ascending + input length 7	0.312	0.429	0.987
SSMI Wind speed Descending + input length 7	0.380	0.645	0.972
AMSR2 SST Ascending + Training Strategies + input length 1	0.527	0.795	0.993
AMSR2 SST Descending + Training Strategies + input length 1	0.581	0.875	0.992
AMSR2 Water vapor Ascending + Training Strategies + input length 1	0.768	1.367	0.993
AMSR2 Water vapor Descending + Training Strategies + input length 1	0.692	1.300	0.993
AMSR2 Cloud liquid Ascending + Training Strategies + input length 1	0.025	0.048	0.860
AMSR2 Cloud liquid Descending + Training Strategies + input length 1	0.045	0.078	0.601
AMSR2 Rain rate Ascending + Training Strategies + input length 1	0.034	0.214	0.871
AMSR2 Rain rate Descending + Training Strategies + input length 1	0.034	0.200	0.881
AMSR2 Wind speed Ascending + Training Strategies + input length 1	0.284	0.545	0.979
AMSR2 Wind speed Descending + Training Strategies + input length 1	0.292	0.549	0.978
SSMI Water vapor Ascending + Training Strategies + input length 1	0.647	0.763	0.989
SSMI Water vapor Descending + Training Strategies + input length 1	0.645	0.794	0.991

SSMI Cloud liquid Ascending + Training Strategies + input length 1	0.033	0.054	0.840
SSMI Cloud liquid Descending + Training Strategies + input length 1	0.042	0.077	0.719
SSMI Rain rate Ascending + Training Strategies + input length 1	0.041	0.111	0.994
SSMI Rain rate Descending + Training Strategies + input length 1	0.060	0.111	0.990
SSMI Wind speed Ascending + Training Strategies + input length 1	0.298	0.369	0.984
SSMI Wind speed Descending + Training Strategies + input length 1	0.302	0.557	0.978
MODIS+ Training Strategies + input length 7	0.743	0.922	0.927
MODIS + input length 7	0.805	0.973	0.867
MODIS+ Training Strategies + input length 1	2.712	6.583	0.062

195