https://doi.org/10.5194/essd-2026-180
Preprint. Discussion started: 17 April 2026

Open Access Earth System Science Data Discussions

TPHH: A long-term (1901–2023) high-resolution (1/30°) near-surface humidity dataset for the Tibetan Plateau generated via spatial downscaling based on hybrid-structure deep learning

Zheng Jin[1], Zezhou Chen[2,3], Qinglong You[4], Zhaoxiang Liu[2,3], Jintao Zhang[5], Huan Hu[2,3], Ping Chen[2,3], Xiang Liu[2,3], Zipeng Wang[2,3], Kai Wang[2,3], Shiguo Lian[2,3], and Shichang Kang[6]

[1]College of Geography and Planning, Chengdu University of Technology, Chengdu 610059, China
[2]Data Science & Artificial Intelligence Research Institute, China Unicom, No. 21 Financial Street, Beijing 100013, China
[3]Unicom Data Intelligence, China Unicom, No. 21 Financial Street, Beijing 100013, China
[4]Department of Atmospheric and Oceanic Sciences & Institute of Atmospheric Sciences, Fudan University, Shanghai 200438, China
[5]Yunnan Key Laboratory of Plateau Geographical Process and Environmental Changes, Faculty of Geography, Yunnan Normal University, Kunming 650050, China
[6]Institute of Mountain Hazards and Environment, Chinese Academy of Sciences, Chengdu 610299, China

Correspondence: Zezhou Chen (chenzz51@chinaunicom.cn), Zhaoxiang Liu (liuzx178@chinaunicom.cn), and Shiguo Lian (liansg@chinaunicom.cn)

Abstract. The Tibetan Plateau acts as the "Asian Water Tower" and faces regional amplified warming compared to the global climate change baseline. Given the Tibet Plateau's pronounced alpine terrain, i.e., significant elevation gradients within short horizontal distances, studies on climate changes/dynamics over this mountainous region fundamentally depend on spatially high-resolution datasets. However, most of currently available spatially high-resolution datasets only extend back to the 1980s, with prolonged temporal coverage data of pre-satellite era remaining scarce, especially for near surface atmospheric humidity. Thus, our study implements a hybrid-structure-based deep learning framework to generate monthly 2 m specific humidity, 2 m temperature and surface pressure at $1/30° \times 1/30°$ horizontal resolution during 1901–2023. Briefly, employing a hybrid-structure model (FourCastNet by NVIDIA®), historical high-resolution fields ($1/30° \times 1/30°$ covering 1901–2023) are generated based on long-range low-resolution ($0.5° \times 0.5°$ covering 1901–2023 from CRU) and short-range high-resolution fields ($1/30° \times 1/30°$ covering 1978–2023 from TPMFD) via spatial downscaling. The produced data were validated against multiple related datasets, with independent in-situ site observations serving as the reference, and showed superior performance compared to most of them. Our study demonstrates that in topographically complex regions like the Tibetan Plateau, where meteorological fields exhibit strong physical dependencies on terrain, the synergistic mapping between total-field signals and subregional terrain constraints can be effectively achieved through hybrid-structure deep learning, thereby enabling this physically-consistent downscaling approach. Open access to this dataset is at https://doi.org/10.57760/sciencedb.36169.

https://doi.org/10.5194/essd-2026-180
Preprint. Discussion started: 17 April 2026

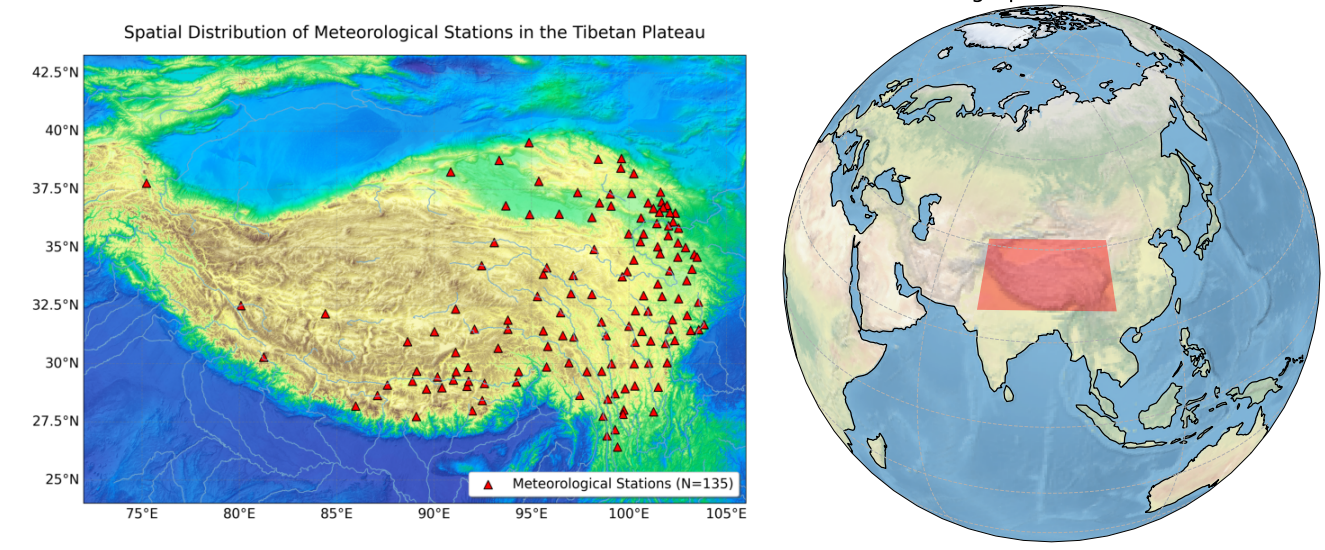


Figure 1. Geographical setting of the Tibetan Plateau. The map illustrates the complex topography and elevation gradients (left), and the study area's location within the broader Asian context (right). Spatial distribution of the 87 independent ground-based meteorological stations used for historical validation (1901–1978).

1 Introduction

The Tibetan Plateau (TP), hailed as the "Asian Water Tower" (Figure 1), is the source of major rivers in Asia, and provides water security and ecosystem services to billions of people residing in downstream regions (Yao et al., 2022). Moreover, the TP is a sensitive region to global climate change. Observational studies show a pronounced "Elevation-Dependent Warming" (EDW) pattern over the TP (Kang et al., 2010; Pepin et al., 2015; You et al., 2020). This EDW pattern results in a warming signature over the TP that is substantially higher than the global average, a phenomenon referred to as "TP amplification" (Zhang et al., 2023). This enhanced warming has triggered a series of environmental changes, including glacier retreat, permafrost degradation (Ran et al., 2018) and lake expansion (Song et al., 2014). Besides, discussions on wet/dry climate trends in this warming context have also received more attention (Dong et al., 2024; Yu et al., 2024). Thus, a long-term high-resolution climate baseline including both temperature and humidity from the early 20th century onwards is essential to characterizing these changes across the highly rugged terrain of TP.

In recent years, various studies have attempted to push gridded climate data toward higher spatial resolution and longer temporal coverage (Table 1). However, except for Yang et al. (2023), most dataset studies have relied on reference data from the 1950s onward. Moreover, current products are predominantly focused on surface air temperature and precipitation. High-resolution long-term datasets for near-surface humidity remain scarce, despite its importance for studying the TP's topographically complex land surface processes.

https://doi.org/10.5194/essd-2026-180
Preprint. Discussion started: 17 April 2026

Open Access Earth System Science Data Discussions

Table 1. Overview of recent algorithm-derived gridded products over the Tibetan Plateau.

Source	Variable	Time range	Resolution	Featured algorithm
Hu et al. (2025)	Precipitation	1951–2023	$0.1^\circ \times 0.1^\circ$	Light gradient boosting machine
Chen et al. (2022)	Rainfall erosivity	1950–2020	$0.25^\circ \times 0.25^\circ$	Inverse distance-weighted interpolation
Zhou et al. (2024)	Precipitation, Temperature	1961–2022	9 km × 9 km	Dynamic downscaling with spectral nudging by WRF
Yang et al. (2023)	Temperature	1901–2020	$0.1^\circ \times 0.1^\circ$	GAN-based Deep learning
Qin et al. (2023)	Temperature	1960–2020	$0.05^\circ \times 0.05^\circ$	Stacking-based ensemble learning
Qin et al. (2022)	Temperature	1961–2020	1 km × 1 km	Bayesian ridge regression for ensemble fusion

Since the beginning of the satellite observation era (1980s to present), regional high-resolution gridded datasets, such as the meteorological forcing dataset for the TP (TPMFD, (Jiang et al., 2025)) product developed through machine learning and multi-source data fusion, provide a high-fidelity benchmark for model training and performance validation. On the other side, the Climatic Research Unit Time-Series (CRU_TS_4.08, (Harris et al., 2020)), utilizing the climate anomaly method and angular-distance weighting, offers unparalleled temporal consistency extending back to 1901. Establishing a efficient spatial downscaling framework to map these centennial-scale CRU signals onto high-resolution grids (Figure 2), such as the TPMFD, would enable the reconstruction of a long-term, high-resolution climate record for the TP.

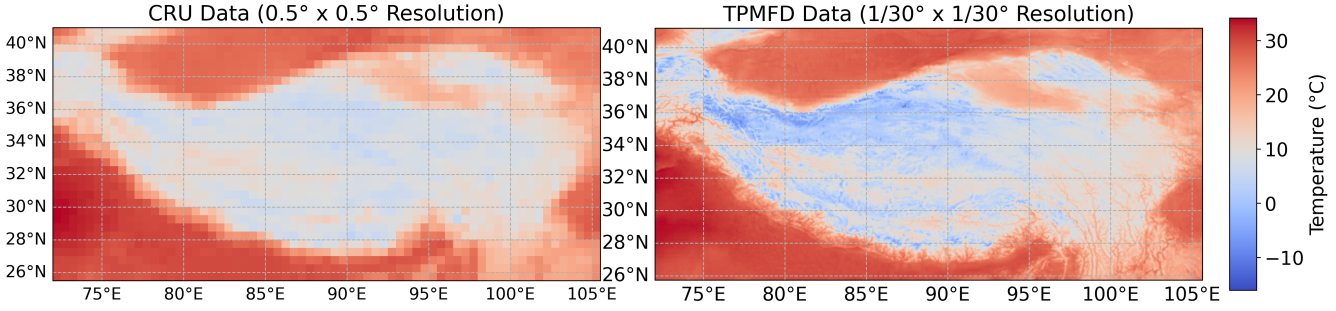


Figure 2. Visual comparison of spatial resolution between the predictor dataset and the target dataset. Left: Coarse-resolution CRU_TS_4.08 ($0.5^\circ \times 0.5^\circ$). Right: High-resolution TPMFD target ($1/30^\circ \times 1/30^\circ$). The contrast highlights the necessity for super-resolution reconstruction to capture local topographic climate details.

Bridging the gap between long-term coarse historical data and short-term high-resolution modern instrument data has long been a challenge, and is often dealt with by downscaling. Dynamical downscaling (e.g., RCM WRF) (Prein et al., 2015) has the advantage of physical consistency but is not now feasible to be used in centennial-scale simulations at convection-permitting (e.g., at 3 km × 3 km)) resolution due to extremely high computational demand. Meanwhile, Deep Learning (DL) has recently become a popular and effective alternative for statistical downscaling (Vandal et al., 2017) and climate reconstruction. DL models differ from conventional statistical approaches in that they are able to represent intricate, non-linear spatial relationships in large-scale atmospheric fields non-linearly. In this sense, FourCastNet (Pathak et al., 2022), a recently proposed global weather

https://doi.org/10.5194/essd-2026-180
Preprint. Discussion started: 17 April 2026

forecasting model exploiting the Transformer architecture developed by NVIDIA®, has several distinctive characteristics. Its architecture is based on the composition of Vision Transformers (ViT) Dosovitskiy (2020) and Adaptive Fourier Neural Operators (AFNO) Guibas et al. (2021), to efficiently capture not only local land surface process but also global atmospheric circulation patterns. Although it is primarily designed for forecasting, the ability to represent multi-scale physical fields also applies to the spatial refinement problems encountered in historical reconstruction.

There are two physical bases for achieving this downscaling over the TP. First is the terrain governance of the regional meteorological fields. Unlike plain areas where meteorological fields are primarily shaped by complex interactions across varying synoptic scales, the near surface temperature, humidity, and surface pressure all exhibit a strong and persistent dependency on the rugged terrain. Consequently, the terrain features themselves provide the primary source of high-frequency spatial information. Second is the synergistic global-local mapping capability of the FourCastNet model architecture. Rather than merely redistributing coarse-resolution ($0.5° \times 0.5°$) information onto finer grids ($1/30° \times 1/30°$) via static terrain-based allocation, this architecture employs AFNO and ViT to achieve cross-scale mapping. These algorithms enable the reliable inference of physically coherent local fields within a broader total-field context, ensuring that the reconstructed fields are both locally refined and synoptically consistent.

Therefore, we leverage the FourCastNet model to reconstruct the historical high-resolution climate fields over the TP. We utilize the long-term consistency of CRU_TS_4.08 (1901–2023) as the predictor and the spatial precision of the TPMFD dataset ($1/30° \times 1/30°$, $\sim 3\text{km} \times 3\text{km}$) as the ground truth target. By adapting the FourCastNet framework to learn the mapping between coarse-resolution historical patterns and fine-scale topographic climate features, we produced the TPHH (Tibetan Plateau Historical High) dataset. This dataset provides reconstructed monthly 2 m temperature, 2 m specific humidity, and surface pressure at $1/30° \times 1/30°$ resolution during 1901–2023. Comparative validations between TPHH and various related products against independent in-situ site-observations show TPHH has leading performances both for 2 m specific humidity and 2 m temperature.

2 Data

The reconstruction framework in this study is based on two datasets: CRU serving as the predictors and the TPMFD as the reference truth. The key attributes of these two datasets are detailed in the table below (Table 2).

2.1 Climatic Research Unit Time-Series

The variables used in this study were sourced from the CRU_TS_4.08 dataset (Harris et al., 2020), a widely used global gridded climate dataset produced by the Climatic Research Unit (CRU) at the University of East Anglia. The generation process for this dataset is both complex and rigorous, with its foundation being the monthly observations from thousands of meteorological stations worldwide. Its core construction methodology is analogous to an "anomaly" or "delta" method, involving the following steps:

https://doi.org/10.5194/essd-2026-180
Preprint. Discussion started: 17 April 2026

Open Access Earth System Science Data Discussions

Table 2. Overview of the predictor and reference datasets.

Attribute	Predictor Dataset (CRU_TS v4.08)	Reference Dataset (TPMFD)
Full Name	Climatic Research Unit Time-Series v4.08	Tibetan Plateau's Multi-source ground meteorological forcing dataset
Variables Used	tmp (mean temp), tmx (max temp), tmn (min temp), dtr (diurnal temp range), pre (precipitation), wet (wet days), frs (frost days), vap (vapour pressure), cld (cloud cover), pet (potential evapotranspiration)	temp (2 m temperature), shum (2 m specific humidity), pres (surface pressure)
Temporal Res.	Monthly	Monthly (aggregated from daily/hourly)
Temporal Coverage	1901–2023	1979–2023
Spatial Res.	$0.5^\circ \times 0.5^\circ$	$1/30^\circ \times 1/30^\circ$ (approx. 3 km × 3 km)
Spatial Coverage	Global land (excluding Antarctica)	Tibetan Plateau
Generation Method	Interpolation of station observations (ADW)	Fusion of WRF simulation, ERA5 reanalysis, and dense station observations
Source/Reference	Harris et al. (2020)	Jiang et al. (2025)

1) Calculation of station anomalies: Initially, observational series from over 5,000 weather stations globally are utilized. The monthly observational data from each station is converted into an anomaly series by comparing it against the station's monthly mean for the 1961–1990 standard period. Anomalies for temperature-based variables are calculated by subtraction, while precipitation anomalies are calculated by division.

2) Spatial interpolation: Then, the anomaly time series of all stations are interpolated on a 0.5º × 0.5º (30 arc-minute) grid over the global land surface.

3) Reconstitution of absolute values: Finally, these gridded anomaly results are merged with an independent, high-resolution 1961–1990 reference climatology product. This process converts the anomalies to absolute values of climate. Notably, the reference climatology dataset used for reconstitution itself accounts for the effects of geographical latitude, longitude, and elevation. The elevation data it incorporates is a 30 arc-minute resolution Digital Elevation Model (DEM), which was derived by averaging a finer 5 arc-minute global DEM. It is through this mechanism that the CRU dataset can effectively represent the influence of topography and orographic effects on the spatial distribution of climate variables at its $0.5^\circ \times 0.5^\circ$ resolution. The dataset has a temporal span from January 1901 to December 2023. In this work, all 10 variables were considered as input

https://doi.org/10.5194/essd-2026-180
Preprint. Discussion started: 17 April 2026

features to the model: mean near-surface air temperature (tmp), maximum air temperature (tmx), minimum air temperature (tmn), diurnal temperature range (dtr), precipitation (pre), wet-day frequency (wet), frost-day frequency (frs), vapor pressure (vap), cloud cover (cld), and potential evapotranspiration (pet).

2.2 Tibetan Plateau's Multi-source ground meteorological forcing dataset

The TPMFD dataset was used as the "ground truth" reference for model training and validation in this study. It is a high-resolution ground meteorological forcing dataset developed specifically for the Tibetan Plateau, with a temporal coverage from 1979–2023 and a spatial resolution as high as 1/30º. Its high accuracy stems from its unique generation method: 2 m temperature, 2 m specific humidity, surface pressure, and wind speed are derived from the fusion of short-term high-resolution WRF simulations, long-term ERA5 data, and dense ground station observations (Jiang et al., 2025). This multi-source data fusion strategy effectively combines the physical consistency of dynamical models, the spatiotemporal continuity of reanalysis data, and the ground-level reality of station observations, making it an ideal benchmark for training and evaluating the reconstruction model in this study. We use the 2 m temperature (temp), 2 m specific humidity (shum), and surface pressure (pres) from TPMFD as the target variables for reconstruction. This dataset is available through the https://doi.org/10.57760/sciencedb.36169.

2.3 Site observations

To evaluate the performance of our reconstruction, we collected monthly mean 2 m temperature, surface pressure, and relative humidity from 135 quality-controlled site-observations over the Tibetan Plateau. These site-observations have been validated for consistency and homogenized to ensure long-term reliability (Cao et al., 2016; Zhu et al., 2015). The spatial distribution of the collected sites within the Tibetan Plateau is shown in **Figure 1**.

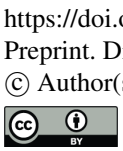
The site-observations were divided into two periods for evaluation. During the historical validation period (prior to 1979), 125 stations contained valid data, yielding 28,800 site-month records for temperature, 25,570 for pressure, 28,077 for relative humidity, and 25,422 for specific humidity. For the calibration period (1979–2023), 125 stations provided 68,840 records for temperature, 70,411 for pressure, 68,714 for relative humidity, and 68,661 for specific humidity. Detailed metadata for the station data in both periods are provided in an Excel (.xlsx) table documentation in the supplementary materials, including the start time, end time, and the exact number of valid records for each variable at each station.

2.4 Comparable datasets

To comprehensively evaluate the reliability of the reconstructed TPHH dataset, we compared it with a wide range of related products, including reanalysis datasets, gridded observations, and Global Climate Model (GCM) simulations.

In addition to ERA5-Land (Muñoz-Sabater et al., 2021) and 20CRv3 (Slivinski et al., 2019), we included two widely used reanalysis datasets: the Japanese Reanalysis for Three Quarters of a Century (JRA-3Q) (Harada et al., 2025) and the NCEP/NCAR Reanalysis 1 (NCEP-R1) (Kalnay et al., 1996). For gridded observations, we utilized the UDel (University of Delaware) air temperature and precipitation dataset (V5.01) (Willmott, 2000). For downscaling-based reconstructions, we utilized the

https://doi.org/10.5194/essd-2026-180
Preprint. Discussion started: 17 April 2026



Open Access Earth System Science Data Discussions

1-km monthly temperature developed by Peng et al. (2019). Furthermore, to assess the model's performance against traditional climate models, we selected historical simulations from six GCMs participating in the Coupled Model Intercomparison Project Phase 6 (CMIP6) (O'Neill et al., 2025): CESM2, UKESM1-0-LL, MPI-ESMI-2-HR, IPSL-CM6A-LR, BCC-CSM2-MR, and CanESM5. All CMIP6 data were bilinearly interpolated to the target resolution for consistent comparison.

3 Methodology

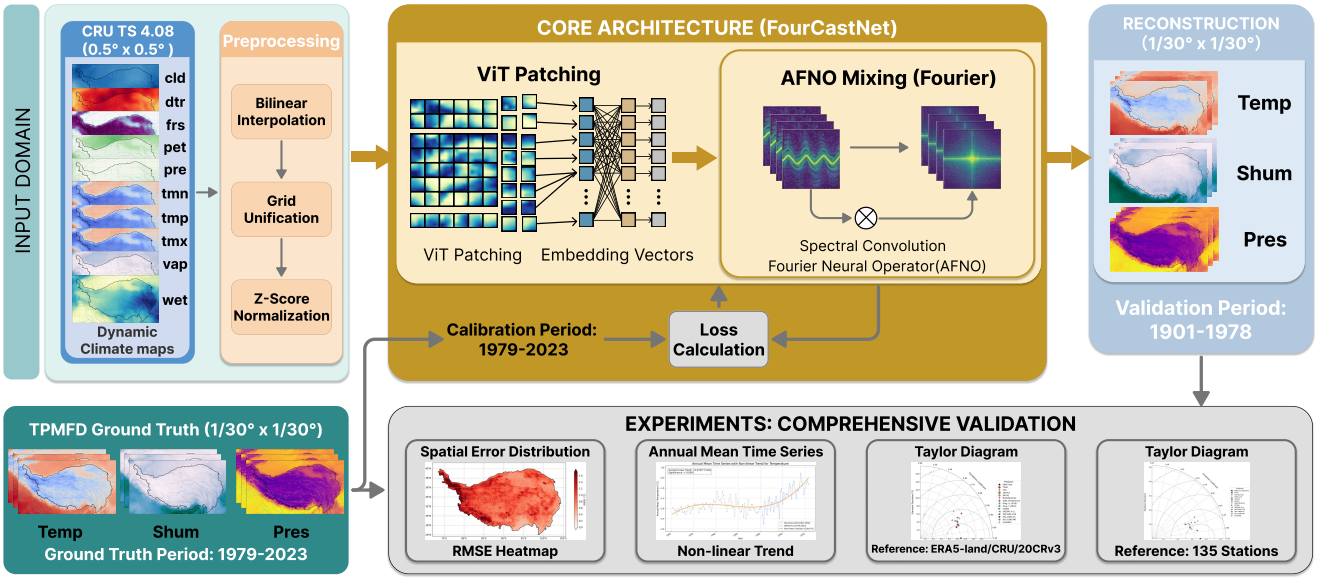


Figure 3. Schematic diagram of the methodology to generate the TPHH. The pipeline begins at the input domain with coarse climate fields. Major components are (1) Input processing, where coarse climate variables are prepared and embedded, (2) The main FourCastNet architecture, based on a ViT (Vision Transformer) backbone for patching and an AFNO (Adaptive Fourier Neural Operator) as local-global information mixing operator in the frequency domain, (3) Training with TPMFD ground truth (1979–2023) through loss computation; and (4) The reconstruction output based on validation over spatial error maps, long-term trends and independent station observations (validation base).

3.1 Hybrid-structure Downscaling Model: FourCastNet

An efficient downscaling model is key to bridging the resolution gap. In this study, we employ FourCastNet as the core downscaling model for handling the complex spatiotemporal physical fields over the TP. As a data-driven weather model, FourCastNet is one of the state-of-the-art deep learning models originally designed for global machine-learning-based weather forecasting. Its architecture combines the strengths of computer vision and Fourier neural operators, making it particularly suitable for the climate data downscaling task in our study. Its core advantage lies in its ability to efficiently learn multi-scale

https://doi.org/10.5194/essd-2026-180
Preprint. Discussion started: 17 April 2026

spatial features and total-field dependencies, which is crucial for downscaling climate fields from coarse CRU gridded data over the complex terrain of the TP.

The hybrid-structure of the model is primarily composed of two key components: a Vision Transformer (ViT) backbone and the Adaptive Fourier Neural Operator (AFNO) serving as the core mixing layer.

3.1.1 Vision Transformer (ViT) Backbone

FourCastNet leverages successes from the field of computer vision, treating the input two-dimensional multi-channel climate field (where each channel represents a CRU variable like temperature, precipitation, etc.) as an "image". The ViT backbone first divides this "image" into a series of non-overlapping, fixed-size patches. Each patch is then linearly projected into a one-dimensional vector. This approach offers two main advantages:

1) Dimensionality reduction and efficiency: By aggregating pixels into patches, the model significantly reduces the spatial resolution of the input data, transforming high-dimensional climate fields into a manageable sequence of tokens. This lowers the computational complexity, making it feasible to process high-resolution data.

2) Global context modeling: Transforming the 2D grid into a sequence of tokens allows the subsequent Transformer architecture to effectively capture long-range dependencies between distant patches, rather than being limited to local receptive fields.

3.1.2 Core Mixing Layer: Adaptive Fourier Neural Operator (AFNO)

In a standard ViT model, the mechanism for mixing information between different patches is multi-head self-attention. However, its computational complexity grows quadratically with the number of patches, making it computationally expensive for high-resolution climate data.

The core innovation of FourCastNet is the use of AFNO as a replacement mixing layer. The principle of AFNO is based on the Fourier Neural Operator (FNO) (Li et al., 2023), which performs computations in the Fourier domain to capture global information with remarkable efficiency. The basic process is as follows:

1) Fourier transform: Initially the information in the spatial domain, i.e., the sequence of patches, is converted to the frequency domain via a Fast Fourier Transform (FFT).

2) Frequency-domain learning: The model, in the frequency domain, learns the correlations of different frequency modes by channel mixing. Additionally, comparing to the local receptive fields in Convolutional Neural Networks (CNNs), the basic functions for the Fourier transform (sine and cosine waves) are global. Hence, an operation in the frequency domain inherently has a global spatial processing.

3) Inverse Fourier transform: At last, the learned and modified frequency domain representation is mapped back to the spatial domain by an Inverse Fast Fourier Transform (IFFT).

The "Adaptive" characteristic of AFNO is an improvement upon FNO. It introduces a soft-thresholding and sparsity mechanism, which allows the model to adaptively filter and retain the frequency information that is most important for the reconstruc-

https://doi.org/10.5194/essd-2026-180
Preprint. Discussion started: 17 April 2026

tion task, rather than simply truncating high-frequency signals as in the original FNO. This makes AFNO more computationally efficient and powerful.

3.1.3 Model Application in This Study

Within the framework of this study, FourCastNet operates as follows:

1) Input: A 3D tensor of shape (number of latitudes) $\times$ (number of longitudes) $\times$ (number of input variables), interpreted as a multi-channel 2D spatial grid field. Each channel represents a climate variable, for instance temperature, vapor pressure, or cloud cover, from the CRU dataset.

2) Processing: Prior to patch embedding, the coarse CRU input is spatially interpolated to match the high-resolution target grid. The upsampled CRU data is then patched by the ViT backbone and processed through multiple stacked layers of AFNO mixers. The model learns the complex, non-linear physical and statistical relationships between the input and target variables in both the frequency and spatial domains.

3) Output: A multi-channel 2D array with the same high spatial resolution as TPMFD ($1/30^\circ \times 1/30^\circ$). Each output channel corresponds to a target variable we aim to reconstruct (2 m temperature, 2 m specific humidity, surface pressure).

3.2 Reconstruction Workflow

The workflow of our approach is visually displayed in Figure 3, showing the online processing from I/O processing to final validation. As shown in Figure 3, the reconstruction pipeline is composed of four successive steps:

1) Domain of input and optimization: The information used is dynamic climate maps from CRU. Next, following a bilinear interpolation, the coarse climate fields are concatenated to form the multi-channel input tensor for the deep learning model.

2) Core architecture (FourCastNet): The refined features are input into the FourCastNet model. The model employs a ViT Patching module, which splits the grid into local patch tokens and embeds these tokens into vectors, then the AFNO Mixing applies Fast Fourier Transforms (FFT) and spectral convolutions for an efficient capture of total-field information in the frequency domain.

3) Training and loss computation: We train the model to predict these features from the high-resolution ground truth from the TPMFD dataset (1979–2023) to minimize the downscaling loss.

4) Output and validation: Model outputs during 1979–2023 are validated against TPMFD via k-fold validation, and historical outputs (1901–1978) are validated against site-observations. This is subjected to a multi-dimensional validation encompassing spatial error heat map analysis, long-term trend agreement among different products and direct comparison with 87 independent observation sites.

Below are more detailed descriptions for the individual steps.

https://doi.org/10.5194/essd-2026-180
Preprint. Discussion started: 17 April 2026

3.2.1 Data Pre-processing

1) Spatial interpolation: For the spatial resolutions of the predictor and predictand data to be consistent, the original $0.5^\circ \times 0.5^\circ$ CRU data was downscaled to the target resolution of $1/30^\circ \times 1/30^\circ$ of the TPMFD data via bilinear interpolation.

2) Data normalization: To avoid the effect of differing scales of variables and make the model converge faster, we normalized all the input data. In particular, for each grid point we computed its monthly mean and standard deviation during 1979–2023. Then, each value on that grid was normalized by subtracting the mean and dividing by the standard deviation. Such per-grid-cell normalization is important to ensure that the local climatology for each point is preserved while training.

3.2.2 Model Calibration

1) Calibration period: We took advantage of the intersection period of the two data sources, 1979–2023 for the model training and cross-validation (45 years in total).

2) Cross-validation: To exploit data to its best and to get a reliable assessment of performance, we considered a k-Fold cross-validation scheme with k=9. The 45-year data was split into 9 folds, each fold had the span of 5 years. Then, the model was trained on 8 folds (40 years) and tested on 1-fold (5 years) in an iterative manner. This procedure was repeated 9 times so that the data from each year was used as the validation data once. The overall model performance which is reported at the end is averaged over the 9 folds.

3.2.3 Historical Validation

Upon completion of the training and validation phase, the best-performing model configuration from the cross-validation was selected as the final model for reconstruction. The CRU data for the period 1901–1978 were pre-processed using the same normalization parameters derived from the 1979–2023 period. This historical data was then fed into the trained model to generate the high-resolution ($1/30^\circ \times 1/30^\circ$), monthly reconstructions of 2 m temperature, 2 m specific humidity, and surface pressure.

3.3 Accuracy Metrics

To evaluate the quality of the reconstruction model quantitatively, we compared the simulated (reconstructed) data with the reference data from the TPMFD dataset in the calibration period (1979–2023). A suite of common statistical metrics was applied, namely coefficient of determination (R^2), Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and the Index of Agreement (IOA).

In the following equations, S_i is the simulated value at time step i, O_i is the observed (reference) value, and $\bar{S}$ and $\bar{O}$ are the mean of the simulated and observed values, respectively. The number of all samples is n.

1) Coefficient of determination (R^2): This measure indicates how much variance in the observed data can be accounted for by the model. It shows how strongly linear simulated and observed values are related. Larger values of R^2 indicate better model

https://doi.org/10.5194/essd-2026-180
Preprint. Discussion started: 17 April 2026

fits. Its value lies between 0 and 1, and 1 denotes a perfect correlation.

$$R^2 = \left(\frac{\sum_{i=1}^{n}(S_i - \bar{S})(O_i - \bar{O})}{\sqrt{\sum_{i=1}^{n}(S_i - \bar{S})^2}\sqrt{\sum_{i=1}^{n}(O_i - \bar{O})^2}} \right)^2 \quad (1)$$

2) Root Mean Square Error (RMSE): RMSE is the standard deviation of the residuals (prediction errors, observed - simulated). It can be interpreted as the average size of the error in the same units as the predicted variable. Smaller RMSE is better, and it means a more accurate model prediction. The best value is 0.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(S_i - O_i)^2} \quad (2)$$

3) Mean Absolute Error (MAE): The MAE measures the average size of the errors in a set of forecasts, without considering their direction. In contrast to Mean Bias, it does not take into account the sign of the error (i.e., whether the prediction is too high or too low). It gives a clean and unambiguous indication of how large, on average, the errors are. A smaller MAE value indicates the better performance of the model with the best value of 0.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|S_i - O_i| \quad (3)$$

4) Index of Agreement (IOA): The IOA is a standardized measure of the degree of model prediction error. It is more sensitive to differences in the observed and simulated means and variances than the R^2 value. The IOA ranges from 0 to 1, where 1 indicates perfect agreement between the model and the observations, and 0 indicates no agreement.

$$\text{IOA} = 1 - \frac{\sum_{i=1}^{n}(S_i - O_i)^2}{\sum_{i=1}^{n}(|S_i - \bar{O}| + |O_i - \bar{O}|)^2} \quad (4)$$

3.4 Spatial, Temporal, and Statistical Evaluation of the Reconstructed Dataset

In addition to the primary accuracy metrics, a series of graphical and comparative analyses were conducted to provide a more comprehensive, multi-faceted evaluation of the reconstructed dataset. These evaluations assess the model's performance in the spatial, temporal, and statistical domains, comparing its output against both the primary reference dataset and other established climate data products.

1) Density scatter plots: To evaluate the validity of point-to-point matching and error distribution (1979–2023) for 2 m temperature (temp), 2 m specific humidity (shum), and surface pressure (pres), density scatter plots were used. The first shows the predictions of the nine individual models produced by k-fold cross-validation against the true values of the TPMFD reference dataset. This plot can be useful in spotting systematic errors, examining scatter, and also the overall distribution concentrated the 1:1 line of the perfect agreement.

2) Long-term time series validation: To evaluate the temporal consistency and stability of the reconstruction, time series plots were created for temp, shum, and pres covering the entire 1901–2023 period. These plots compare the output from our

https://doi.org/10.5194/essd-2026-180
Preprint. Discussion started: 17 April 2026

final ensemble model for the historical validation period (1901–1978) with the ground truth data from the TPMFD dataset for the calibration period (1979–2023). The primary goals of this analysis are to visually inspect for any artificial discontinuities at the 1978/1979 boundary and to assess the model's ability to capture plausible long-term trends and inter-annual variability.

3) Spatial distribution of model performance: To investigate the spatial differences in model performance, maps of the four key performance metrics (R^2, RMSE, MAE, and IOA) were constructed. This analysis is done for all the three prediction variables (temp, shum, and pres). With these maps we can see if there are particular areas of the TP for which the model is more or less skillful, and so gain some insight into the spatial nature of the model's performance.

4) Taylor diagrams for inter-comparison: To offer a concise statistical summary of model performance, Taylor diagrams were employed. This analysis was performed in two distinct ways. First, a Taylor diagram was constructed using the 87 independent station observations as the reference, providing a direct assessment against ground-based measurements.

Second, to place our model's performance in the context of other widely-used climate datasets, a series of Taylor diagrams were produced using multiple gridded data products as references. This multi-reference evaluation provides a robust assessment of how well our reconstruction reproduces key statistical properties of the climate variables relative to different established datasets. The specific comparisons were as follows: temp was compared against 20CRv3, ERA5, and CRU_TS_4.08; shum was compared against 20CRv3 and ERA5; pres was compared against 20CRv3 and ERA5.

3.5 Validation using tree-ring proxy data

To evaluate the temporal fidelity of the TPHH product over the pre-instrumental era, we conducted a comprehensive validation against tree-ring width chronologies. Recognizing that tree radial growth on the TP is co-limited by the thermal and moisture conditions and characterized by biological memory, we established a physically constrained ternary regression model. Prior to modeling, the TPHH summer temperature (T_{JJA}), the early-summer specific humidity (Q_{MJJ}), and the tree-ring width index (TRWI) (Fang et al., 2022), were standardized using Z-score normalization to eliminate dimensional discrepancies. To isolate the independent climatic forcing from the physiological lag effect, the current-year thermal and moisture variables were combined with a first-order autoregressive term (the lag -1 $TRWI_{t-1}$). The normalized ternary model is expressed as:

$$TRWI_t = \beta_0 + \beta_1 \times T_{JJA} + \beta_2 \times Q_{MJJ} + \beta_3 \times TRWI_{t-1} \quad (5)$$

This standardized modeling effectively estimates the interannual theoretical tree growth driven by the TPHH historical climate. To further assess the decadal-to-multidecadal reliability of the dataset, an 11-year moving average was applied to both the model-fitted series and the actual tree-ring proxy records, filtering out high-frequency biological noise.

https://doi.org/10.5194/essd-2026-180
Preprint. Discussion started: 17 April 2026

Open Access Earth System Science Data Discussions

4 Results

4.1 Model performance cross-validation

To assess the ability of the hybrid-structure deep-learning model (FourCastNet) to reconstruct high-resolution climate fields from CRU predictors, we performed a comprehensive validation of its output using the TPMFD dataset during 1979–2023 as the ground truth.

4.1.1 Overall Quantitative Metrics

We first evaluated the model's overall performance during k-fold cross-validation. In the 9-fold cross-validation process, the model was independently evaluated on the test set of each fold. During 1979–2023, over a period of 45 years, there are a total of 540 months serving as time steps. We collected the metric values reported in the logs for all k-folds and calculated the arithmetic mean for four core metrics (R^2, RMSE, MAE, and IOA) across the three variables: temperature (temp), specific humidity (shum), and surface pressure (pres).

Table 3. Mean performance metrics from k-fold cross-validation (1979–2023, Reference: TPMFD).

Variable	R^2	RMSE	MAE	IOA
Temp (°C)	0.9762	0.9338	0.7001	0.9938
Shum (g/kg)	0.9741	0.2807	0.2152	0.9930
Pres (hPa)	0.9996	1.4422	1.0763	0.9999

The cross-validation results in Table 1 demonstrate that the model achieves similar simulation accuracy for all three variables. The mean IOA values for all variables exceed 0.99, and the mean R^2 for all variables exceed 0.97, with surface pressure reaching nearly 1.00 (0.9996), indicating that the model explains the vast majority of the variance in the high-resolution target fields. The RMSE for Temperature is 0.9338 °C, for specific humidity is 0.28 g/kg, and for surface pressure is 1.44 hPa. Thus, further confirms the model's capability to stably reconstruct fine-scale climate fields from low-resolution predictors.

4.1.2 Consistency analysis of reconstructed vs. aim values

To more intuitively assess the consistency between the model's predictions and the TPMFD reference data, we generated density scatter plots (2D histograms). This analysis aggregates all data points from the test sets of the 9-fold cross-validation, with 540 months serve as time steps in totally 45 years.

Firstly, for each variable, we calculated the "overall validation plots" (**Figure 4**), which include all data points of all months. It can be seen that the data for temp, shum and pres are tightly clustered around the 1:1 diagonal line. The high density of points in the core of the distribution and the lack of dramatic outliers imply little or no systematic bias in the reconstruction of the model.

https://doi.org/10.5194/essd-2026-180
Preprint. Discussion started: 17 April 2026

Open Access Earth System Science Data Discussions

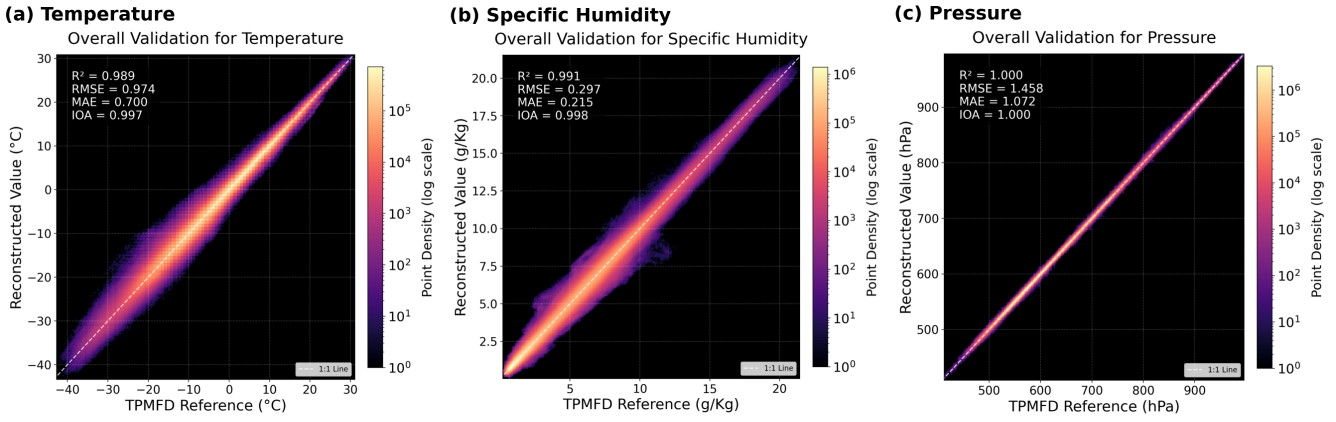


Figure 4. Overall density scatter plots for (a) temp, (b) shum, and (c) pres based on all 540 test time steps from the k-fold cross-validation (1979–2023). Colors indicate point density on a logarithmic scale.

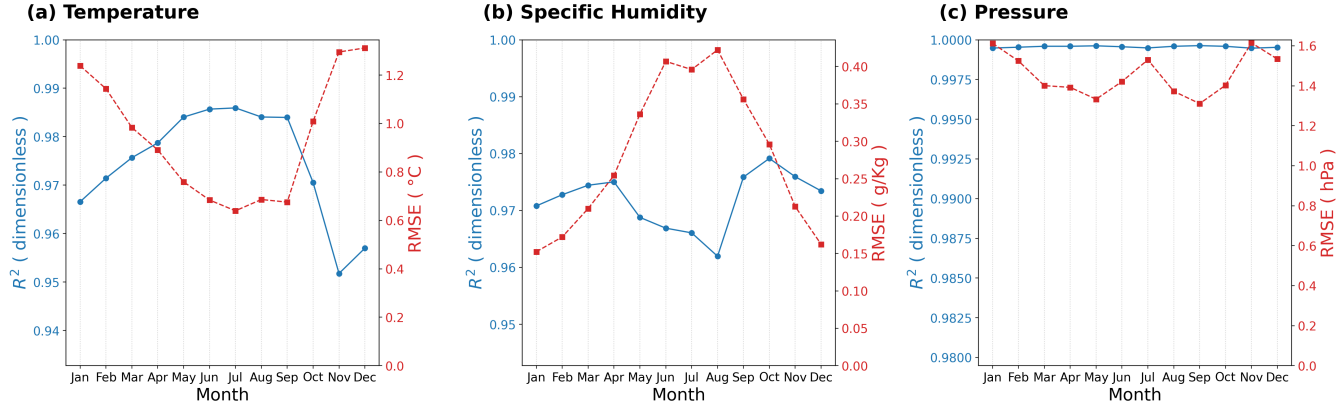


Figure 5. Seasonal cycle of model performance (1979–2023). Each panel shows the selected metrics (R^2 and RMSE) calculated for each month of the year, based on the k-fold validation results for (a) temp, (b) shum, and (c) pres.

Next, to investigate potential seasonal variations in model fidelity, we analyzed the performance metrics for each of the 12 months. A full set of 36 monthly density scatter plots is provided in the Supplementary Materials (**Figures S1–S3**). To summarize the seasonal cycle of model performance, we plot the monthly-stratified metrics in **Figure 5**.

This examination makes apparent the regular seasonality. For temp (a), the reconstruction error (RMSE) is slightly larger in winter than in summer, possibly because of complicated surface energy feedbacks over snow-covered surface in winter months. In contrast, the performance exhibits larger errors for the precipitation-related shum (b) during the major monsoon months (June, July, August), indicating that high-frequency variations of specific humidity associated with convective activities are more difficult to be captured. Yet, the R^2 stays above 0.95 for each month through the year, indicating the soundness of the model for all seasons.

https://doi.org/10.5194/essd-2026-180
Preprint. Discussion started: 17 April 2026

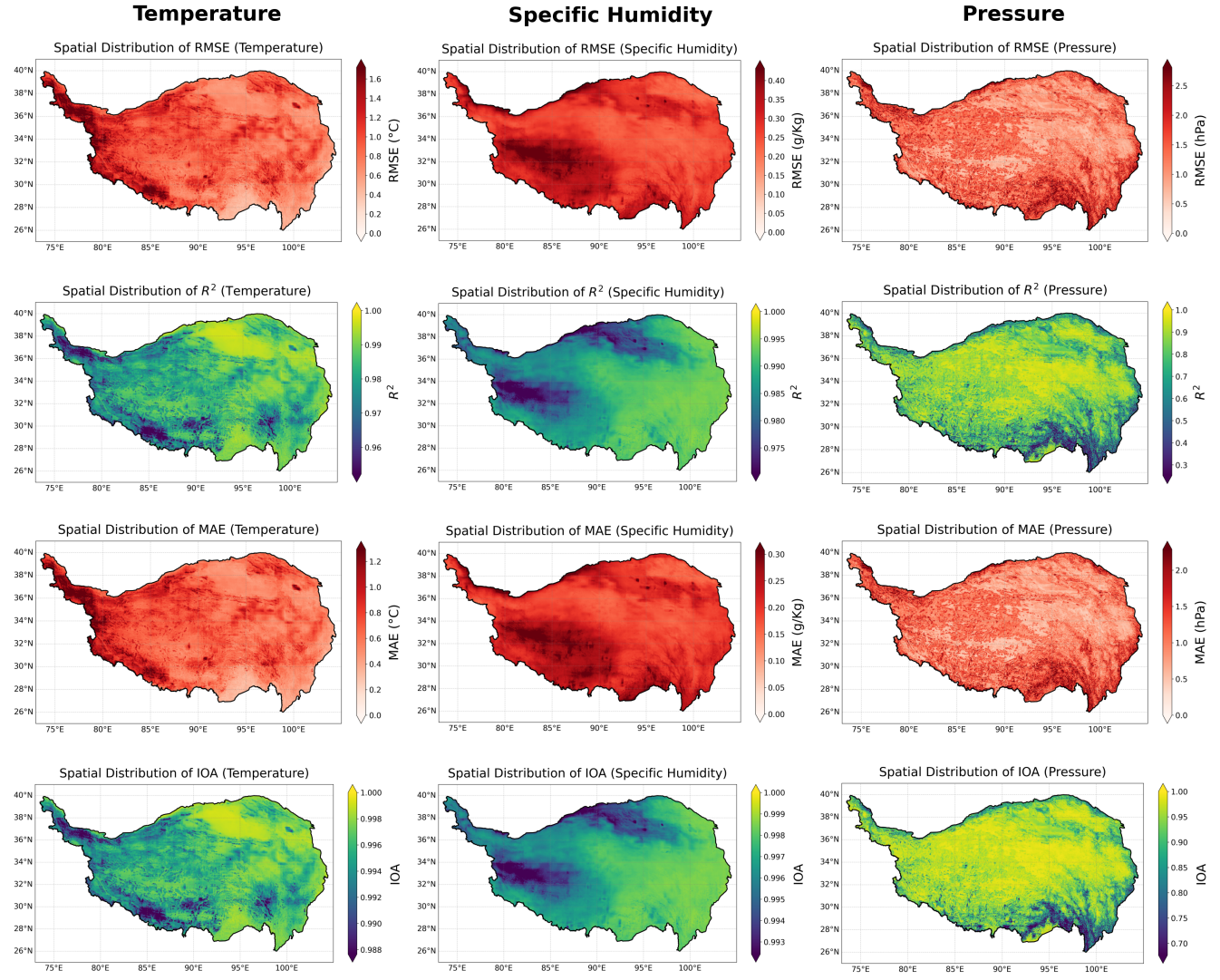


Figure 6. Spatial patterns of metrics (1979–2023). The matrix contains the performance metrics in the rows and the variables in the columns: Rows: (1) RMSE, (2) R^2, (3) MAE, and (4) IOA. Columns: (Left) surface temperature, (Middle) specific humidity, and (Right) surface pressure.

4.1.3 Spatial distribution of model performance

To investigate the spatial variations in model performance, we mapped the spatial distribution of performance metrics for the study region.. This analysis is also based on all 540 months serving as time steps from the 9-fold cross-validation. The results for all 12 metric-variable combinations are compiled in **Figure 6**.

https://doi.org/10.5194/essd-2026-180
Preprint. Discussion started: 17 April 2026

Figure 6 demonstrates spatial variations of model performance. In general, the model is more accurate over the central interior basins and the more level portions of the TP. This is shown by consistently high R^2 and IOA (second and fourth rows) and low error values (first and third rows).

Relatively larger errors (e.g., higher RMSE/MAE) concentrate mainly at the margins of the plateau, such as the southern slopes of the Himalayas and Hengduan Mountains. These steep topographic gradients, make it challenging to extract fine $1/30° \times 1/30°$ fields from $0.5° \times 0.5°$ predictors. Yet the texture is well represented even in these complicated landscapes by the model. As demonstrated in the third column of **Figure 6**, the metrics of pres are spatially uniform, which is to be expected as surface pressure is primarily dictated by elevation, a static feature that the model has learned very well.

4.2 Evaluation of the reconstruction

Having demonstrated the model's calibration performance, we now turn to performance validation of reconstruction. We employed two benchmark validation strategies: 1) direct comparison with independent site-observations, 2) statistical comparison with other related reanalysis/model products.

4.2.1 Validation against site-observations

We used the 135 independent in-situ meteorological stations on the TP as the reference to benchmark our reconstruction (TPHH) against other comparable products (e.g., ERA5, CRU, 20CRv3). These in situ observations are derived from 125 stations prior to 1979, and 135 stations thereafter. For the validation during 1979–2023, the TPHH product was evaluated against independent hold-out samples generated via a 10-fold cross-validation process, which recombined the entire dataset. We interpolated the values from gridded products corresponding to the site locations and calculated the standard deviation (STD) and correlation coefficient of their time series against the site-observation values.

The Taylor diagrams (**Figure 7**) provide a concise summary of the statistical consistency of each dataset with the in-situ site-observations. In the comparisons for temp, shum, and pres, our reconstruction (TPHH) outperformed most comparable products specifically:

1) For Surface air temperature (**Figure 7a,d**), most products have standard deviations between 8–10 (°C), close to the reference value of ~ 9 (°C). The main strength of TPHH is its high correlation coefficient. All products exhibited nearly unchanged relative diagram positions in the two periods before and after 1979.

2) In the case of specific humidity (**Figure 7b,e**), the ERA, JRA, 20CR products all have substantially smaller standard deviations than site-observations, while the other products have closer standard deviation but lower correlation coefficients. The TPHH ranked first in both of these two metrics and held a leading advantage of its position.

3) For surface pressure (**Figure 7c,f**), in all products, the metric with the largest differences is the correlation coefficient. The TPHH ranked first again and expanded its leading advantage.

Therefore, we conclude that the accuracy of our reconstruction, at least during the two validation periods, ranks highly among comparable products in capturing point-scale variability, which substantially supports its credibility.

https://doi.org/10.5194/essd-2026-180
Preprint. Discussion started: 17 April 2026

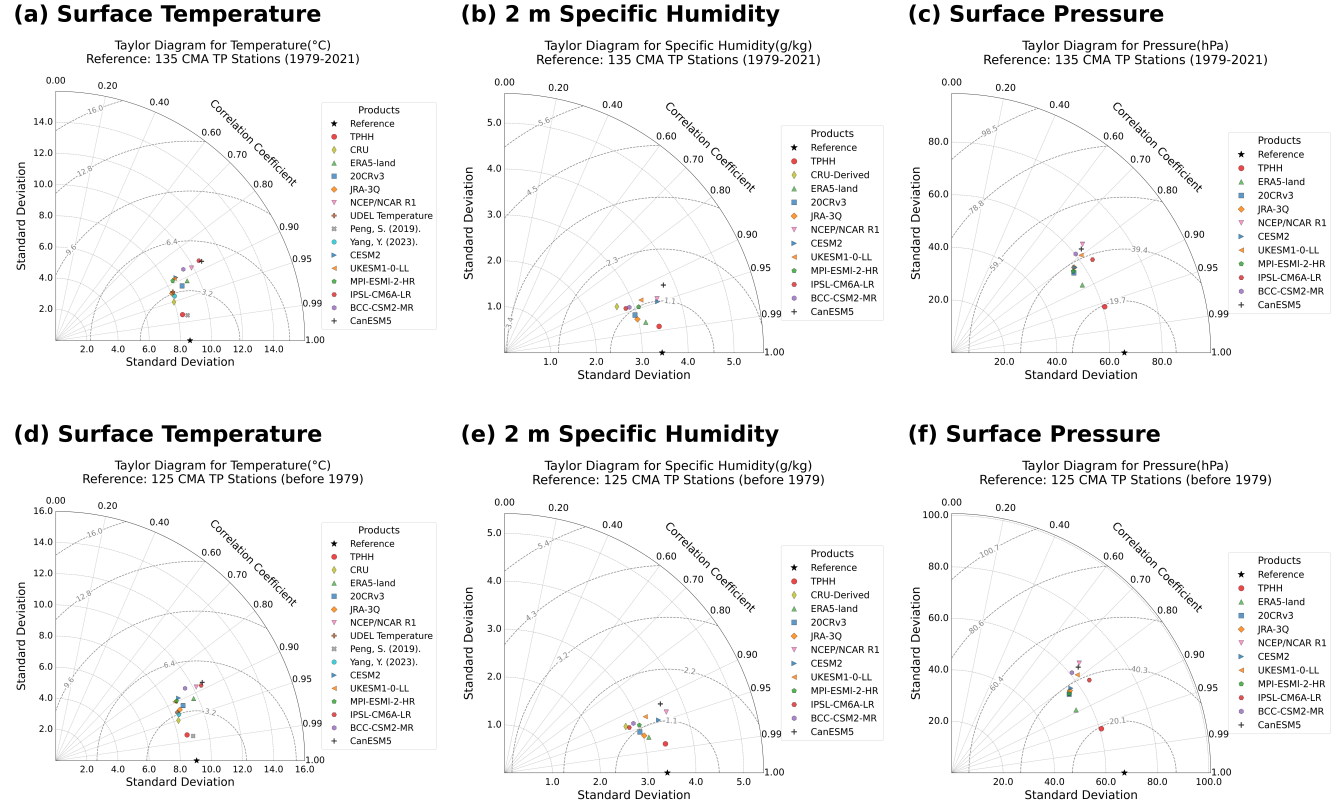


Figure 7. Taylor diagrams comparing our reconstruction (TPHH, red circle) and other climate products against independent CMA site observations. The top row (a–c) evaluates the calibration period (1979–2023) based on 135 stations, and the bottom row (d–f) evaluates the historical validation period (1901–1978) based on 125 stations for surface temperature, 2 m specific humidity, and surface pressure, respectively. The reference point (black star) represents in-situ site observations.

4.2.2 Inter-comparison with comparable products

To further assess the statistical plausibility (e.g., spatial variability, correlation) of the reconstructed data, we conducted a broader comparison of TPHH against more than ten related products. We performed seven Taylor diagram analyses, each using a different dataset (e.g., ERA5-Land, JRA-3Q) as the benchmarking reference.

As shown in **Figure 8**, the statistical performance of TPHH varies depending on the reference benchmark, reflecting the distinct characteristics of the datasets being compared.

1) High fidelity to modern reanalysis (vs. ERA5). When ERA5 (0.1°) is used as the reference (**Figure 8c, e, g**), TPHH demonstrates remarkably high statistical similarity, clustering closely with other top-tier datasets. Specifically, only CRU and TPHH have correlation coefficients (corr.) with ERA5 exceeding 0.95 in temperature. Besides, in case of specific humidity validation, only JRA-3Q and TPHH exceeded 0.95 with prominent advantages against other products. This confirms that,

https://doi.org/10.5194/essd-2026-180
Preprint. Discussion started: 17 April 2026

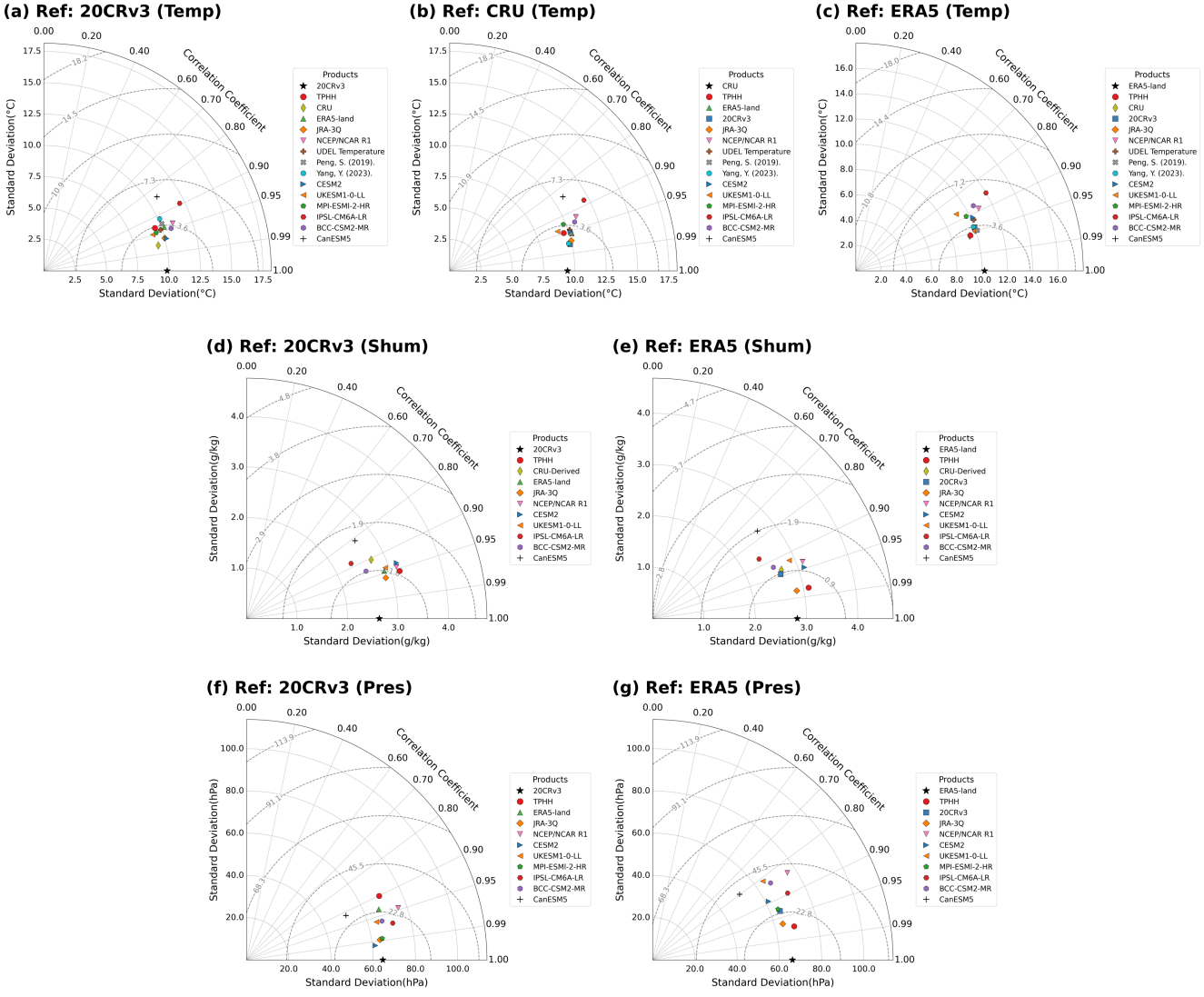


Figure 8. Taylor diagrams for multi-product inter-comparison (before 1979). (a-c) Top row: temperature (temp, units: °C) comparison against (a) 20CRv3, (b) CRU, and (c) ERA5. (d-e) Middle row: specific humidity (shum, units: g/kg) comparison against (d) 20CRv3 and (e) ERA5. (f-g) Bottom row: surface pressure (pres, units: hPa) comparison against (f) 20CRv3 and (g) ERA5. The TPHH dataset is marked by a red circle.

for the overlapping period where high-quality observational constraints are available, our model efficiently reproduces the synoptic-scale climatic patterns.

2) Reasonable consistency with historical baselines (vs. 20CRv3): In comparisons against the coarser 20CRv3 (1.0°), TPHH does not always occupy the position closest to the reference point (**Figure 8a, d, f**). However, TPHH becomes one of the only

https://doi.org/10.5194/essd-2026-180
Preprint. Discussion started: 17 April 2026

two humidity products (another is JRA-3Q) with corr. > 0.95 again. As for the temperature, TPHH's performance level is also around the median among multiple products.

3) Fidelity to Input forcing (vs. CRU): Finally, the comparison against the predictor dataset CRU (**Figure 8b**) shows an exceptionally high correlation coefficient. This is a critical validation, confirming that the deep learning model preserves the reliable long-term temporal trends inherent in the observational inputs, without introducing significant artifacts during the downscaling process.

4.3 Analysis of spatial-temporal consistency

Having thoroughly validated the reliability of the TPHH dataset, this section utilizes this new dataset to analyze the long-term climatic characteristics of the Tibetan Plateau since 1901.

4.3.1 Spatial patterns for long-term historical climatology (1901–1978)

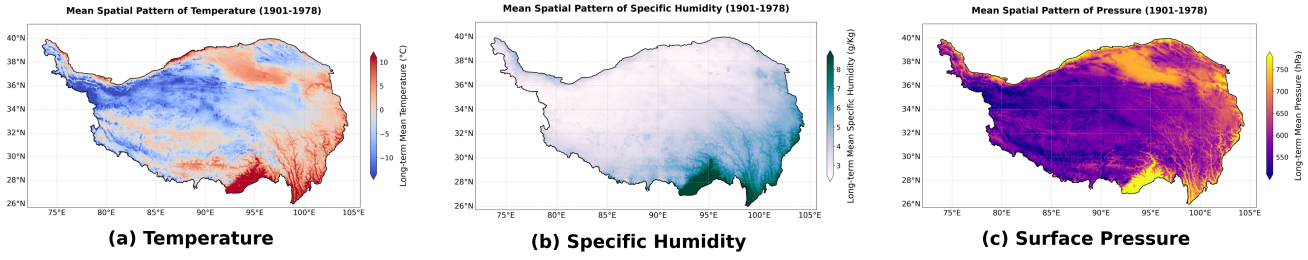


Figure 9. Long-term mean spatial distribution (1901–1978) reconstructed by TPHH. Panels show (a) surface temperature, (b) specific humidity, and (c) surface pressure. All data are at $1/30^\circ \times 1/30^\circ$ resolution, masked by the TP boundary.

To present the reconstructed historical climate baseline, we show the temporal mean over the 78-year period from 1901 to 1978, generating long-term mean spatial distribution maps for temp, shum, and pres.

Figure 9 presents the average climatic conditions over the Tibetan Plateau for most of the 20th century.

1) Surface air temperature (**Figure 9a**): The distribution is predominantly governed by elevation and latitude, as seen in the illustration, with the temperature decreasing toward the high-altitude northwestern borders and lower Qaidam Basin and southeastern valley experiencing warmer climate.

2) Specific humidity (**Figure 9b**): The pattern exhibits an obvious decreasing trend from southeast to northwest. This two-dimensional spatial heterogeneity well describes the water vapor transport streams within the South Asian Monsoon system, which penetrates into the plateau through the "Grand Canyon" of Yarlung Zangbo River.

3) Surface pressure (**Figure 9c**): The surface pressure field exhibits fine structures consistent with high-resolution terrain, and the valley-mountain differences resolved here are usually overlapped/averaged in coarse resolution fields.

https://doi.org/10.5194/essd-2026-180
Preprint. Discussion started: 17 April 2026

Open Access Earth System Science Data Discussions

These climatology maps at $1/30° \times 1/30°$ resolution from reanalysis and satellite data offer a prescribed refined background field for TP's ecological, hydrological, and glaciological research.

4.3.2 Analysis of Long-term time series continuity and trends

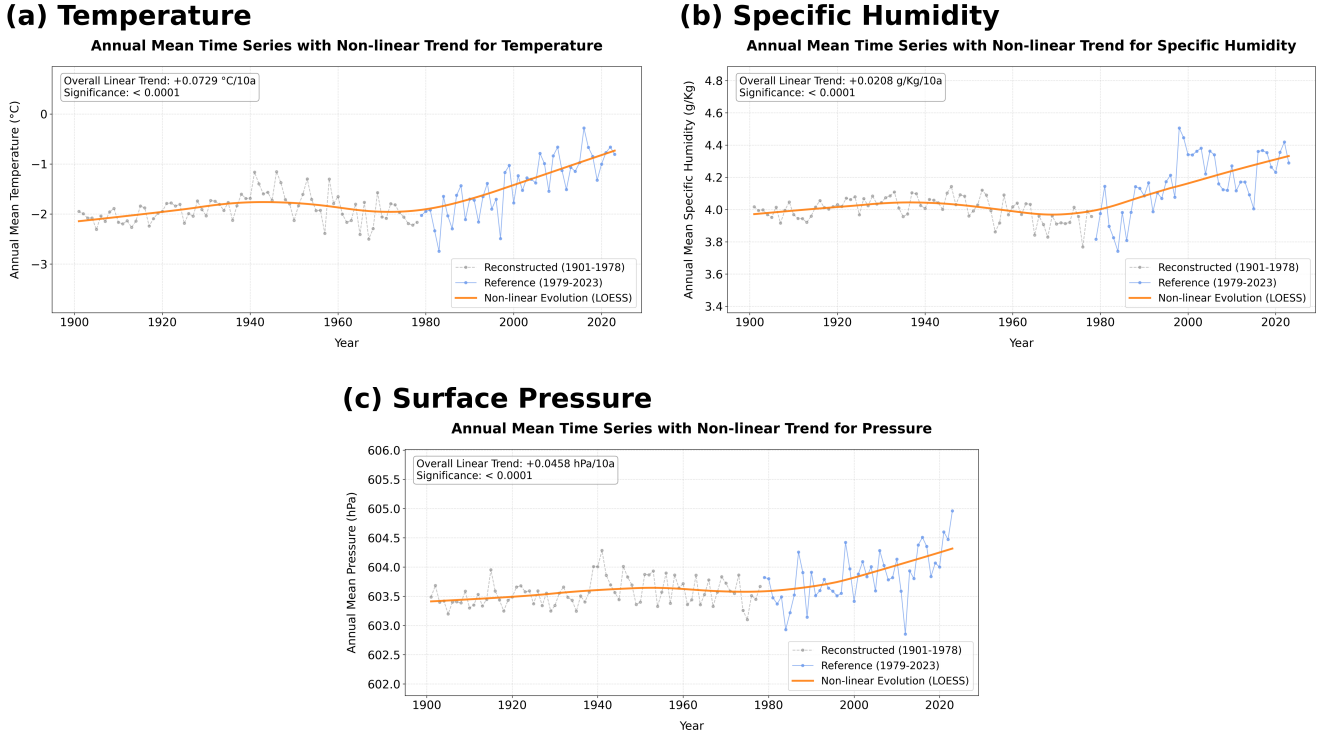


Figure 10. Annual Mean Time Series with LOESS trend for (a) temp (b) shum (c) pres, derived from spatial averages over the TP.

A key quality check is to assess the continuity at the junction between the historical reconstruction (1901–1978) and the modern reference data (1979–2023). We first extracted the time series at a specific coordinate point (Longitude 91.1°, Latitude 31.5°) (**Figure 10**) and then calculated the spatial mean time series over the entire study area (**Figure 11**).

As can be seen in **Figure 10** and **11**, the reconstructed and TPMFD reference data are well matched at the 1979 boundary, without apparent "jumps" or artificial discontinuities. This smooth transition demonstrates that our model did not overfit on the calibration period (1979–2023) and that the reconstructed historical climate is rational.

The long-term trend derived via LOESS in **Figure 11** prominently displays the climate variability during the last 123 years. The temp variable also exhibits a notable warming trend throughout the 1901–2023 period. The shum variable also exhibits a trend towards wetting, with that trend being stronger in recent decades. To delve deeper into this long-term trend seasonality, the linear trend spatial pattern was assessed on a month-by-month basis. Because of the size of the visual data (36 maps), detailed monthly trend maps are given in the Supplementary Materials (**Figures S4–S6**). Analysis of these supplemental Figures shows

https://doi.org/10.5194/essd-2026-180
Preprint. Discussion started: 17 April 2026

Open Access Earth System Science Data Discussions

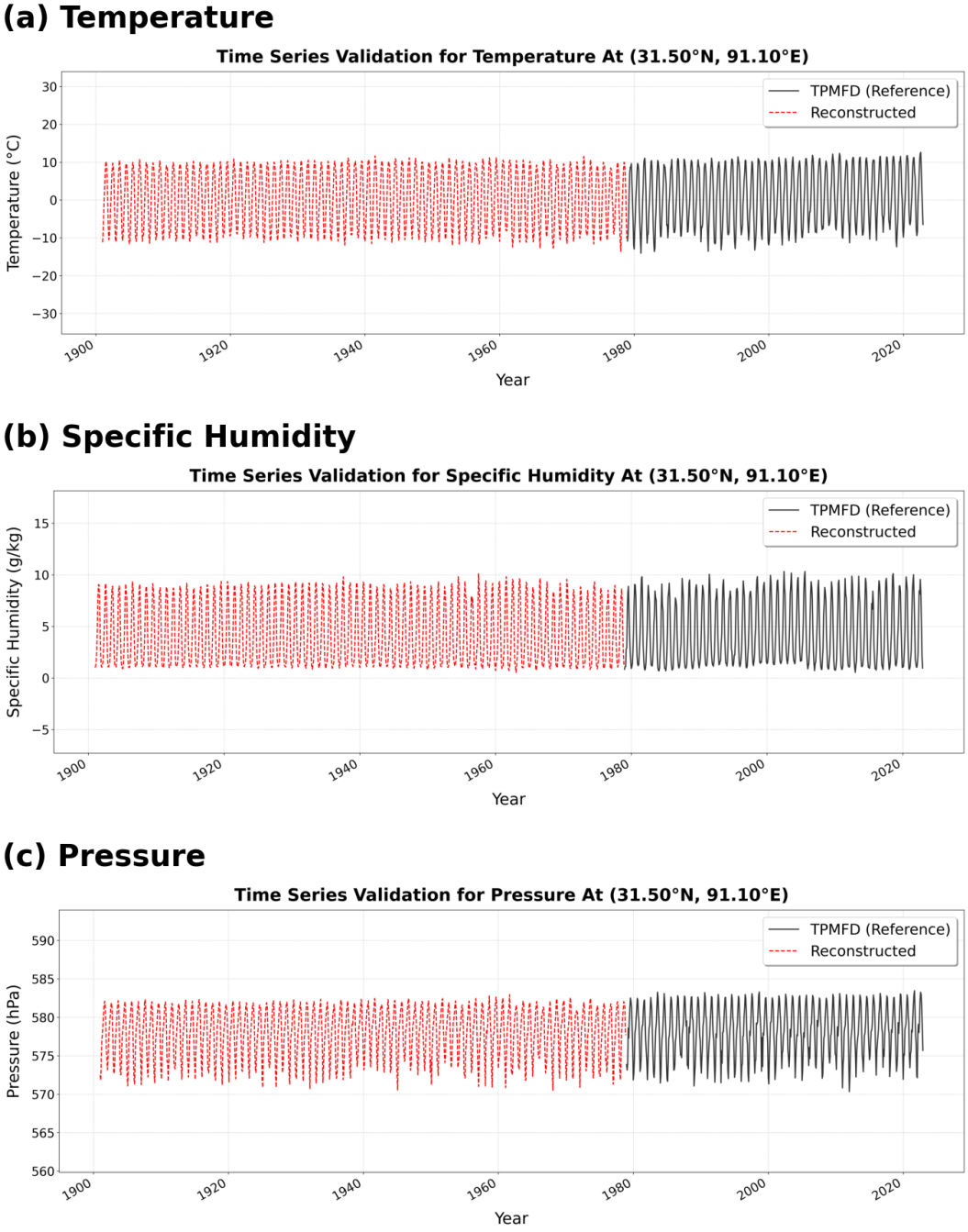


Figure 11. Single-point time series validation (Lon: 91.1°, Lat: 31.5°) for (a) `temp`, (b) `shum`, and (c) `pres`. The plot shows the reconstructed values (1901–1978) joined with the TPMFD reference values (1979–2023).

https://doi.org/10.5194/essd-2026-180
Preprint. Discussion started: 17 April 2026

that the warming trend is strongest in winter and spring, likely associated with snow-albedo feedbacks, while increases in specific humidity are greatest in the summer monsoon.

4.4 Efficiency of the reconstruction in tree-ring proxy signal fitting

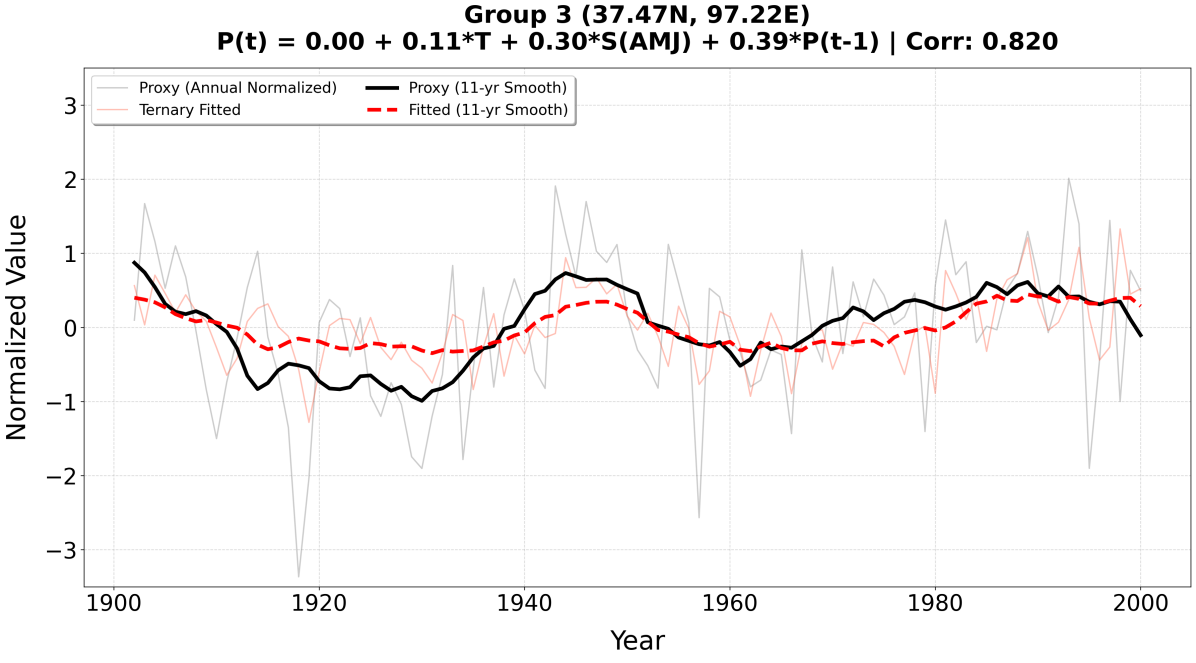


Figure 12. Temporal consistency between the actual tree-ring proxy and the TPHH-based ternary fitted series (1901–2000). The fitted series is generated via a standardized ternary regression model integrating TPHH JJA temperature (T), AMJ specific humidity (S), and lag -1 biological memory of the tree-ring proxy (P). Bold lines denote the 11-year moving average, highlighting the low-frequency correlation. See Supplementary **Figures S9–S13** for the complete validation across all 41 proxy groups.

The fitting result demonstrates a high temporal consistency with the annual tree-ring proxy series. As illustrated in **Figure 12** for a representative chronology, the smoothed climate-driven fitted series aligns tightly with the actual proxy record, yielding a Pearson correlation coefficient exceeding 0.8. Specifically, the lag -1 autoregressive term exhibits the largest standardized coefficient ($\beta = 0.39$), highlighting the dominant role of physiological inertia. This indicates that radial growth on the TP is heavily subsidized by non-structural carbohydrates stored in the preceding year. Early-summer specific humidity emerges as the primary limiting factor ($\beta = 0.30$), with a contribution nearly three times greater than that of peak growing-season temperature ($\beta = 0.11$). This long-term coherence confirms that the TPHH dataset captures a part of the annual climate fluctuations. The spatial distribution and fitting results for totally 41 groups of proxy locations are detailed in the Supplementary Materials **Figure S8-13**.

https://doi.org/10.5194/essd-2026-180
Preprint. Discussion started: 17 April 2026

5 Conclusions

Through a series of qualitative evaluations, our study has demonstrated that the TPHH dataset possesses high accuracy during the modern validation period (1979–2023) and superior fidelity to in-situ observations compared to related products during the historical period (1901–1978). Analysis based on this dataset has revealed the fine-scale, long-term mean climatology of the Tibetan Plateau and has shown significant and continuous climate change trends since 1901.

The TPHH fill the essential gap of high-resolution climate information concerning near surface atmospheric specific humidity over the TP in the pre-satellite era. By combining the historical coherence of the CRU data with the high resolution of the TPMFD data in a FourCastNet-based deep learning framework, we can produce the dataset TPHH ($1/30° \times 1/30°$, monthly, 1901–2023), and can be updated as the temporal range of CRU product extends. Key validation results are summarized as follows:

1) Model performance: The FourCastNet model achieved superior performance in the task of super-resolution for coarse climate data. Cross-validation (1979–) produced mean R^2 of 0.97 both for humidity and temperature, with RMSE shown in humidity as 0.28 g/kg and in temperature as 0.93 °C.

2) Historical soundness: Our validation with 135 independent meteorological sites demonstrated that the reconstruct surpasses other related products (ERA5-Land, JRA-3Q, 20CRv3, etc.) when capturing local climate details. The Taylor diagram indicates that both humidity and pressure from TPHH has the best performance in terms of the highest correlation and lowest RMSE when compared with in-situ observation.

3) Climate trends: The reconstructed series indicates a persistent long-term warming pattern over the TP since 1901. In addition, the spatial trend analysis reveals a warming pattern dependent on elevation (strongest in winter/spring) and a wetting trend mainly under the control of summer monsoon intensification.

The TPHH dataset stands for a high-scientific standard, high-resolution climate baseline of the early twentieth century for the “Asian Water Tower”. This can provide alternative materials to study long-term hydrological cascades, glacier dynamics, and ecosystem responses to climate change. Further studies can generalize this approach to finer temporal scales (e.g., daily) by taking into account extreme weather events.

6 Discussion

6.1 Physical basis of TPHH’s downscaling approach

The superior performance of the TPHH dataset is fundamentally rooted in the physical premise of "terrain governance" over the TP’s climate. The complex terrain of the TP acts as the primary modulator of high-frequency local climate variability, which provides potential information for the model to extract. Unlike traditional downscaling methods that often rely on static interpolation or linear regression, which merely redistribute coarse-resolution signals based on elevation lapse rates, the TPHH framework employs a hybrid-structure deep-learning model that can synergize total-field and sub-regional information mapping to the predictor. By leveraging the FourCastNet architecture (hybrid structure of AFNO and ViT), our model goes beyond point-

https://doi.org/10.5194/essd-2026-180
Preprint. Discussion started: 17 April 2026

to-point estimation. It learns the non-linear, synergistic relationship between the synoptic-scale background and the fine-scale terrain-driven features. This allows TPHH to infer physically coherent sub-regional fields where the high-resolution terrain footprints are implicitly captured during the learning of temperature and humidity gradients, ensuring that the generated fields are not only statistically accurate but also dynamically consistent with the underlying landscape.

6.2 Comparison with two other 1901–present reconstructions

Although the primary objective of TPHH is to reconstruct near-surface atmospheric humidity fields via machine-learning downscaling, this product also demonstrates advantages in air temperature reconstruction performance. Traditional reconstruction methods, such as the delta downscaling interpolation employed by Peng et al. (2019), rely heavily on the assumption of spatially continuous lapse rates. While effective in homogenous terrain, these methods treat the "terrain-climate" relationship as a static, local linear function of elevation. They fundamentally lack the capacity to reflect various terrain-atmosphere feedback conditions. In the complex hinterlands of the TP, where terrain governance induces various features, this method struggles to capture the high-frequency variability, leading to the "smoothing effect" observed in their results. Recent advancements by Yang et al. (2023) introduced a deep-learning model based on generative adversarial network (GAN) to reconstruction. However, standard convolutional neural networks (CNNs) used in such studies typically operate with limited "receptive fields". They excel at extracting local features but often fail to capture long-range dependencies, that is the critical "teleconnections" between the large-scale synoptic background and sub-regional micro-climates. As a result, while the GAN-based product improves local texture, it may lose physical consistency across the broader "total-field" when the input station data is sparse. Shown as spatial patterns for temperature RMSE compared to TPMFD, the delta-downscaling-based product, GAN-based product, and TPHH all exhibit increased bias in regions with high elevation gradients and complex mountainous terrain (**Figure S7**). Due to its insufficient spatial resolution (~ 10 km $\times$ 10 km), the GAN-based product even shows widespread noise in the extremely rugged southeastern TP, failing to follow the actual terrain patterns. Notably, the bias amplification of the delta-downscaling-based product in high-relief areas is entirely and substantially greater than that of TPHH. In the high-elevation central-southern TP (around 29°N, 93°E), TPHH even effectively mitigated the influence of mountainous terrain and did not display elevation-dependent bias, providing a stark contrast to the delta-downscaling-based product.

6.3 Uncertainty and prospects

Despite the advancements in downscaling reconstruction, TPHH has certain limitations. First, the reconstruction quality is inherently bounded by the temporal consistency of the predictor dataset. While CRU provides a reliable long-term baseline, any systemic biases in the early 20th-century observational records could propagate into the TPHH product. Second, while the $1/30° \times 1/30°$ resolution is a major advancement, it may still be insufficient to resolve micro-climate features in extremely narrow valleys or glaciated peaks smaller than the grid scale. Future iterations of this work will aim to incorporate multi-source predictors (e.g., paleo-climate proxies) to further constrain uncertainties in the pre-1950 observational void.

https://doi.org/10.5194/essd-2026-180
Preprint. Discussion started: 17 April 2026

Data availability

The TPHH dataset provides monthly 2 m specific humidity, 2 m air temperature, and surface air pressure data with a resolution of $1/30° \times 1/30°$, covering the Tibetan Plateau (25–40° N, 75–105° E). This product spanned over the period during 1901–2023 and will be continuously updated as the TPMFD or CRU product is updated. All variables are provided in NetCDF format with open access. The dataset can be accessed at https://doi.org/10.57760/sciencedb.36169 and is formally cited in this paper as Chen (2026).

Financial support. This work was supported by the National Natural Science Foundation of China (grant no. 42505115), the Sichuan Science and Technology Program (grant no. 2025ZNSFSC1135), the Shanghai Key Laboratory of Ocean-land-atmosphere Boundary Dynamics and Climate Change (grant no. FDAOS-OP202411), and the Everest Initiative Interdisciplinary Team Project of Chengdu University of Technology (grant no. 2024ZF11422).

Author contributions. ZJ contributed to designing the research. ZC and JZ implemented data curation and visualization. ZJ and ZC implemented the research and wrote the original draft. QY, ZL, SL, and SK supervised the research. All co-authors revised the manuscript and contributed to the writing.

Competing interests. The contact author has declared that none of the authors has any competing interests.

https://doi.org/10.5194/essd-2026-180
Preprint. Discussion started: 17 April 2026

References

Cao, L., Zhu, Y., Tang, G., Yuan, F., and Yan, Z.: Climatic warming in China according to a homogenized data set from 2419 stations, Int. J. Climatol, 36, 4384–4392, https://doi.org/10.1002/joc.46391, 2016.

Chen, Y., Duan, X., Ding, M., Qi, W., Wei, T., Li, J., and Xie, Y.: New gridded dataset of rainfall erosivity (1950–2020) on the Tibetan Plateau, Earth System Science Data, 14, 2681–2695, https://doi.org/10.5194/essd-14-2681-2022, eSSD, 2022.

Chen, Z.: TPHH: A long-term (1901–1978) high-resolution (1/30°) reconstruction of meteorological variables over the Tibetan Plateau (1.0) [dataset], https://doi.org/10.57760/sciencedb.36169, 2026.

Dong, N., Xu, X., Zhang, R., Sun, C., Cai, W., and Zhao, R.: Mechanism underlying the correlation between the warming-wetting of the Qinghai-Tibet Plateau and atmospheric energy changes in high-impact oceanic areas, npj Climate and Atmospheric Science, 7, 324, https://doi.org/10.1038/s41612-024-00849-1, 2024.

Dosovitskiy, A.: An image is worth 16×16 words: Transformers for image recognition at scale, arXiv preprint arXiv:2010.11929, 2020.

Fang, M., Li, X., Chen, H. W., and Chen, D.: Arctic amplification modulated by Atlantic Multidecadal Oscillation and greenhouse forcing on multidecadal to century scales, Nature Communications, 13, 1865, https://doi.org/10.1038/s41467-022-29523-x, 2022.

Guibas, J., Mardani, M., Li, Z., Tao, A., Anandkumar, A., and Catanzaro, B.: Adaptive Fourier Neural Operators: Efficient Token Mixers for Transformers, arXiv preprint arXiv:2111.13587, 2021.

Harada, Y., Kobayashi, S., Kosaka, Y., Chiba, J., and Tanaka, T. Y.: Quality evaluation of the precipitation over the tropical oceans in the JRA-3Q reanalysis, Quarterly Journal of the Royal Meteorological Society, 152, e70 026, https://doi.org/10.1002/qj.70026, 2025.

Harris, I., Osborn, T. J., Jones, P., and Lister, D.: Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset, Scientific Data, 7, 109, https://doi.org/10.1038/s41597-020-0453-3, 2020.

Hu, J., Miao, C., Su, J., Zhang, Q., Gou, J., and Sun, Q.: An upgraded high-precision gridded precipitation dataset for the Chinese mainland considering spatial autocorrelation and covariates, Earth System Science Data, 17, 3987–4004, https://doi.org/10.5194/essd-17-3987-2025, eSSD, 2025.

Jiang, Y., Tang, W., Yang, K., He, J., Shao, C., Zhou, X., Lu, H., Chen, Y., Li, X., and Shi, J.: Development of a high-resolution near-surface meteorological forcing dataset for the Third Pole region, Science China Earth Sciences, 68, 1274–1290, https://doi.org/10.1007/s11430-024-1507-6, 2025.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., et al.: The NCEP/NCAR 40-year reanalysis project, Bulletin of the American Meteorological Society, 77, 437–471, 1996.

Kang, S., Xu, Y., You, Q., Flügel, W.-A., Pepin, N., and Yao, T.: Review of climate and cryospheric change in the Tibetan Plateau, Environmental Research Letters, 5, 015 101, https://doi.org/10.1088/1748-9326/5/1/015101, 2010.

Li, Z., Huang, D. Z., Liu, B., and Anandkumar, A.: Fourier Neural Operator with Learned Deformations for PDEs on General Geometries, Journal of Machine Learning Research, 24, 1–26, 2023.

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles, D. G., Piles, M., Rodríguez-Fernández, N. J., Zsoter, E., Buontempo, C., and Thépaut, J.-N.: ERA5-Land: a state-of-the-art global reanalysis dataset for land applications, Earth System Science Data, 13, 4349–4383, https://doi.org/10.5194/essd-13-4349-2021, eSSD, 2021.

https://doi.org/10.5194/essd-2026-180
Preprint. Discussion started: 17 April 2026

O'Neill, B., Tebaldi, C., Van Vuuren, D., Eyring, V., Friedlingstein, P., Hurtt, G., Knutti, R., Kriegler, E., Lamarque, J., Lowe, J., Meehl, G., Moss, R., Riahi, K., and Sanderson, B.: The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6, https://ore.exeter.ac.uk/articles/online_resource/The_Scenario_Model_Intercomparison_Project_ScenarioMIP_for_CMIP6/29724578, 2025.

Pathak, J., Subramanian, S., Harrington, P., Raja, S., Chattopadhyay, A., Mardani, M., Kurth, T., Hall, D., Li, Z., and Azizzadenesheli, K.: FourCastNet: A Global Data-driven High-resolution Weather Model using Adaptive Fourier Neural Operators, arXiv preprint arXiv:2202.11214, 2022.

Peng, S., Ding, Y., Liu, W., and Li, Z.: 1 km monthly temperature and precipitation dataset for China from 1901 to 2017, Earth System Science Data, 11, 1931–1946, https://doi.org/10.5194/essd-11-1931-2019, eSSD, 2019.

Pepin, N., Bradley, R. S., Diaz, H. F., Baraer, M., Caceres, E. B., Forsythe, N., Fowler, H., Greenwood, G., Hashmi, M. Z., Liu, X. D., Miller, J. R., Ning, L., Ohmura, A., Palazzi, E., Rangwala, I., Schöner, W., Severskiy, I., Shahgedanova, M., Wang, M. B., Williamson, S. N., Yang, D. Q., and Mountain Research Initiative, E. D. W. Working Group: Elevation-dependent warming in mountain regions of the world, Nature Climate Change, 5, 424–430, https://doi.org/10.1038/nclimate2563, 2015.

Prein, A. F., Langhans, W., Fosser, G., Ferrone, A., Ban, N., Goergen, K., Keller, M., Tölle, M., Gutjahr, O., Feser, F., Brisson, E., Kollet, S., Schmidli, J., van Lipzig, N. P. M., and Leung, R.: A review on regional convection-permitting climate modeling: Demonstrations, prospects, and challenges, Reviews of Geophysics, 53, 323–361, https://doi.org/10.1002/2014RG000475, 2015.

Qin, J., He, M., Jiang, H., and Lu, N.: Reconstruction of 60-year (1961–2020) surface air temperature on the Tibetan Plateau by fusing MODIS and ERA5 temperatures, Science of The Total Environment, 853, 158 406, https://doi.org/10.1016/j.scitotenv.2022.158406, 2022.

Qin, J., He, M., Yang, W., Lu, N., Yao, L., Jiang, H., Wu, J., Yang, K., and Zhou, C.: Temporally extended satellite-derived surface air temperatures reveal a complete warming picture on the Tibetan Plateau, Remote Sensing of Environment, 285, 113 410, https://doi.org/10.1016/j.rse.2022.113410, 2023.

Ran, Y., Li, X., and Cheng, G.: Climate warming over the past half century has led to thermal degradation of permafrost on the Qinghai–Tibet Plateau, The Cryosphere, 12, 595–608, https://doi.org/10.5194/tc-12-595-2018, tC, 2018.

Slivinski, L. C., Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Giese, B. S., McColl, C., Allan, R., Yin, X., Vose, R., Titchner, H., Kennedy, J., Spencer, L. J., Ashcroft, L., Brönnimann, S., Brunet, M., Camuffo, D., Cornes, R., Cram, T. A., Crouthamel, R., Domínguez-Castro, F., Freeman, J. E., Gergis, J., Hawkins, E., Jones, P. D., Jourdain, S., Kaplan, A., Kubota, H., Blancq, F. L., Lee, T.-C., Lorrey, A., Luterbacher, J., Maugeri, M., Mock, C. J., Moore, G. W. K., Przybylak, R., Pudmenzky, C., Reason, C., Slonosky, V. C., Smith, C. A., Tinz, B., Trewin, B., Valente, M. A., Wang, X. L., Wilkinson, C., Wood, K., and Wyszyński, P.: Towards a more reliable historical reanalysis: Improvements for version 3 of the Twentieth Century Reanalysis system, Quarterly Journal of the Royal Meteorological Society, 145, 2876–2908, https://doi.org/10.1002/qj.3598, 2019.

Song, C., Huang, B., Richards, K., Ke, L., and Hien Phan, V.: Accelerated lake expansion on the Tibetan Plateau in the 2000s: Induced by glacial melting or other processes?, Water Resources Research, 50, 3170–3186, https://doi.org/10.1002/2013WR014724, 2014.

Vandal, T., Kodra, E., Ganguly, S., Michaelis, A., Nemani, R., and Ganguly, A. R.: DeepSD: Generating High Resolution Climate Change Projections through Single Image Super-Resolution, https://doi.org/10.1145/3097983.3098004, 2017.

Willmott, C. J.: Terrestrial air temperature and precipitation: Monthly and annual time series (1950-1996), https://climate.udel.edu/resources/, 2000.

Yang, Y., You, Q., Jin, Z., Zuo, Z., Zhang, Y., and Kang, S.: The reconstruction for the monthly surface air temperature over the Tibetan Plateau during 1901–2020 by deep learning, Atmospheric Research, 285, 106 635, https://doi.org/10.1016/j.atmosres.2023.106635, 2023.

https://doi.org/10.5194/essd-2026-180
Preprint. Discussion started: 17 April 2026

Yao, T., Bolch, T., Chen, D., Gao, J., Immerzeel, W., Piao, S., Su, F., Thompson, L., Wada, Y., Wang, L., Wang, T., Wu, G., Xu, B., Yang, W., Zhang, G., and Zhao, P.: The imbalance of the Asian water tower, Nature Reviews Earth & Environment, 3, 618–632, https://doi.org/10.1038/s43017-022-00299-4, 2022.

You, Q., Chen, D., Wu, F., Pepin, N., Cai, Z., Ahrens, B., Jiang, Z., Wu, Z., Kang, S., and AghaKouchak, A.: Elevation dependent warming over the Tibetan Plateau: Patterns, mechanisms and perspectives, Earth-Science Reviews, 210, 103 349, https://doi.org/10.1016/j.earscirev.2020.103349, 2020.

Yu, Y., You, Q., Zhang, Y., Jin, Z., Kang, S., and Zhai, P.: Integrated warm-wet trends over the Tibetan Plateau in recent decades, Journal of Hydrology, 639, 131 599, https://doi.org/10.1016/j.jhydrol.2024.131599, 2024.

Zhang, C., Qin, D.-H., and Zhai, P.-M.: Amplification of warming on the Tibetan Plateau, Advances in Climate Change Research, 14, 493–501, https://doi.org/10.1016/j.accre.2023.07.004, 2023.

Zhou, P., Tang, J., Ma, M., Ji, D., and Shi, J.: High resolution Tibetan Plateau regional reanalysis 1961-present, Scientific Data, 11, 444, https://doi.org/10.1038/s41597-024-03282-4, 2024.

Zhu, Y., Cao, L., Tang, G., and Zhou, Z.: Homogenization of surface relative humidity over China (In Chinese), Advances in Climate Change Research, 11, 379, https://doi.org/10.3969/j.issn.1673-1719.2015.06.001, 2015.