remote sensing

MDPI

Data Descriptor

A 1 km Monthly Multi-Timescale SPEI Dataset for Mainland China Reconstructed from Station Observations and Multi-Source Environmental Predictors

Haoxiang Peng [1,†], Wenxing Li [2,†], Kun Zhang [3], Chengrong Wu [4], Hui Liu [1,*], Yihang Chu [1] and Xidong Chen [5]

[1] School of Computer Science and Mathematics, Central South University of Forestry and Technology, Changsha 410004, China; 20235223@csuft.edu.cn (H.P.); yhchu@xjtufh.edu.cn (Y.C.)
[2] College of Systems Engineering, National University of Defense Technology, Changsha 410073, China
[3] School of Electronics and Information, Northwestern Polytechnical University, Xi'an 710072, China; kunzhang@nwpu.edu.cn
[4] Department of Computer Science, University of Manchester, Manchester M13 9PL, UK; chengrong.wu@student.manchester.ac.uk
[5] Future Urbanity & Sustainable Environment (FUSE) Lab, Division of Landscape Architecture, Department of Architecture, Faculty of Architecture, The University of Hong Kong, Hong Kong SAR 999007, China; xdchenrs@hku.hk
* Correspondence: t20030970@csuft.edu.cn
† These authors contributed equally to this work.

Highlights

What are the main findings?

- A publicly available 1 km monthly multi-timescale SPEI dataset was produced for mainland China, covering 1979–2018 and including SPEI-1, SPEI-3, SPEI-6, SPEI-12, and SPEI-24.
- Technical validation yielded random five-fold R^2 values of 0.8807–0.9122, spatial-block R^2 values of 0.7857–0.8739, and forward blocked-period R^2 values of 0.7577–0.8337.

What are the implications of the main findings?

- The dataset expands the available high-resolution SPEI resources for mainland China by providing an independently reconstructed 1 km monthly product at five accumulation timescales, integrating station observations with CMFD meteorological forcing and NASADEM terrain.
- The open GeoTIFF archive supports drought monitoring and hydroclimatic, agricultural, ecological, and remote-sensing analyses.

Abstract

High-resolution multi-timescale SPEI data are important for drought monitoring and climate-impact studies, but long-term high-resolution datasets remain limited over large regions. This study developed and technically validated a station-supervised monthly multi-timescale SPEI dataset for mainland China from 1979 to 2018 on a 1 km grid, including SPEI-1, SPEI-3, SPEI-6, SPEI-12, and SPEI-24. Monthly labels were extracted from daily records at 427 stations using an exact source-calendar month-end rule. Monthly 0.1° CMFD predictors were mapped to stations and 1 km pixel centres using a NaN-aware bilinear sampler and combined with complete-window temporal features and NASADEM terrain covariates. Five timescale-specific LightGBM models generated 2359 nationwide monthly rasters. Random five-fold cross-validation yielded mean R^2 values of 0.8807–0.9122; 200 km spatial-block validation with a 100 km exclusion buffer yielded 0.7857–0.8739; and forward blocked-period validation yielded 0.7577–0.8337. Full-period product comparison gave

Check for updates

Academic Editor: Gabriel Senay

Received: 13 August 2026
Revised: 30 September 2026
Accepted: 20 October 2026
Published: 22 October 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

Pearson r = 0.9295 against Zhang-SPEI for SPEI-3, while comparison with HSPEI across five timescales yielded r = 0.7772–0.3267. Spatially buffered drought-condition validation produced F1 scores of 0.7021–0.7838 for moderate-or-worse drought, whereas extreme-drought Recall was 0.2397–0.3869. The dataset provides a long-term, multi-timescale resource for drought monitoring and climate-impact research across mainland China, with validation results characterizing its performance across regions, periods, and drought-severity levels.

Keywords: SPEI; drought monitoring; 1 km gridded dataset; mainland China; machine learning

1. Introduction

Drought is one of the most widespread and persistent natural hazards under global change, and its occurrence and evolution can exert profound impacts on regional water security, agricultural production, ecosystem stability, and socioeconomic activities [1–5]. Compared with abrupt disasters such as floods, drought is generally characterized by slow onset, long duration, broad spatial extent, and pronounced cumulative effects. Therefore, its monitoring and characterization at multiple timescales and fine spatial resolution are of great importance [6–9]. The Standardized Precipitation Evapotranspiration Index (SPEI), which simultaneously accounts for precipitation supply and atmospheric evaporative demand, enables the characterization of drought processes at different timescales within a unified framework and has become one of the most widely used indices in meteorological, agricultural, and hydrological drought studies [10].

In recent years, with the continuous accumulation of station observations, reanalysis products [11,12], land-surface meteorological forcing datasets [13–15], and multi-source environmental factors [16,17], substantial progress has been made in the development of SPEI and related drought datasets. The 5 km high-resolution SPEI dataset developed for Central Asia has demonstrated the potential of high-resolution drought products for regional studies and risk analysis [18]. For mainland China, previous studies have released a daily multi-timescale station-based SPEI dataset derived from 427 meteorological stations and covering 1961–2018 (Wang-SPEI), which provides a reliable basis for station-scale drought characterization [19]. In addition, researchers have further developed the Zhang-SPEI dataset, a daily SPEI raster product at 0.1° spatial resolution covering mainland China from 1979 to 2018 [20], as well as the 1 km resolution GPR-SPEI and HSPEI datasets for mainland China [21,22], indicating the continued advancement of high-resolution SPEI product development.

Existing datasets have substantially improved drought data availability over China, and complementary high-resolution SPEI products based on alternative supervision sources and reconstruction designs can further broaden the available data resources [23]. A dataset directly supervised by station-based SPEI observations can help maintain consistency with ground-based drought characterization. Combining meteorological forcing, timescale-dependent memory features, and consistently defined topographic information can represent regional hydroclimatic conditions and terrain-related spatial heterogeneity. This design yields an independently constructed and station-validated data source for cross-product comparison and spatially detailed drought analysis over mainland China, particularly in complex terrain, climatic transition zones, and areas with strong land-surface heterogeneity [24,25]. Moreover, SPEI has a pronounced multi-timescale nature, and different timescales correspond to different hydroclimatic accumulation and memory processes. Therefore, high-resolution SPEI dataset construction should consider not only contempora-

neous meteorological signals, but also cumulative background conditions and timescale-dependent temporal memory features [26].

The continued accumulation of multi-source meteorological and environmental data has provided a favorable basis for information-fusion reconstruction. For mainland China, the China Meteorological Forcing Dataset (CMFD) provides meteorological forcing data covering the entire country since 1979 at 0.1° spatial resolution, offering a consistent basis for representing regional-scale dynamic meteorological conditions [27]. Predictor-sensitivity tests also evaluated SoilGrids 2.0 covariates [28], whereas the production models and released rasters use the consistently defined NASADEM terrain stack without soil covariates.

Machine learning has gradually become an important technical pathway for constructing high-resolution drought datasets [29–33]. Compared with traditional empirical interpolation or single statistical downscaling methods, machine learning can more flexibly capture the nonlinear relationships among meteorological factors, land-surface environmental conditions, and drought indices, thus providing new technical support for high-resolution drought product reconstruction. For example, Fu et al. reconstructed a 1 km long-term SPEI product over Southwest China using machine learning, demonstrating that data-driven approaches based on multi-source environmental factors can effectively enhance the spatial detail of drought indices [34]. Similarly, AI-based downscaling studies have shown that, under limited observational conditions, intelligent methods that integrate satellite information and meteorological factors can improve the simulation of drought indices at higher spatial resolution [35,36].

These studies collectively indicate that machine learning provides an effective tool for high-resolution drought product construction, while its reconstruction performance still depends on the support of reliable supervision and the proper integration of dynamic meteorological background, temporal accumulation effects, and local environmental constraints [32]. Mainland China is characterized by vast territory, pronounced topographic relief, diverse climate types, and complex land-surface heterogeneity, which impose higher requirements on the spatial resolution and local representation capability of drought products [37]. In this context, this study uses exact source-calendar month-end SPEI observations from 427 meteorological stations across mainland China as supervision and combines 0.1° monthly dynamic meteorological factors from CMFD with their corresponding time-memory features and consistently defined NASADEM terrain predictors to establish a station-supervised reconstruction framework on a 1 km target grid. Following this design, we constructed and technically validated a monthly multi-timescale SPEI dataset on a 1 km grid for mainland China covering 1979–2018. By linking station-level drought characterization with nationwide gridded reconstruction, the dataset provides a complementary, spatially detailed SPEI product for mainland China.

This study aims to develop and validate a station-supervised monthly multi-timescale SPEI dataset for mainland China on a 1 km grid. The objectives are to: (1) construct monthly station-based SPEI labels from daily station observations; (2) integrate monthly meteorological forcing, temporal memory features, and topographic factors for machine-learning reconstruction; (3) generate nationwide monthly SPEI fields at five accumulation timescales; and (4) evaluate the dataset through sample-level, spatial, temporal, product-consistency, drought-condition, and predictor-contribution analyses. The resulting dataset provides a station-validated SPEI resource for drought monitoring, hydroclimatic assessment, ecosystem analysis, and climate-impact research, and demonstrates a practical reconstruction strategy for regions where long-term fine-resolution dynamic meteorological data are limited.

2. Materials and Methods

2.1. Study Area

The study area covers mainland China, which spans a broad range of climatic, topographic, and ecological conditions. The region includes humid and semi-humid monsoon areas in eastern and southern China, arid and semi-arid regions in northern and northwestern China, and high-altitude areas such as the Tibetan Plateau. These diverse environmental conditions lead to substantial regional differences in drought occurrence, persistence, and intensity. Therefore, high-resolution drought datasets are important for representing local dry-wet variability and supporting regional drought monitoring and climate-impact assessment.

The spatial domain of the dataset was defined over mainland China, and the final SPEI products were generated on a 1 km grid. The station-based SPEI observations used for supervision were derived from 427 meteorological stations distributed across mainland China. Their spatial distribution is shown in Figure 1.

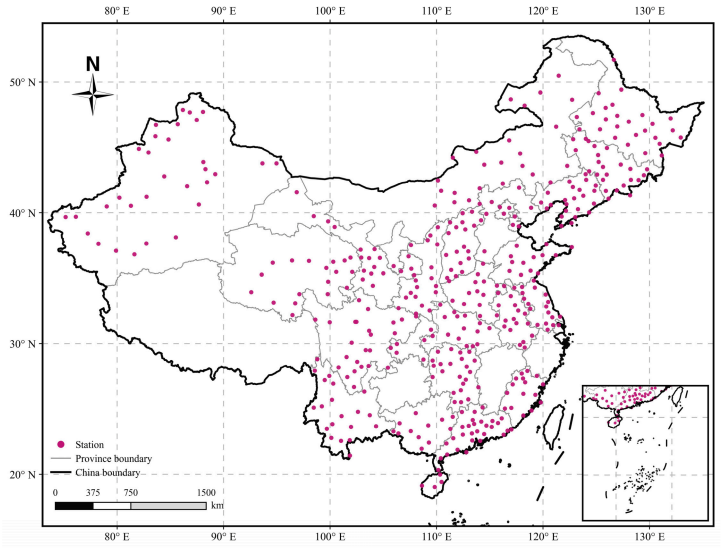


Figure 1. Study area and spatial distribution of the 427 meteorological stations used for SPEI-supervised reconstruction over mainland China.

2.2. Station-Based SPEI Observations

The station-based SPEI observations used in this study were derived from the published daily multi-timescale SPEI dataset for mainland China. This dataset was constructed based on 427 meteorological stations across mainland China and covers the period 1961–2018. It provides daily SPEI series at multiple accumulation timescales, offering a reliable station-level reference for drought characterization.

In this study, the station-based SPEI observations were used as the supervisory source for reconstructing the monthly gridded SPEI dataset. Specifically, daily SPEI series at the selected accumulation timescales were extracted from the station dataset and subsequently converted into monthly station labels during the data processing stage. Although the original station-based SPEI dataset covers 1961–2018, the final gridded product developed in this study covers 1979–2018, consistent with the temporal availability of the CMFD meteorological forcing data used for nationwide reconstruction.

2.3. Meteorological Forcing Data

Monthly meteorological forcing data were obtained directly from the 0.1° monthly CMFD v2.0 files (01mo_010deg_197901-201812.nc) for 1979–2018. No aggregation from

daily or sub-daily CMFD records was performed in this workflow. The monthly files were used as the primary dynamic predictors for reconstructing the SPEI fields.

Seven monthly CMFD variables were used: longwave radiation (lrad), precipitation (prec), air pressure (pres), specific humidity (shum), shortwave radiation (srad), air temperature (temp), and wind speed (wind). Precipitation was converted from the monthly mean rate stored in the source files (mm h^{-1}) to calendar-month accumulation (mm $month^{-1}$) by multiplying by 24 and the number of days in the corresponding month; the other six variables retained their source-file units. Station locations and 1 km pixel centres were sampled with one shared NaN-aware bilinear implementation. Bilinear weights were renormalized over finite corner cells, and nearest-valid fallback was used only when all four interpolation corners were invalid. This sampler provided consistent spatial matching of CMFD forcing for station extraction and nationwide inference.

2.4. Static Environmental Data

Static environmental predictors were used to represent local spatial heterogeneity and environmental constraints under coarse-resolution meteorological forcing. The topographic data were derived from the NASADEM elevation product [38]. Based on the original elevation data, terrain-related variables were further generated, including elevation, local terrain relief, and local elevation variability. These topographic factors were used to characterize elevation background and local terrain differences, which may influence regional hydroclimatic conditions and drought spatial patterns.

NASADEM_HGT.001 tiles are provided in geographic coordinates (EPSG:4326) at approximately 1 arc-second resolution. The tiles were assembled into a virtual raster and warped once to the common 1000 m target grid. GDAL Warp resampling modes average, max, min, and rms were used to derive mean elevation, maximum elevation, minimum elevation, and root-mean-square elevation, respectively. Terrain relief was calculated as maximum minus minimum elevation, and elevation standard deviation was calculated as $\mathrm{sqrt}(\max(\mathrm{RMS}^2 - \mathrm{mean}^2, 0))$. The resulting elevation, relief, and elevation-standard-deviation rasters form one canonical terrain stack used without definitional changes in both station-sample construction and nationwide inference.

NASADEM terrain predictors were represented by dem_elev_m.tif, dem_relief_1km.tif, and dem_std_1km.tif on the 1 km target grid. Station values were sampled from these rasters, and nationwide inference used the same three layers without further terrain calculation. Training and gridded inference therefore shared identical terrain definitions, CRS, transform, extent, pixel alignment, mask, and NoData treatment.

The data sources used for SPEI reconstruction are summarized in Table 1.

Table 1. Data sources used for SPEI reconstruction.

Data Source	Variables/Information	Spatial Resolution	Temporal Coverage	Role in This Study
Station-based SPEI observations	Daily SPEI series at multiple accumulation timescales from 427 meteorological stations	Station level	1961–2018	Supervisory source for monthly SPEI label generation
CMFD v2.0 meteorological forcing	Monthly lrad, prec, pres, shum, srad, temp, and wind	0.1° (NaN-aware bilinear mapping)	1979–2018	Dynamic meteorological forcing for station-level training and nationwide gridded reconstruction
NASADEM	Original elevation and derived mean elevation, relief, and elevation variability	≈1 arc-second source; terrain predictors at ≈1 km support	Static	Static topographic constraints for representing local terrain heterogeneity

2.5. Reconstruction Framework

2.5.1. Dataset Production Process

This study used daily station-based SPEI observations from 427 meteorological stations as supervision. Within each station, accumulation timescale, year, and month, the monthly target was defined as the SPEI value on the exact final day of the source dataset's fixed 365-day no-leap calendar. 28 February was used as the February endpoint in every year, and the usual calendar month-end dates were used for all other months. No fallback to an earlier available or non-missing daily observation was permitted. Records with missing or invalid values on the required month-end date were excluded from model fitting. The resulting labels were combined with the timescale-specific predictors described below.

Based on these inputs, five independent LightGBM regression models were fitted, one for each SPEI accumulation timescale, and applied over mainland China on the target 1 km grid. Monthly CMFD predictors were mapped to 1 km pixel centres using the same NaN-aware bilinear sampler used for station extraction. Spatial variation below the CMFD grid scale was estimated from fitted relationships with location, temporal features, and fine-grid terrain predictors. This design combines regional meteorological forcing with temporal and topographic context to estimate monthly SPEI on the target grid.

Following the above procedure, a monthly multi-timescale SPEI dataset for mainland China was produced on a 1 km grid. The overall workflow is shown in Figure 2.

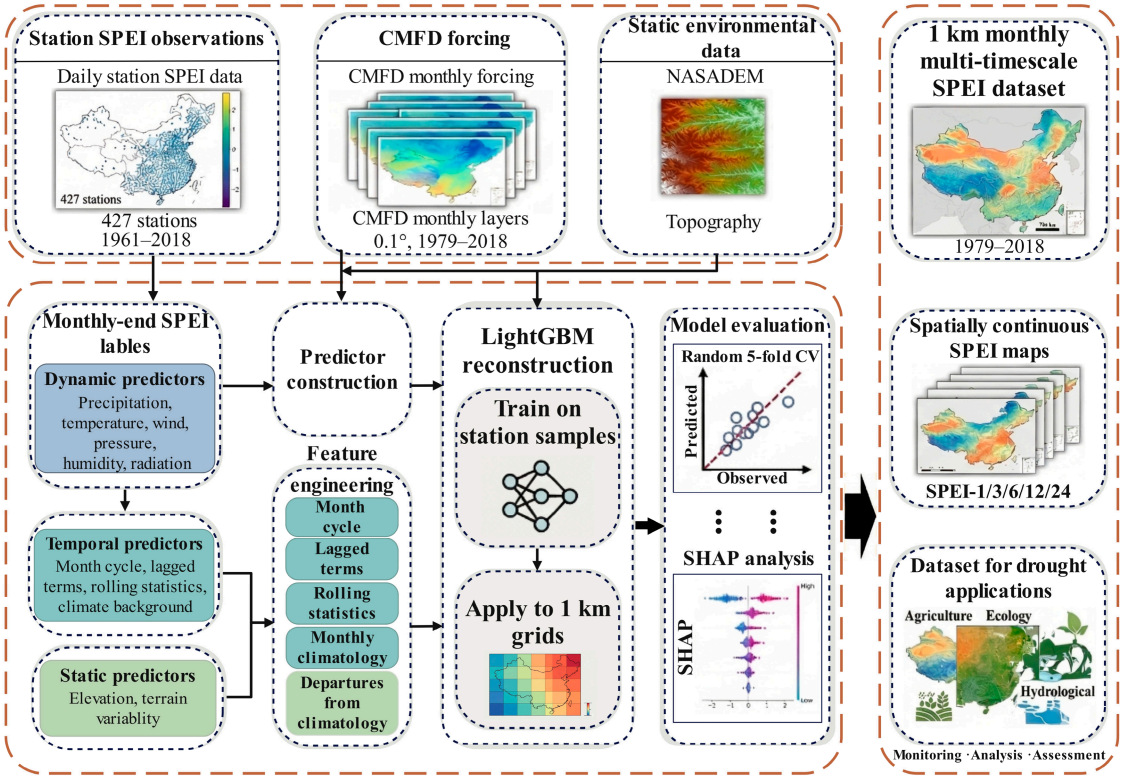


Figure 2. Workflow for constructing the monthly multi-timescale SPEI dataset over mainland China on a 1 km grid. Daily station-based SPEI observations were converted into monthly labels at five accumulation timescales. Monthly CMFD predictors, timescale-dependent temporal features, and 1 km terrain covariates were then constructed. Five station-supervised LightGBM models were used to generate the nationwide products.

2.5.2. Monthly Station Label Generation

This study directly used the published daily multi-timescale station-based SPEI dataset for mainland China as the supervisory source rather than recalculating SPEI from mete-

orological variables. For each of the 427 stations and each accumulation timescale, the value on the exact final day of each month in the source 365-day no-leap calendar was retained. Station ID, year, month, source date, and timescale were preserved as sample identifiers, and duplicate station-timescale-month records were rejected. This procedure yielded 204,960 candidate month-end labels per timescale over 1979–2018. After requiring complete timescale-specific lag and rolling windows, the final training tables contained 204,960, 204,106, 202,825, 200,263, and 195,139 samples for SPEI-1, SPEI-3, SPEI-6, SPEI-12, and SPEI-24, respectively. Each monthly label is a single sample from a precomputed daily SPEI series, not a sum or mean of the 28–31 daily values within that month; therefore, variation in calendar-month length does not redefine the accumulation window supplied by the source product.

Monthly labels were constructed separately for SPEI-1, SPEI-3, SPEI-6, SPEI-12, and SPEI-24. The fixed source-calendar month-end rule provided a consistent sampling date across stations and aligned the precomputed daily SPEI series with monthly CMFD predictors.

2.5.3. Predictor Construction

The predictor system used for monthly multi-timescale SPEI reconstruction comprised dynamic meteorological factors, time-related factors, and static topographic factors. The composition of each predictor group, its temporal treatment, spatial resolution, and role in reconstruction are summarized in Table 2. Together, these predictors represent regional-scale hydroclimatic conditions, seasonal cyclicity, multi-timescale cumulative memory, and local topographic constraints.

Table 2. Predictor groups used for monthly multi-timescale SPEI reconstruction.

Predictor Group	Variables	Temporal Treatment	Spatial Resolution	Role in Reconstruction
Dynamic meteorological predictors	Precipitation, temperature, wind speed, pressure, specific humidity, shortwave radiation, longwave radiation	Monthly values; no lag for SPEI-1; lag 1 to s − 1 for s = 3, 6, 12, and 24; rolling mean for all variables; rolling sum for precipitation; climatology and anomalies	0.1° CMFD sampled by NaN-aware bilinear interpolation	Describe regional hydroclimatic forcing and temporal variability
Cyclic temporal predictors	Month sine, month cosine	Cyclic encoding of calendar month	Not spatially explicit	Represent annual seasonality
Static topographic predictors	Elevation, local terrain relief, local elevation variability	Time-invariant; identical canonical 1 km rasters for station training and inference	≈1 km spatial support in both stages	Represent topographic background and sub-grid spatial heterogeneity

The dynamic predictors were derived from the monthly 0.1° CMFD files. For each station-month, lrad, prec, pres, shum, srad, temp, and wind were extracted with the shared NaN-aware bilinear sampler and retained with the temporal identifiers as the basis for constructing scale-dependent temporal features.

Time-related predictors represented long-term trend, seasonality, and cumulative hydroclimatic behaviour. Raw year and month, together with sine and cosine encodings of month, were retained. SPEI-1 used no lagged CMFD predictors. For SPEI-3, SPEI-6, SPEI-12, and SPEI-24, monthly lags 1–2, 1–5, 1–11, and 1–23, respectively, were generated for all seven CMFD variables. Rolling means used the complete target window for every variable, and precipitation additionally received a complete-window rolling sum (min_periods = scale). Samples lacking any required lag or rolling value were removed, which determined the scale-specific starting months and sample counts. Remaining sporadic predictor missing values were handled within each model fit by a median imputer

estimated from the corresponding training subset only; no standardization or normalization was applied.

Monthly climatologies and anomalies were constructed in addition to lagged and rolling statistics. The seven CMFD variables were represented by their monthly values, multi-year calendar-month climatologies, and departures from those climatologies. Multi-year background states were also calculated for timescale-specific rolling means, while rolling precipitation totals were represented by their background accumulations and corresponding departures. These features jointly characterize regional climatic background, contemporaneous anomalies, and multi-timescale cumulative processes.

The static reconstruction predictors consisted of the three NASADEM terrain covariates. Station values were sampled from the 1 km elevation, relief, and elevation-standard-deviation rasters described in Section 2.4, and nationwide inference used the same raster layers with identical CRS, transform, extent, pixel alignment, and NoData mask.

Monthly labels and predictors were assembled independently for the five timescales. After complete-window filtering, the SPEI-1, SPEI-3, SPEI-6, SPEI-12, and SPEI-24 training tables contained 204,960, 204,106, 202,825, 200,263, and 195,139 rows, respectively. The corresponding reconstruction models used 54, 68, 89, 131, and 215 predictors. Nationwide inference followed each timescale-specific training schema with identical feature names, column order, and 1 km terrain rasters.

2.5.4. LightGBM-Based SPEI Reconstruction

After the construction of monthly station labels and multi-source predictors, a station-supervised machine-learning reconstruction model was further established, with LightGBM selected as the unified core model [39]. For each timescale $s \in \{1, 3, 6, 12, 24\}$, the model takes the feature vector $x_i^{(s)}$ of the monthly station sample as input and the corresponding monthly station SPEI value $y_i^{(s)}$ as the supervisory target, thereby learning the direct mapping between the input features and the target variable. The corresponding prediction function can be expressed as

$$\hat{y}_i^{(s)} = f^{(s)}\left(x_i^{(s)}\right) \tag{1}$$

where $f^{(s)}$ denotes the LightGBM model trained for timescale s, and $\hat{y}_i^{(s)}$ denotes the predicted SPEI value of the i-th sample at timescale s. The corresponding training objective is to minimize the squared error between predictions and observations, which can be written as

$$\mathcal{L}^{(s)} = \frac{1}{N_s}\sum_{i=1}^{N_s}\left(y_i^{(s)} - \hat{y}_i^{(s)}\right)^2 \tag{2}$$

where N_s represents the number of training samples at timescale s.

Five independent LGBMRegressor models were trained, one for each timescale. The fixed parameters were objective = 'regression', metric = 'rmse', n_estimators = 800, learning_rate = 0.03, num_leaves = 63, max_depth = −1, min_child_samples = 30, subsample = 0.8, subsample_freq = 0, colsample_bytree = 0.8, reg_alpha = 0.0, reg_lambda = 0.5, random_state = 42, n_jobs = 1, and verbose = −1. No early stopping or hyperparameter search was used. Following validation, each timescale-specific model was fitted to all valid complete-window samples available for nationwide inference. The workflow was implemented in Python 3.13 using LightGBM 4.6.0, scikit-learn 1.8.0, GDAL 3.11.4, Rasterio 1.5.0, and SHAP 0.51.0.

2.5.5. Nationwide 1 km Product Generation

During the generation of the nationwide 1 km product, the trained LightGBM model was further applied to 1 km grid cells across mainland China. Specifically, nationwide 1 km static environmental grids and the corresponding monthly CMFD dynamic meteorological

data were first prepared. Then, following the same feature-construction procedure used in the training stage, a feature vector $X_{g,t}^{(s)}$ was constructed for each 1 km grid cell at a given month t and timescale s. The trained model was subsequently applied to generate grid-level SPEI predictions:

$$\hat{Y}_{g,t}^{(s)} = f^{(s)}\left(X_{g,t}^{(s)}\right) \quad (3)$$

where $\hat{Y}_{g,t}{}^{(s)}$ denotes the predicted SPEI value for grid cell g in month t at timescale s. Month-by-month and timescale-specific inference for 1979–2018 generated the 1 km monthly multi-timescale SPEI dataset for mainland China.

Nationwide inference was performed month by month and timescale by timescale on an Interrupted Goode Homolosine target grid (ESRI:54052). The grid contains 5251 columns and 3869 rows at a 1000 m pixel size, with an affine transform (1000, 0, 7082250, 0, −1000, 5894000). For every target pixel, the corresponding monthly CMFD values were obtained with the shared NaN-aware bilinear sampler and then integrated with temporal predictors, coordinates, and the canonical 1 km terrain covariates.

Product output began only after the complete accumulation window for each timescale: January 1979 for SPEI-1, March 1979 for SPEI-3, June 1979 for SPEI-6, December 1979 for SPEI-12, and December 1980 for SPEI-24. The five products contain 480, 478, 475, 469, and 457 monthly rasters, respectively, for a total of 2359 GeoTIFF files.

The training and inference tables used the same timescale-specific feature names and column order. CMFD predictors were sampled from the 0.1° grid using NaN-aware bilinear interpolation, while terrain predictors for station samples and nationwide inference were obtained from the same 1 km raster stack. The inference code checked feature names and column order before prediction, ensuring consistent covariate definitions throughout model fitting and gridded reconstruction.

3. Results

3.1. Dataset Overview

The dataset described in this article is named China 1 km Monthly Multiscale SPEI Dataset over Mainland China (1979–2018), Version 2. It provides model-reconstructed monthly SPEI estimates over mainland China on a 1 km grid. The dataset includes SPEI-1, SPEI-3, SPEI-6, SPEI-12, and SPEI-24, representing drought conditions over different hydroclimatic accumulation windows.

The dataset provides spatially continuous monthly SPEI fields on a 1 km grid. SPEI is standardized and dimensionless: negative values indicate drier-than-normal conditions and positive values indicate wetter-than-normal conditions. The reconstruction combines temporally consistent CMFD forcing, timescale-dependent memory features, coordinates, and NASADEM terrain covariates. It represents national and regional dry–wet variations on the target grid, while local patterns remain model estimates constrained by the coarser dynamic forcing.

3.2. Data Format and File Organization

The released dataset is organized as six compressed files in Science Data Bank, accompanied by a SHA-256 checksum text file. Five archive files correspond to SPEI-1, SPEI-3, SPEI-6, SPEI-12, and SPEI-24 and collectively contain 2359 monthly GeoTIFF rasters. An additional documentation and metadata archive provides README files, metadata tables, checksums, release notes, and preview information required for data interpretation and reuse.

Inside each SPEI archive, the monthly rasters are stored as single-band GeoTIFF files. Each GeoTIFF file represents one monthly SPEI field over mainland China. The file naming rule follows spei_XXmo_YYYY-MM.tif, where XX denotes the SPEI accumulation timescale

and YYYY-MM denotes the corresponding month. Therefore, users can download a specific timescale archive and extract monthly GeoTIFF files according to their research needs. The file organization of the released dataset is summarized in Table 3.

Table 3. File organization of the released 1 km monthly multi-timescale SPEI dataset.

Released File	Description	Internal Data Format	Temporal Coverage	Notes
China_1km_monthly_multiscale_SPEI_v2_SPEI01_1979-01_2018-12.tar	SPEI-1 monthly raster archive	GeoTIFF	1979-01 to 2018-12	480 monthly files
China_1km_monthly_multiscale_SPEI_v2_SPEI03_1979-03_2018-12.tar	SPEI-3 monthly raster archive	GeoTIFF	March 1979–December 2018	478 monthly files
China_1km_monthly_multiscale_SPEI_v2_SPEI06_1979-06_2018-12.tar	SPEI-6 monthly raster archive	GeoTIFF	1979-06 to 2018-12	475 monthly files
China_1km_monthly_multiscale_SPEI_v2_SPEI12_1979-12_2018-12.tar	SPEI-12 monthly raster archive	GeoTIFF	1979-12 to 2018-12	469 monthly files
China_1km_monthly_multiscale_SPEI_v2_SPEI24_1980-12_2018-12.tar	SPEI-24 monthly raster archive	GeoTIFF	1980-12 to 2018-12	457 monthly files
China_1km_monthly_multiscale_SPEI_v2_docs_metadata.zip	Documentation and metadata archive	CSV, TXT, MD, preview figures	Dataset-wide	Includes README, metadata tables, checksums, release notes, and preview information

3.3. Variables and Metadata

The dataset contains five SPEI variables corresponding to five accumulation timescales: SPEI-1, SPEI-3, SPEI-6, SPEI-12, and SPEI-24. All variables are provided as monthly single-band GeoTIFF rasters on the same 1 km grid. SPEI is standardized and dimensionless, with negative values indicating drier-than-normal conditions and positive values indicating wetter-than-normal conditions. Product start dates reflect the complete accumulation window: SPEI-1 begins in January 1979, SPEI-3 in March 1979, SPEI-6 in June 1979, SPEI-12 in December 1979, and SPEI-24 in December 1980.

All monthly rasters share the same grid geometry, with 5251 columns and 3869 rows in ESRI:54052 (Interrupted Goode Homolosine), a 1000 m pixel size, and affine transform (1000, 0, 7082250, 0, −1000, 5894000). Files are single-band float32 GeoTIFFs with NoData = −9999, LZW compression, and tiled storage. The metadata includes raster inventories, archive checksums, alignment checks, validation summaries, and documentation. Alignment checks confirmed consistent raster shape, projection, transform, and NoData settings across all five timescales.

3.4. Data Access

The dataset has been deposited in Science Data Bank. The public dataset DOI is https://doi.org/10.57760/sciencedb.34500. The repository provides the release package, including monthly SPEI GeoTIFF archives at five accumulation timescales, metadata files, documentation, and preview figures. Detailed citation and access information are also provided in the Data Availability section.

3.5. Technical Validation

3.5.1. Validation Metrics

To evaluate the technical reliability of the reconstructed SPEI dataset, several groups of validation metrics were used. For station-level reconstruction accuracy and out-of-station spatial generalization, the coefficient of determination (R^2), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Bias were used. R^2 measures the explanatory

ability of the reconstructed values for the observed variance, RMSE and MAE characterize the magnitude of reconstruction errors, and Bias reflects the systematic deviation between reconstructed and observed SPEI values. These metrics are defined as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \overline{y})^2} \tag{4}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2} \tag{5}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|\hat{y}_i - y_i| \tag{6}$$

$$\text{Bias} = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i) \tag{7}$$

where y_i and $\hat{y}_i$ denote the observed and predicted values of the i-th sample, respectively, $\overline{y}$ denotes the mean of the observed values, and N denotes the total number of validation samples.

For product-level consistency assessment with existing SPEI datasets, Pearson correlation coefficient, RMSE, MAE, and Bias were calculated using common valid pixels and months. Pearson correlation coefficient was used to describe the temporal consistency between the reconstructed dataset and reference SPEI products.

For drought-condition detection validation, the reconstructed and observed SPEI values were converted into binary drought-condition labels under different drought severity thresholds. Precision, Recall, F1-score, False Alarm Ratio (FAR), and Critical Success Index (CSI) were used to evaluate the ability of the dataset to identify drought conditions. These metrics are defined as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{8}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{9}$$

$$\text{F1} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{10}$$

$$\text{FAR} = \frac{\text{FP}}{\text{TP} + \text{FP}} \tag{11}$$

$$\text{CSI} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \tag{12}$$

where TP, FP, FN, and TN denote true positives, false positives, false negatives, and true negatives, respectively. Recall is also referred to as hit rate or probability of detection in drought-condition detection.

For the interpretation of dry and wet conditions, the commonly used SPEI classification standard was adopted. SPEI values of ≤ -2.00, -1.99 to -1.50, -1.49 to -1.00, and -0.99 to -0.50 correspond to extreme, severe, moderate, and mild drought, respectively; values from -0.49 to 0.49 indicate near-normal conditions; and values from 0.50 to 0.99, 1.00 to 1.49, 1.50 to 1.99, and ≥ 2.00 correspond to mild, moderate, severe, and extreme wet conditions, respectively [40]. In the drought-condition detection validation, cumulative drought thresholds were used to assess moderate-or-worse drought (SPEI ≤ -1.0), severe-or-worse drought (SPEI ≤ -1.5), and extreme drought (SPEI ≤ -2.0).

3.5.2. Station-Level Reconstruction Accuracy and Spatial Generalization

Station-level validation used three complementary five-fold designs. Random sample-level K-Fold quantified interpolation performance when months from the same station could occur in training and validation subsets. Station-grouped GroupKFold kept every station entirely within one fold and was retained as a secondary unseen-station comparison. The primary spatial-transfer assessment used 200 km spatial blocks together with a 100 km exclusion buffer around every test station.

For random five-fold cross-validation, individual monthly station samples at each timescale were partitioned using KFold. Consequently, observations from the same station could occur in both training and validation folds. The LightGBM model was refitted in each fold, the four metrics were calculated separately for that fold, and the five fold-level metric values were then averaged to obtain the reported result for each timescale.

Random five-fold cross-validation results are presented in Table 4. Across SPEI-1, SPEI-3, SPEI-6, SPEI-12, and SPEI-24, mean R^2 ranged from 0.8807 to 0.9122, mean RMSE ranged from 0.2941 to 0.3436, and Bias remained close to zero. These values characterize sample-level interpolation performance; spatial transfer is evaluated separately using the spatial-block design below.

Table 4. Random 5-fold cross-validation performance of LightGBM for monthly multi-timescale SPEI reconstruction.

Timescale	R^2	RMSE	MAE	Bias
SPEI-1	0.8884	0.3352	0.2369	−0.0001
SPEI-3	0.8818	0.3436	0.2462	0.0002
SPEI-6	0.8807	0.3434	0.2469	0.0000
SPEI-12	0.9035	0.3061	0.2205	0.0001
SPEI-24	0.9122	0.2941	0.2158	0.0001

Under random KFold, SPEI-24 achieved the highest mean R^2 (0.9122) and lowest mean RMSE (0.2941), followed by SPEI-12. All five timescales had mean R^2 above 0.88. The more conservative spatial-block results reported below are used for claims about transfer to geographically separated station locations.

Because ordinary station GroupKFold can retain nearby stations or stations sharing similar CMFD forcing in the training set, it was not described as fully spatially independent. For the stricter design, stations were assigned to 200 km spatial blocks, complete blocks formed each test fold, and all nominal training stations within 100 km of any test station were removed. Each fold contained 85–86 test stations and 291–316 retained training stations. The minimum test-to-training distance was 100.0–100.4 km, and no station occurred in both subsets. Fold-level metrics were averaged and are reported in Table 5.

Table 5. Five-fold spatial-block validation performance using 200 km blocks and a 100 km exclusion buffer.

Timescale	R^2	RMSE	MAE	Bias
SPEI-1	0.8739	0.3563	0.2541	0.0005
SPEI-3	0.8594	0.3746	0.2700	0.0009
SPEI-6	0.8470	0.3889	0.2805	−0.0006
SPEI-12	0.8276	0.4089	0.2934	−0.0019
SPEI-24	0.7857	0.4592	0.3333	0.0002

Under spatial-block validation, mean R^2 decreased from 0.8739 for SPEI-1 to 0.7857 for SPEI-24, while mean RMSE increased from 0.3563 to 0.4592. Mean Bias ranged from

−0.0019 to 0.0009. These results provide a stronger assessment of transfer to geographically separated stations than random KFold or ordinary station GroupKFold.

The systematic reduction in R^2 toward longer accumulation timescales indicates that long-memory SPEI is more difficult to transfer across spatially separated regions. Performance remained substantially above a mean-only prediction at every timescale, with near-zero average Bias. Spatial-block validation therefore provides the primary assessment of spatial transfer, while station-grouped GroupKFold provides a complementary sensitivity analysis.

Figure 3 presents a station-level comparison of SPEI-3 among the station observations and the three gridded products at five selected stations. For Our SPEI, every station-month value is a spatial-block out-of-fold prediction generated by a model that excluded the station's complete spatial block and all training stations within the 100 km buffer. Thus, none of the displayed observations were used to train the corresponding model. GPR-SPEI and Zhang-SPEI were sampled at the same station locations and months for product-level context.

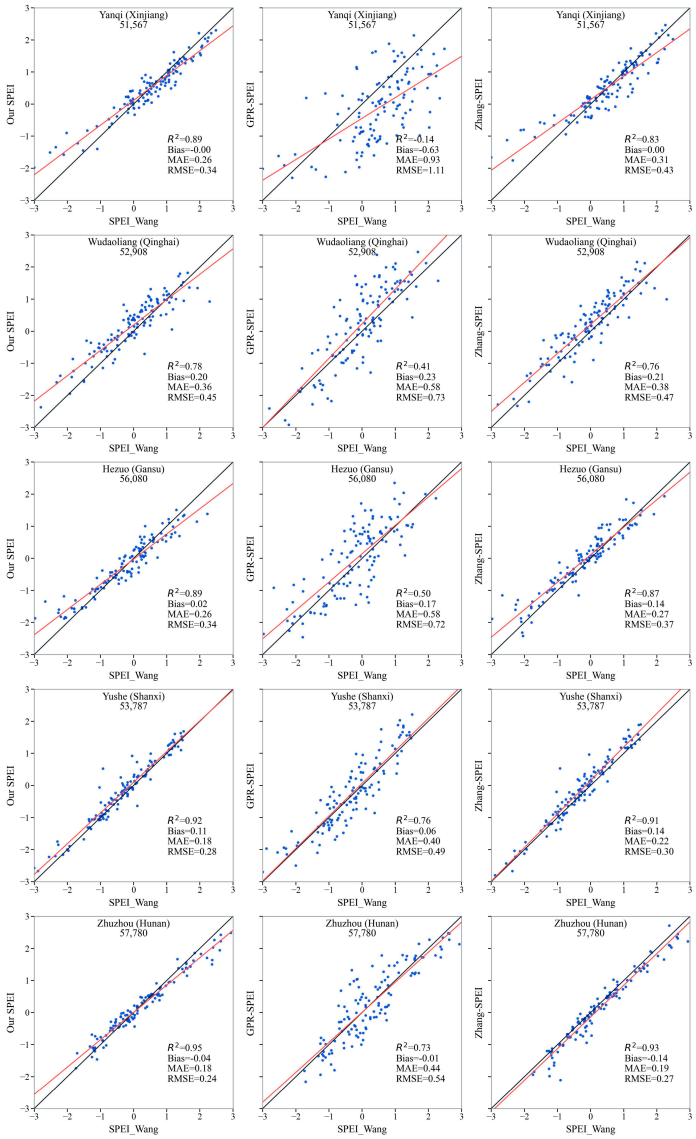


Figure 3. Station-level scatter comparison of SPEI-3 between Wang station observations and three gridded SPEI products at five selected stations. Our SPEI values are spatial-block out-of-fold predictions from the 200 km block and 100 km exclusion-buffer validation. Blue circles denote paired monthly observations and predictions, the black line denotes the 1:1 reference, and the red line denotes the fitted linear regression.

3.5.3. Spatial Characteristics and Region-Specific Uncertainty

Region-specific uncertainty was quantified from spatial-block out-of-fold predictions by province, elevation zone, long-term precipitation zone, and station-support class. Figure 4 shows the reconstructed fields for February and August 1998, selected as contrasting dry and wet months from the national SPEI-3 screening, at all five accumulation timescales.

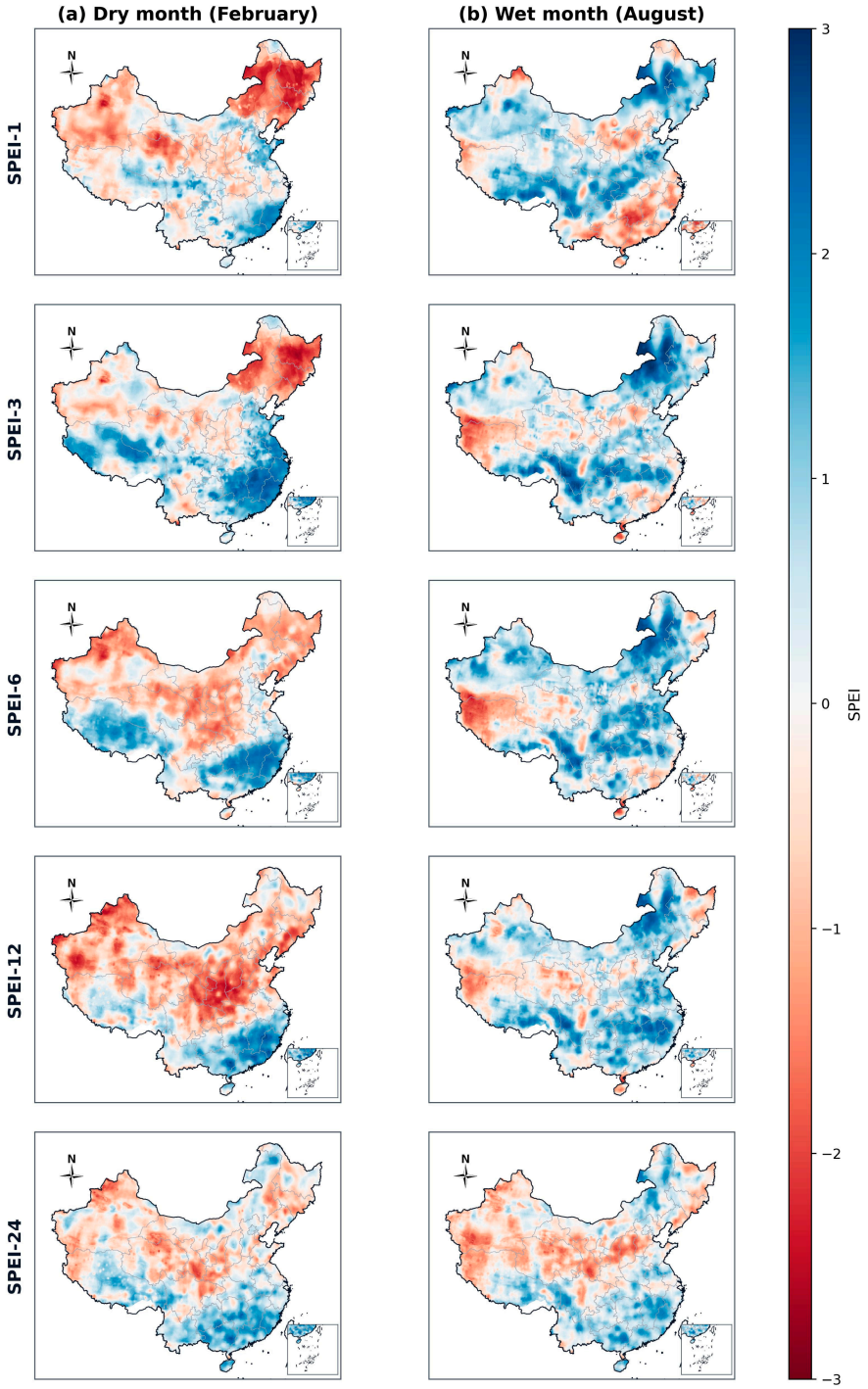


Figure 4. Spatial patterns of the reconstructed monthly multi-timescale SPEI dataset on the 1 km target grid for two contrasting months in 1998: (**a**) February and (**b**) August. The same colour scale is used for SPEI-1, SPEI-3, SPEI-6, SPEI-12, and SPEI-24.

In February 1998, the short-timescale fields show extensive negative SPEI over Northeast China and parts of northern and central China, with wetter conditions over parts

of southern and western China. In August 1998, positive SPEI is more extensive across northern, northeastern, and several southern regions, while dry pockets remain in western and southern transition zones. These maps illustrate coherent national-scale contrasts and scale-dependent spatial patterns.

The spatial pattern becomes progressively more persistent from SPEI-1 to SPEI-24 because longer accumulation windows integrate antecedent forcing. Quality control confirmed that all 2359 rasters share the same grid geometry and contain 9,428,399 valid mainland pixels. The observed scale dependence is therefore not caused by inconsistent masks or raster alignment.

Supplementary Figure S4 provides zoomed SPEI-3 maps for the southwest mountains, eastern Tibetan Plateau, and northern arid-transition zone. The ratios of mean pixel-to-pixel jumps at 0.1° CMFD boundaries to jumps inside CMFD cells were 1.0053, 1.0020, and 1.0023, respectively, indicating that bilinear mapping did not systematically amplify cell-boundary steps in these selected regions. A separate northeast diagnostic identified a band near 50°N associated with a strong gradient or discontinuity in the source CMFD precipitation field. Together, these diagnostics characterize both the spatial continuity of the reconstructed fields and the source-scale structure inherited from CMFD.

3.5.4. Product-Level Consistency with Existing SPEI Datasets

Product-level consistency was assessed against HSPEI for SPEI-1, SPEI-3, SPEI-6, SPEI-12, and SPEI-24 over 2001–2018. Zhang-SPEI and GPR-SPEI provided additional SPEI-3 comparisons over 1979–2018 and 1991–2000, respectively. Because the products use different forcing data, standardization baselines, and reconstruction methods, the analysis quantified inter-product consistency on common grids and periods.

For each product-timescale pair, comparisons used identical months, a common mainland-China valid mask, and a common grid. Our 1 km values were aggregated or reprojected to the reference grid rather than interpolating a coarser reference product to 1 km. Pearson correlation, RMSE, MAE, and Bias were calculated from common valid month-pixel pairs over the complete overlap, and the numbers of months and valid spatial cells were recorded. The resulting full-period consistency metrics are summarized in Table 6.

Table 6. Full-period consistency between the reconstructed SPEI product and existing SPEI products across all available common timescales.

Reference Product	Timescale	Common Period	Months	Valid Pixels	Pearson r	RMSE	MAE	Bias
Zhang-SPEI	SPEI-3	March 1979–December 2018	478	95,160	0.9295	0.3708	0.2711	0.0042
GPR-SPEI	SPEI-3	January 1991–December 2000	120	13,627,528	0.7350	0.7551	0.5795	−0.0174
HSPEI	SPEI-1	January 2001–December 2018	216	11,712,784	0.7772	0.6111	0.4695	0.0450
HSPEI	SPEI-3	January 2001–December 2018	216	11,712,784	0.6287	0.7441	0.5780	0.0448
HSPEI	SPEI-6	January 2001–December 2018	216	11,712,784	0.5059	0.8345	0.6462	−0.0482
HSPEI	SPEI-12	January 2001–December 2018	216	11,712,784	0.4547	0.8515	0.6673	−0.0561
HSPEI	SPEI-24	January 2001–December 2018	216	11,712,784	0.3267	0.9254	0.7279	−0.2031

Our SPEI showed strong full-period consistency with Zhang-SPEI at SPEI-3 (r = 0.9295, RMSE = 0.3708, MAE = 0.2711, Bias = 0.0042) and moderate consistency with GPR-SPEI (r = 0.7350). Against HSPEI, Pearson r decreased from 0.7772 for SPEI-1 to 0.3267 for SPEI-24, while RMSE increased from 0.6111 to 0.9254. The scale-dependent divergence indicates increasing sensitivity to differences in forcing, standardization baselines, and reconstruction methods at longer accumulation windows.

Figure 5 compares all 12 months of 1998 for SPEI-3 on harmonized display grids. Our SPEI and Zhang-SPEI show broadly similar national dry–wet organization, whereas GPR-SPEI is spatially smoother and exhibits larger regional contrasts in several months. These visual differences are consistent with the full-period metrics and the distinct source data and modelling strategies.

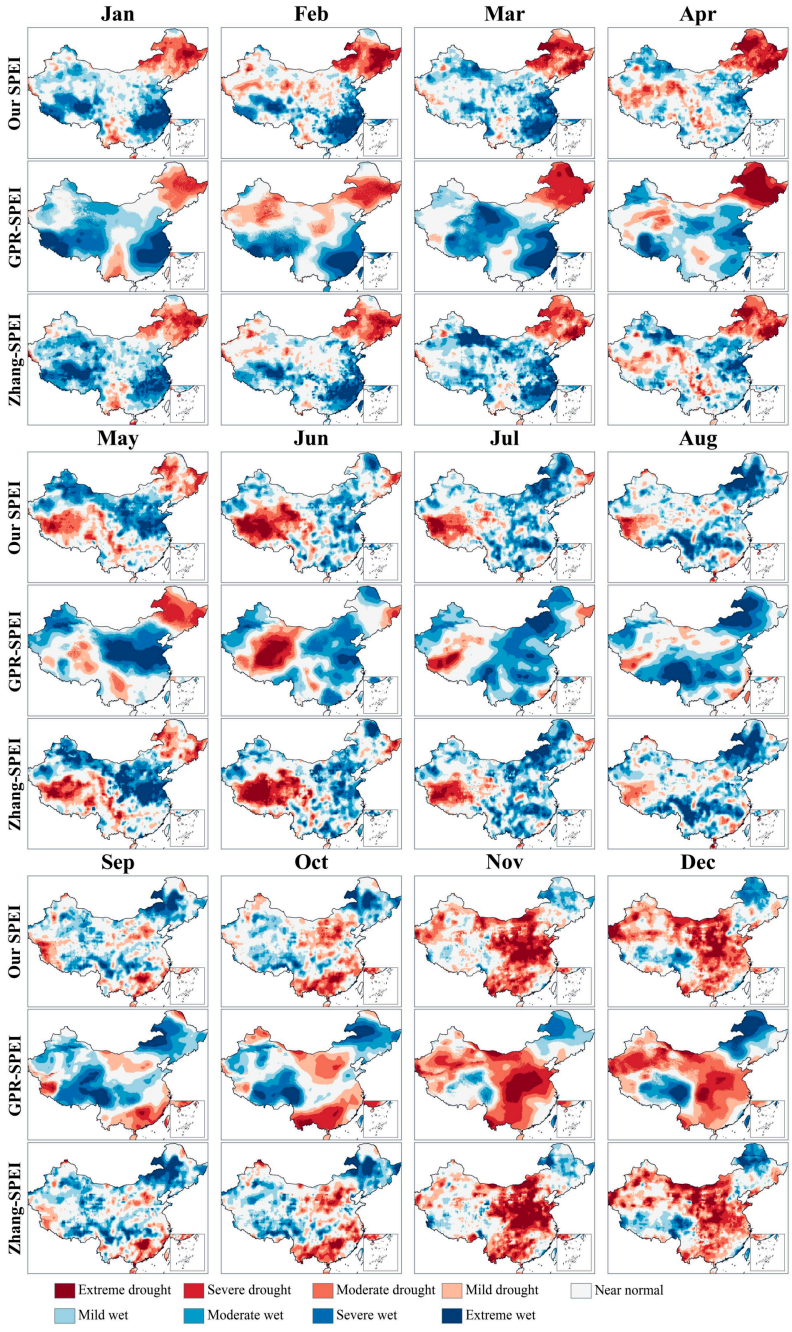


Figure 5. Monthly SPEI-3 spatial comparison among Our SPEI, GPR-SPEI, and Zhang-SPEI for January–December 1998. All products were harmonized to a common display grid and classified using the same SPEI severity intervals.

Figure 6 maps absolute product differences for the same months. Differences from Zhang-SPEI are generally smaller and more spatially localized than those from GPR-SPEI. Supplementary Figure S1 extends the grid-cell correlation comparison to all five HSPEI timescales, and Supplementary Figures S2, S3 and S5 show national monthly mean series over the corresponding overlap periods.

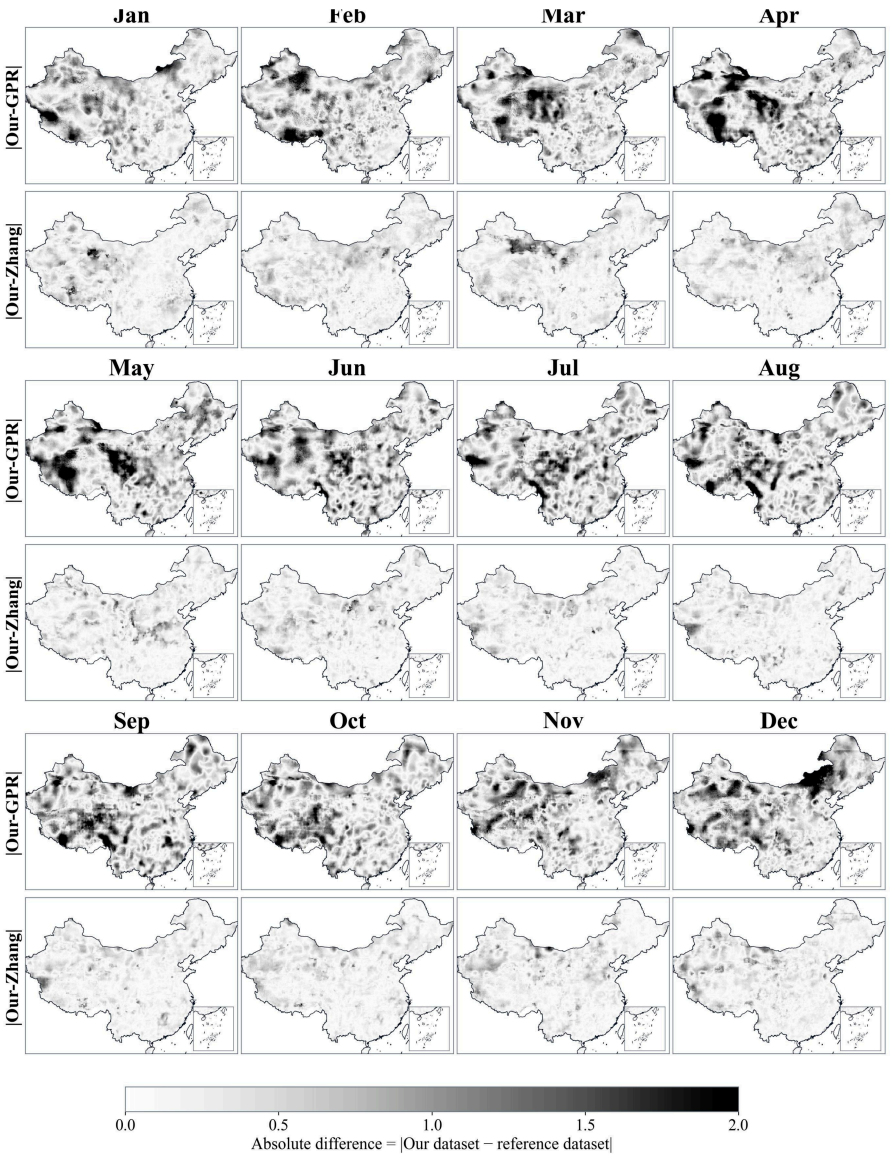


Figure 6. Monthly absolute SPEI-3 differences between Our SPEI and GPR-SPEI and between Our SPEI and Zhang-SPEI for January–December 1998. A common greyscale range is used for both comparisons.

The spatial contrasts in Figures 5 and 6 reflect differences in source data, standardization baselines, effective resolution, and reconstruction methods among the three products.

Spatial-block out-of-fold errors confirm greater uncertainty in western, high-altitude, arid, and sparse-support settings. For SPEI-24, R^2 was 0.5053 in Xinjiang and 0.6044 in Qinghai, 0.6053 at elevations of at least 3000 m, and 0.5836 in areas receiving less than 200 mm yr^{-1}. R^2 decreased from 0.8042 in the dense-support class to 0.7683 in the sparse-support class, whose mean nearest-station distance was 116.8 km. Detailed stratified results are reported in Supplementary Figure S7 and Tables S7 and S8.

Spatial uncertainty was non-uniform, with larger errors in western arid, high-elevation, and sparse-support regions, particularly at long accumulation timescales. These patterns provide a spatially explicit basis for interpreting regional differences in dataset performance.

3.5.5. Forward Temporal Transfer and Regional Time-Series Validation

Temporal transferability was evaluated using four expanding-window blocked tests. The training periods were 1979–1998, 1979–2003, 1979–2008, and 1979–2013, and the corresponding non-overlapping five-year test periods were 1999–2003, 2004–2008, 2009–2013, and 2014–2018. In every fold, all test years occurred after all training years. Monthly station climatologies and anomaly predictors were re-estimated using training years only, so information from the held-out period did not enter model fitting or anomaly construction.

Across SPEI-1, SPEI-3, SPEI-6, SPEI-12, and SPEI-24, the four-fold mean R^2 values were 0.8337, 0.8119, 0.7971, 0.7900, and 0.7577, respectively; the corresponding mean RMSE values were 0.4118, 0.4339, 0.4499, 0.4571, and 0.5017. The decline toward SPEI-24 indicates greater uncertainty when transferring long-memory relationships to future periods, while Bias remained between −0.0473 and −0.0155 across timescales.

For the dedicated Figure 7 experiment, the model was trained on 1979–1990 and evaluated on the entirely held-out 1991–2000 period. Pooled SPEI-3 performance over the held-out station-months was R^2 = 0.8484, RMSE = 0.3937, MAE = 0.2890, and Bias = 0.0331. The province-mean series provide a visual complement to these pooled metrics and to the four expanding-window tests.

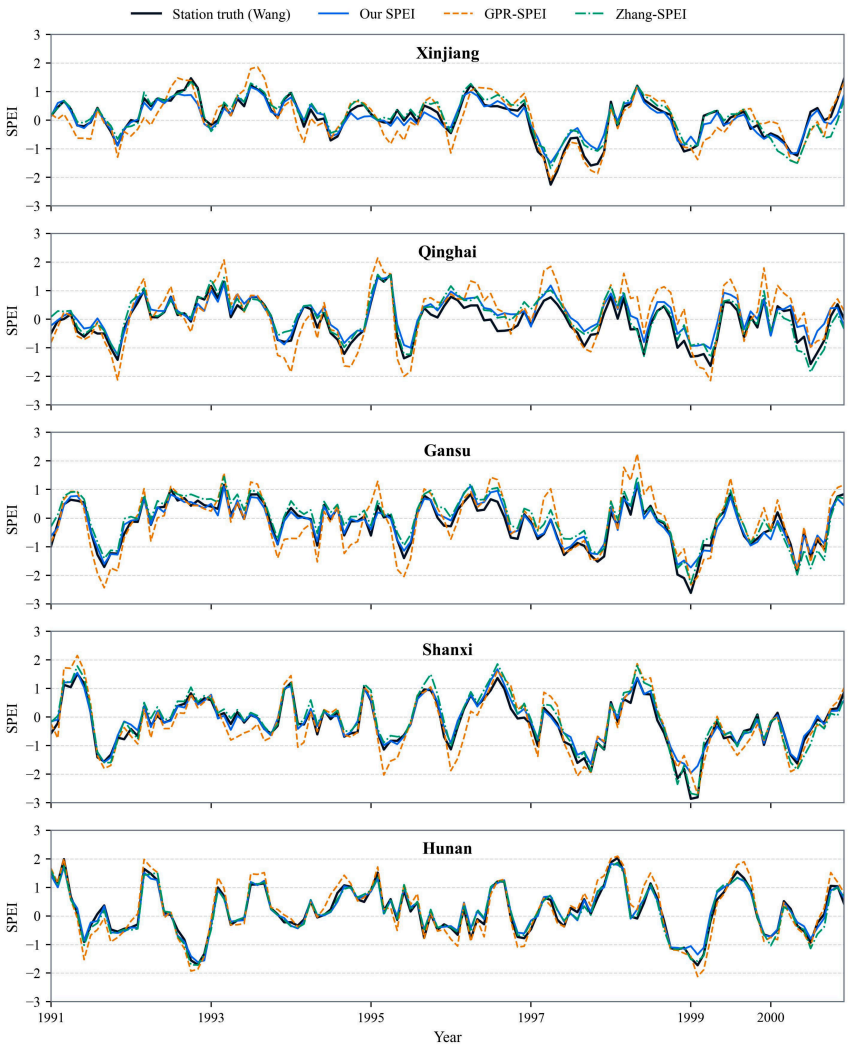


Figure 7. Regional mean time-series comparison of SPEI-3 among Wang station observations, temporally held-out Our SPEI predictions, GPR-SPEI, and Zhang-SPEI for five provinces during 1991–2000. The Our SPEI model was trained only on 1979–1990, and the complete 1991–2000 period was excluded from model fitting and anomaly-baseline estimation.

The forward blocked tests quantify temporal transfer to complete periods that were unavailable during training and show increasing uncertainty toward longer accumulation timescales. The selected province series illustrate regional behaviour during an entirely held-out decade and complement the pooled temporal-validation metrics.

3.5.6. Drought-Condition Detection Validation

Drought-condition detection was evaluated against station-based SPEI observations using the pooled out-of-fold predictions from the five-fold 200 km spatial-block validation with a 100 km exclusion buffer. Each test prediction was produced by a model that withheld the station's complete spatial block and all training stations within 100 km. Three cumulative drought-severity thresholds were considered: moderate-or-worse drought (SPEI ≤ -1.0), severe-or-worse drought (SPEI ≤ -1.5), and extreme drought (SPEI ≤ -2.0). For each timescale and threshold, observed and reconstructed SPEI values were converted into binary drought-condition labels, and Precision, Recall, F1-score, False Alarm Ratio (FAR), and Critical Success Index (CSI) were calculated.

Table 7 reports all three cumulative drought-severity thresholds. For moderate-or-worse drought, F1 ranged from 0.7021 to 0.7838 and CSI from 0.5410 to 0.6444. For severe-or-worse drought, Recall ranged from 0.4927 to 0.6210 and F1 from 0.5729 to 0.6823. For extreme drought, Precision remained between 0.6569 and 0.7401, but Recall decreased to 0.2397–0.3869, F1 to 0.3512–0.5053, and CSI to 0.2130–0.3380.

Table 7. Drought-condition detection performance at moderate-or-worse (SPEI ≤ -1.0), severe-or-worse (SPEI ≤ -1.5), and extreme (SPEI ≤ -2.0) thresholds.

Timescale	Threshold	Drought Level	Precision	Recall	F1-Score	FAR	CSI
SPEI-1	−1.0	Moderate-or-worse	0.8153	0.7545	0.7838	0.1847	0.6444
SPEI-1	−1.5	Severe-or-worse	0.7570	0.6210	0.6823	0.2430	0.5178
SPEI-1	−2.0	Extreme	0.7322	0.3857	0.5053	0.2678	0.3380
SPEI-3	−1.0	Moderate-or-worse	0.8020	0.7298	0.7642	0.1980	0.6184
SPEI-3	−1.5	Severe-or-worse	0.7496	0.5971	0.6647	0.2504	0.4978
SPEI-3	−2.0	Extreme	0.7274	0.3869	0.5051	0.2726	0.3379
SPEI-6	−1.0	Moderate-or-worse	0.7973	0.7222	0.7579	0.2027	0.6101
SPEI-6	−1.5	Severe-or-worse	0.7321	0.5670	0.6391	0.2679	0.4696
SPEI-6	−2.0	Extreme	0.7401	0.3573	0.4819	0.2599	0.3174
SPEI-12	−1.0	Moderate-or-worse	0.7770	0.7140	0.7442	0.2230	0.5926
SPEI-12	−1.5	Severe-or-worse	0.6982	0.5415	0.6100	0.3018	0.4388
SPEI-12	−2.0	Extreme	0.6927	0.2824	0.4012	0.3073	0.2509
SPEI-24	−1.0	Moderate-or-worse	0.7547	0.6563	0.7021	0.2453	0.5410
SPEI-24	−1.5	Severe-or-worse	0.6844	0.4927	0.5729	0.3156	0.4015
SPEI-24	−2.0	Extreme	0.6569	0.2397	0.3512	0.3431	0.2130

The lower Recall at SPEI ≤ -2.0 indicates regression-to-the-mean shrinkage for rare extremes, with the largest missed-event rates at SPEI-12 and SPEI-24. The dataset is therefore most effective for regional monthly drought characterization, while station observations and operational information remain important for high-impact extreme events, particularly in sparsely monitored regions.

3.5.7. Feature Contribution Analysis

SHAP analysis was conducted for the SPEI-1 and SPEI-12 models fitted to all available station-month samples. A random subset of 5000 samples was analysed with TreeExplainer. Figure 8 reports the mean absolute SHAP rankings and beeswarm distributions: panels a and c show overall mean absolute contributions, while panels b and d show the distribution and direction of feature contributions. The SHAP results characterize fitted model behaviour and predictor contributions.

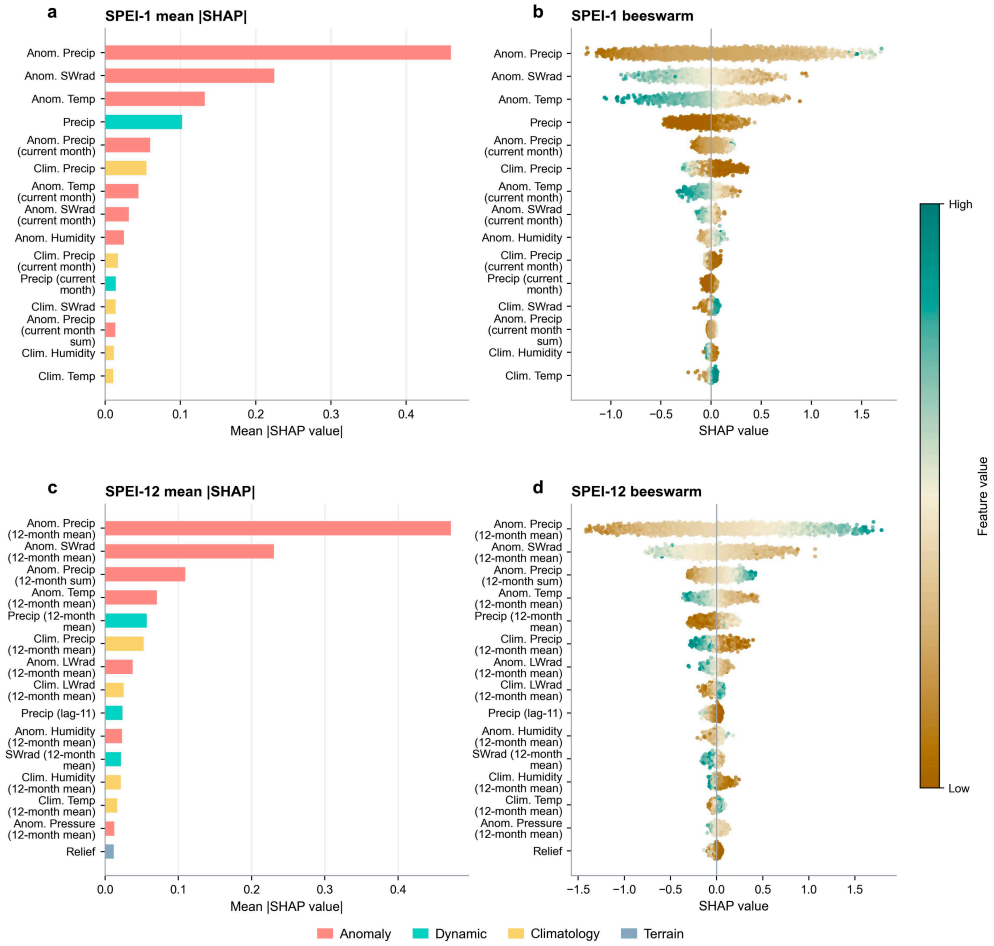


Figure 8. SHAP-based plausibility analysis for the SPEI-1 and SPEI-12 models fitted to all available station-month samples, using a random subset of 5000 samples and TreeExplainer. Panels show mean absolute SHAP rankings and beeswarm distributions.

For SPEI-1, the high-contribution features are mainly related to contemporaneous meteorological anomalies. As shown in Figure 8a, precipitation anomaly has the largest mean absolute SHAP value, followed by shortwave radiation anomaly and air temperature anomaly. Figure 8b further shows that these anomaly-related variables have wide SHAP value distributions, indicating their strong influence on short-timescale SPEI reconstruction. This suggests that SPEI-1 is more sensitive to immediate monthly hydroclimatic fluctuations.

For SPEI-12, the dominant predictors are mainly associated with long-window cumulative or rolling hydroclimatic features. As shown in Figure 8c, 12-month precipitation anomaly, 12-month mean precipitation anomaly, 12-month mean shortwave radiation anomaly, and 12-month mean temperature anomaly rank among the most important predictors. The beeswarm distribution in Figure 8d also shows that these long-window variables contribute substantially to model output. This indicates that long-timescale SPEI reconstruction relies more on cumulative background conditions and persistent hydroclimatic signals rather than single-month anomalies.

Overall, the SHAP results show that the models use distinct predictor information across accumulation timescales. Short-timescale reconstruction relies primarily on contemporaneous meteorological anomalies, whereas long-timescale reconstruction depends more strongly on cumulative hydroclimatic memory. The importance rankings are therefore consistent with the expected transition from immediate atmospheric anomalies to persistent multi-month forcing as the SPEI accumulation window increases.

4. Discussion

4.1. Methodological Advantages and Future Development

A major advantage of the dataset is its station-supervised, multi-source reconstruction strategy. It combines temporally continuous 0.1° CMFD meteorological forcing with monthly station-based SPEI supervision, timescale-dependent memory features, and fine-grid NASADEM terrain covariates. This design provides a practical pathway for producing spatially detailed drought estimates from widely available multi-source information. Bilinear mapping improves continuity across coarse-grid boundaries, while gradients or discontinuities already present in CMFD can remain in the reconstructed fields, as illustrated by the band near 50°N.

The framework integrates complementary spatial and temporal information. CMFD represents regional dynamic meteorological conditions, while terrain, coordinates, and temporal features provide context on the 1 km target grid. Five timescale-specific models represent distinct short- to long-term accumulation processes. Spatial-block and forward-time results support broad reconstruction skill, whereas stratified uncertainty identifies weaker transfer in sparsely observed western, arid, and high-elevation settings. The decreasing agreement with HSPEI from SPEI-1 to SPEI-24 further indicates that long-memory products are sensitive to differences in forcing, baseline standardization, and reconstruction design, making regional evaluation particularly important for long-timescale applications.

Future development can further enrich the dataset by incorporating denser station observations, newly available high-resolution dynamic environmental variables, and anthropogenic factors such as irrigation and water use. These extensions would build on the current framework and broaden its application to more detailed drought-process and impact studies.

4.2. Potential Applications and Usage Notes

The 1 km monthly multi-timescale SPEI dataset developed in this study can support a wide range of drought-related applications in meteorology, hydrology, agriculture, ecology, natural geography, and remote sensing. High-resolution drought information is important for characterizing drought severity, duration, frequency, and spatial extent at regional and local scales [41]. The multi-timescale structure of SPEI also allows users to examine drought processes with different accumulation windows. SPEI-1 and SPEI-3 are more suitable for short-term dry-wet fluctuation analysis and meteorological drought monitoring, whereas SPEI-6, SPEI-12, and SPEI-24 are more suitable for examining seasonal to long-term drought accumulation, hydrological drought evolution, and persistent water-deficit conditions [42].

The 1 km grid of the dataset makes it useful for regional-scale drought impact assessment and remote-sensing-based environmental analysis. The dataset can be combined with remote sensing indicators such as vegetation indices, land-surface temperature, evapotranspiration, soil moisture, gross primary productivity, and solar-induced chlorophyll fluorescence to investigate vegetation stress, agricultural drought risk, ecosystem vulnerability, and carbon-cycle sensitivity under different drought conditions [5,25,43,44]. Recent studies have shown that water constraints can strongly affect vegetation growth and dryland productivity under climate change [45,46], while SIF provides an important remote-sensing proxy for vegetation photosynthesis and ecosystem functioning [47,48]. Therefore, the dataset can provide a spatial drought background for linking meteorological drought conditions with vegetation productivity, crop stress, water-resource availability, and ecosystem responses.

The dataset can also support agricultural drought assessment and climate-impact studies. Climate extremes, including droughts and heatwaves, have been shown to affect crop yields, food security, and agricultural adaptation needs [49–52]. By providing spatially

continuous monthly SPEI fields on a 1 km grid, the dataset can be used together with crop distribution maps, yield statistics, irrigation information, land-cover products, and socioeconomic data to assess drought exposure and vulnerability in agricultural regions. It may also serve as an auxiliary variable for drought-risk mapping, ecological restoration planning, water-resource management, and regional climate-impact assessment.

SPEI values are standardized and dimensionless: negative values indicate drier-than-normal conditions, positive values indicate wetter-than-normal conditions, and values close to zero indicate near-normal conditions. Starting months differ among timescales because output begins after each accumulation window is complete. The 1 km value denotes target-grid spacing, while the dynamic meteorological forcing is supplied by 0.1° CMFD and fine-grid variation is estimated from station supervision, temporal features, coordinates, and terrain. For local applications, integration with station observations, remote-sensing products, and regional information can strengthen interpretation, especially in western, arid, high-elevation, and sparsely monitored regions. The lower Recall at SPEI ≤ -2.0 also indicates that station and operational information remain important when assessing high-impact extreme events.

5. Conclusions

This study developed a monthly multi-timescale SPEI dataset for mainland China from 1979 to 2018 on a 1 km grid. Monthly supervisory labels were obtained from a consistent source-calendar month-end rule applied to daily observations from 427 stations. Monthly 0.1° CMFD predictors were mapped with a NaN-aware bilinear sampler and combined with timescale-dependent temporal features and NASADEM terrain covariates. Five independent LightGBM models produced SPEI-1, SPEI-3, SPEI-6, SPEI-12, and SPEI-24 fields.

Random five-fold validation yielded mean R^2 values of 0.8807–0.9122. Under the stricter 200 km spatial-block design with a 100 km exclusion buffer, mean R^2 values were 0.7857–0.8739, and forward blocked-period validation yielded 0.7577–0.8337. Product-level consistency was high against Zhang-SPEI for SPEI-3 but declined against HSPEI at longer timescales. Region-specific errors were larger in western arid, high-elevation, and sparse-support settings. Drought-condition validation showed useful moderate-drought detection but substantially lower Recall for SPEI ≤ -2.0, and SHAP analysis confirmed the transition from current-month anomalies at SPEI-1 to long-window memory features at SPEI-12.

Overall, the dataset combines a 1 km output grid, 40-year monthly coverage, five accumulation timescales, station-based supervision, and multi-level validation. By coupling long-term CMFD forcing with fine-grid terrain information, the framework provides spatially detailed drought estimates together with explicit spatial, temporal, and event-severity performance assessments. The resulting open dataset can support drought monitoring and hydroclimatic, ecological, agricultural, remote-sensing, and climate-impact analyses across mainland China.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/rs18213639/s1, Figure S1, grid-cell correlation maps for Zhang-SPEI, GPR-SPEI, and all five HSPEI timescales; Figures S2 and S3, national monthly mean SPEI-3 comparisons with Zhang-SPEI and GPR-SPEI; Figure S4, complex-terrain and arid-transition detail with CMFD-grid continuity diagnostics; Figure S5, national monthly mean comparison with HSPEI at five timescales; Figure S6, spatial-block fold allocation; Figure S7, stratified regional uncertainty; Figure S8, multi-threshold drought-detection heatmap; Tables S1–S8, provincial consistency, drought contingency counts, fold-level spatial and temporal validation, predictor-sensitivity tests, and stratified uncertainty results.

Author Contributions: Conceptualization, H.P., W.L. and H.L.; methodology, H.P. and W.L.; software, H.P.; validation, Y.C. and X.C.; formal analysis, H.P., W.L. and C.W.; investigation, H.P., W.L. and Y.C.; resources, K.Z. and H.L.; data curation, H.P., C.W. and Y.C.; writing—original draft preparation, H.P.; writing—review and editing, W.L., K.Z., H.L. and X.C.; visualization, H.P., C.W. and H.L.; supervision, K.Z. and H.L.; project administration, H.L.; funding acquisition, H.L. and X.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China, grant number 42301336; the Open Research Fund of Henan Provincial Key Laboratory of Hydrosphere and Watershed Water Security, grant number HWWSF202302; the Scientific Research Foundation of Hunan Provincial Education Department, grant number 24C0128; and the National College Students Innovation and Entrepreneurship Training Program, grant number 202510538029.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The dataset generated in this study is openly available in Science Data Bank at https://doi.org/10.57760/sciencedb.34500 (accessed on 19 October 2026). The source code is available at https://github.com/Scorpional/China-mainland-1km-SPEI-dataset-construct (accessed on 19 October 2026).

Acknowledgments: The authors are grateful to all members of the research team for their helpful discussions and support during this study. The authors also acknowledge the providers of the CMFD and NASADEM datasets for making these data publicly available. During the preparation of this manuscript, the authors used ChatGPT 5.4 for language polishing and translation assistance. The authors reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Vicente-Serrano, S.M.; Pena-Angulo, D.; Begueria, S.; Dominguez-Castro, F.; Tomas-Burguera, M.; Noguera, I.; Gimeno-Sotelo, L.; El Kenawy, A. Global drought trends and future projections. *Philos. Trans. A Math. Phys. Eng. Sci.* **2022**, *380*, 20210285. [CrossRef] [PubMed]
2. Tang, T.; Ge, J.; Shi, H.; Wang, L.; Cao, J.; Lee, X. Drought frequency, intensity, and exposure have increased due to historical land use and land cover changes. *Commun. Earth Environ.* **2025**, *6*, 398. [CrossRef]
3. Rodell, M.; Famiglietti, J.S.; Wiese, D.N.; Reager, J.T.; Beaudoing, H.K.; Landerer, F.W.; Lo, M.H. Emerging trends in global freshwater availability. *Nature* **2018**, *557*, 651–659, Correction in *Nature* **2019**, *565*, E7. https://doi.org/10.1038/s41586-018-0831-6. [CrossRef] [PubMed]
4. Samaniego, L.; Thober, S.; Kumar, R.; Wanders, N.; Rakovec, O.; Pan, M.; Zink, M.; Sheffield, J.; Wood, E.F.; Marx, A. Anthropogenic warming exacerbates European soil moisture droughts. *Nat. Clim. Change* **2018**, *8*, 421–426. [CrossRef]
5. Sippel, S.; Reichstein, M.; Ma, X.; Mahecha, M.D.; Lange, H.; Flach, M.; Frank, D. Drought, Heat, and the Carbon Cycle: A Review. *Curr. Clim. Change Rep.* **2018**, *4*, 266–286. [CrossRef]
6. Pizzorni, M.; Innocenti, A.; Tollin, N. Droughts and floods in a changing climate and implications for multi-hazard urban planning: A review. *City Environ. Interact.* **2024**, *24*, 100169. [CrossRef]
7. AghaKouchak, A.; Farahmand, A.; Melton, F.S.; Teixeira, J.; Anderson, M.C.; Wardlow, B.D.; Hain, C.R. Remote sensing of drought: Progress, challenges and opportunities. *Rev. Geophys.* **2015**, *53*, 452–480. [CrossRef]
8. Yuan, X.; Wang, Y.; Ji, P.; Wu, P.; Sheffield, J.; Otkin, J.A. A global transition to flash droughts under climate change. *Science* **2023**, *380*, 187–191. [CrossRef] [PubMed]
9. Zscheischler, J.; Martius, O.; Westra, S.; Bevacqua, E.; Raymond, C.; Horton, R.M.; van den Hurk, B.; AghaKouchak, A.; Jézéquel, A.; Mahecha, M.D.; et al. A typology of compound weather and climate events. *Nat. Rev. Earth Environ.* **2020**, *1*, 333–347. [CrossRef]
10. Beguería, S.; Vicente-Serrano, S.M.; Reig, F.; Latorre, B. Standardized precipitation evapotranspiration index (SPEI) revisited: Parameter fitting, evapotranspiration models, tools, datasets and drought monitoring. *Int. J. Climatol.* **2014**, *34*, 3001–3023. [CrossRef]

11. Hersbach, H.; Bell, B.; Berrisford, P.; Hirahara, S.; Horányi, A.; Muñoz-Sabater, J.; Nicolas, J.; Peubey, C.; Radu, R.; Schepers, D.; et al. The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* **2020**, *146*, 1999–2049. [CrossRef]
12. Muñoz-Sabater, J.; Dutra, E.; Agustí-Panareda, A.; Albergel, C.; Arduini, G.; Balsamo, G.; Boussetta, S.; Choulga, M.; Harrigan, S.; Hersbach, H.; et al. ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. *Earth Syst. Sci. Data* **2021**, *13*, 4349–4383. [CrossRef]
13. Abatzoglou, J.T.; Dobrowski, S.Z.; Parks, S.A.; Hegewisch, K.C. TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015. *Sci. Data* **2018**, *5*, 170191. [CrossRef] [PubMed]
14. Funk, C.; Peterson, P.; Landsfeld, M.; Pedreros, D.; Verdin, J.; Shukla, S.; Husak, G.; Rowland, J.; Harrison, L.; Hoell, A.; et al. The climate hazards infrared precipitation with stations--a new environmental record for monitoring extremes. *Sci. Data* **2015**, *2*, 150066. [CrossRef] [PubMed]
15. Harris, I.; Osborn, T.J.; Jones, P.; Lister, D. Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. *Sci. Data* **2020**, *7*, 109. [CrossRef] [PubMed]
16. Fick, S.E.; Hijmans, R.J. WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* **2017**, *37*, 4302–4315. [CrossRef]
17. Karger, D.N.; Conrad, O.; Bohner, J.; Kawohl, T.; Kreft, H.; Soria-Auza, R.W.; Zimmermann, N.E.; Linder, H.P.; Kessler, M. Climatologies at high resolution for the earth's land surface areas. *Sci. Data* **2017**, *4*, 170122. [CrossRef] [PubMed]
18. Pyarali, K.; Peng, J.; Disse, M.; Tuo, Y. Development and application of high resolution SPEI drought dataset for Central Asia. *Sci. Data* **2022**, *9*, 172. [CrossRef] [PubMed]
19. Wang, Q.; Zeng, J.; Qi, J.; Zhang, X.; Zeng, Y.; Shui, W.; Xu, Z.; Zhang, R.; Wu, X.; Cong, J. A multi-scale daily SPEI dataset for drought characterization at observation stations over mainland China from 1961 to 2018. *Earth Syst. Sci. Data* **2021**, *13*, 331–341. [CrossRef]
20. Zhang, R.; Bento, V.A.; Qi, J.; Xu, F.; Wu, J.; Qiu, J.; Li, J.; Shui, W.; Wang, Q. The first high spatial resolution multi-scale daily SPI and SPEI raster dataset for drought monitoring and evaluating over China from 1979 to 2018. *Big Earth Data* **2023**, *7*, 860–885. [CrossRef]
21. He, Q.; Wang, M.; Liu, K.; Wang, B. High-resolution Standardized Precipitation Evapotranspiration Index (SPEI) reveals trends in drought and vegetation water availability in China. *Geogr. Sustain.* **2025**, *6*, 100228. [CrossRef]
22. Xia, H.; Sha, Y.; Zhao, X.; Jiao, W.; Song, H.; Yang, J.; Zhao, W.; Qin, Y. HSPEI: A 1-km spatial resolution SPEI dataset across the Chinese mainland from 2001 to 2022. *Geosci. Data J.* **2024**, *11*, 479–494. [CrossRef]
23. Yuan, M.; Gan, G.; Bu, J.; Su, Y.; Ma, H.; Liu, X.; Zhang, L.; Zhang, Y.; Gao, Y. A global drought dataset for Multivariate Composite Drought Index (MCDI) and its constituent drought indices. *Sci. Data* **2025**, *12*, 2035. [CrossRef] [PubMed]
24. Park, S.; Im, J.; Park, S.; Rhee, J. Drought monitoring using high resolution soil moisture through multi-sensor satellite data fusion over the Korean Peninsula. *Agric. For. Meteorol.* **2017**, *237–238*, 257–269. [CrossRef]
25. Xu, Z.; Sun, H.; Zhang, T.; Xu, H.; Wu, D.; Gao, J. The high spatial resolution Drought Response Index (HiDRI): An integrated framework for monitoring vegetation drought with remote sensing, deep learning, and spatiotemporal fusion. *Remote Sens. Environ.* **2024**, *312*, 114324. [CrossRef]
26. Stagge, J.H.; Tallaksen, L.M.; Gudmundsson, L.; Van Loon, A.F.; Stahl, K. Candidate Distributions for Climatological Drought Indices (SPI and SPEI). *Int. J. Climatol.* **2015**, *35*, 4027–4040. [CrossRef]
27. He, J.; Yang, K.; Tang, W.; Lu, H.; Qin, J.; Chen, Y.; Li, X. The first high-resolution meteorological forcing dataset for land process studies over China. *Sci. Data* **2020**, *7*, 25. [CrossRef] [PubMed]
28. Poggio, L.; de Sousa, L.M.; Batjes, N.H.; Heuvelink, G.B.M.; Kempen, B.; Ribeiro, E.; Rossiter, D. SoilGrids 2.0: Producing soil information for the globe with quantified spatial uncertainty. *Soil* **2021**, *7*, 217–240. [CrossRef]
29. Rolnick, D.; Donti, P.L.; Kaack, L.H.; Kochanski, K.; Lacoste, A.; Sankaran, K.; Ross, A.S.; Milojevic-Dupont, N.; Jaques, N.; Waldman-Brown, A.; et al. Tackling Climate Change with Machine Learning. *ACM Comput. Surv.* **2022**, *55*, 1–96. [CrossRef]
30. Bauer, P.; Dueben, P.D.; Hoefler, T.; Quintino, T.; Schulthess, T.C.; Wedi, N.P. The digital revolution of Earth-system science. *Nat. Comput. Sci.* **2021**, *1*, 104–113. [CrossRef] [PubMed]
31. Irrgang, C.; Boers, N.; Sonnewald, M.; Barnes, E.A.; Kadow, C.; Staneva, J.; Saynisch-Wagner, J. Towards neural Earth system modelling by integrating artificial intelligence in Earth system science. *Nat. Mach. Intell.* **2021**, *3*, 667–674. [CrossRef]
32. Reichstein, M.; Camps-Valls, G.; Stevens, B.; Jung, M.; Denzler, J.; Carvalhais, N.; Prabhat, F. Deep learning and process understanding for data-driven Earth system science. *Nature* **2019**, *566*, 195–204. [CrossRef] [PubMed]
33. Salcedo-Sanz, S.; Ghamisi, P.; Piles, M.; Werner, M.; Cuadra, L.; Moreno-Martínez, A.; Izquierdo-Verdiguier, E.; Muñoz-Marí, J.; Mosavi, A.; Camps-Valls, G. Machine learning information fusion in Earth observation: A comprehensive review of methods, applications and data sources. *Inf. Fusion* **2020**, *63*, 256–272. [CrossRef]
34. Fu, R.; Chen, R.; Wang, C.; Chen, X.; Gu, H.; Wang, C.; Xu, B.; Liu, G.; Yin, G. Generating High-Resolution and Long-Term SPEI Dataset over Southwest China Through Downscaling EEAD Product by Machine Learning. *Remote Sens.* **2022**, *14*, 1662. [CrossRef]

35. Behfar, N.; Sharghi, E.; Nourani, V.; Booij, M.J. Drought index downscaling using AI-based ensemble technique and satellite data. *Theor. Appl. Climatol.* **2024**, *155*, 2379–2397. [CrossRef]
36. Kadow, C.; Hall, D.M.; Ulbrich, U. Artificial intelligence reconstructs missing climate information. *Nat. Geosci.* **2020**, *13*, 408–413. [CrossRef]
37. Cheng, Y.; An, Q.; Liu, L.; Li, H.; Huang, G. Spatially distinct drought patterns and influencing factors across China: A machine learning approach with a comprehensive index. *Ecol. Indic.* **2025**, *179*, 114170. [CrossRef]
38. NASA JPL. *NASADEM Merged DEM Global 1 arc Second V001*; NASA Land Processes Distributed Active Archive Center: Sioux Falls, SD, USA, 2020. [CrossRef]
39. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017; Volume 30, pp. 3146–3154.
40. Vicente-Serrano, S.M.; Beguería, S.; López-Moreno, J.I. A Multiscalar Drought Index Sensitive to Global Warming: The Standardized Precipitation Evapotranspiration Index. *J. Clim.* **2010**, *23*, 1696–1718. [CrossRef]
41. Gebrechorkos, S.H.; Peng, J.; Dyer, E.; Miralles, D.G.; Vicente-Serrano, S.M.; Funk, C.; Beck, H.E.; Asfaw, D.T.; Singer, M.B.; Dadson, S.J. Global high-resolution drought indices for 1981–2022. *Earth Syst. Sci. Data* **2023**, *15*, 5449–5466. [CrossRef]
42. Tyagi, S.; Zhang, X.; Saraswat, D.; Sahany, S.; Mishra, S.K.; Niyogi, D. Flash Drought: Review of Concept, Prediction and the Potential for Machine Learning, Deep Learning Methods. *Earths Future* **2022**, *10*, e2022EF002723. [CrossRef]
43. Piao, S.; Wang, X.; Park, T.; Chen, C.; Lian, X.; He, Y.; Bjerke, J.W.; Chen, A.; Ciais, P.; Tømmervik, H.; et al. Characteristics, drivers and feedbacks of global greening. *Nat. Rev. Earth Environ.* **2019**, *1*, 14–27. [CrossRef]
44. Schwalm, C.R.; Anderegg, W.R.L.; Michalak, A.M.; Fisher, J.B.; Biondi, F.; Koch, G.; Litvak, M.; Ogle, K.; Shaw, J.D.; Wolf, A.; et al. Global patterns of drought recovery. *Nature* **2017**, *548*, 202–205. [CrossRef] [PubMed]
45. Jiao, W.; Wang, L.; Smith, W.K.; Chang, Q.; Wang, H.; D'Odorico, P. Observed increasing water constraint on vegetation growth over the last three decades. *Nat. Commun.* **2021**, *12*, 3777. [CrossRef] [PubMed]
46. Wang, L.; Jiao, W.; MacBean, N.; Rulli, M.C.; Manzoni, S.; Vico, G.; D'Odorico, P. Dryland productivity under a changing climate. *Nat. Clim. Change* **2022**, *12*, 981–994. [CrossRef]
47. Du, Y.; Du, S.; Zhao, D.; Liu, X.; Liu, L. QeRSIF: A quality-enhanced red solar-induced chlorophyll fluorescence dataset for TROPOMI satellite. *GISci. Remote Sens.* **2026**, *63*, 2632999. [CrossRef]
48. Porcar-Castell, A.; Malenovsky, Z.; Magney, T.; Van Wittenberghe, S.; Fernandez-Marin, B.; Maignan, F.; Zhang, Y.; Maseyk, K.; Atherton, J.; Albert, L.P.; et al. Chlorophyll a fluorescence illuminates a path connecting plant molecular biology to Earth-system science. *Nat. Plants* **2021**, *7*, 998–1009. [CrossRef] [PubMed]
49. Hasegawa, T.; Sakurai, G.; Fujimori, S.; Takahashi, K.; Hijioka, Y.; Masui, T. Extreme climate events increase risk of global food insecurity and adaptation needs. *Nat. Food* **2021**, *2*, 587–595. [CrossRef] [PubMed]
50. Kornhuber, K.; Lesk, C.; Schleussner, C.F.; Jagermeyr, J.; Pfleiderer, P.; Horton, R.M. Risks of synchronized low yields are underestimated in climate and crop model projections. *Nat. Commun.* **2023**, *14*, 3528. [CrossRef] [PubMed]
51. Kummu, M.; Heino, M.; Taka, M.; Varis, O.; Viviroli, D. Climate change risks pushing one-third of global food production outside the safe climatic space. *One Earth* **2021**, *4*, 720–729. [CrossRef] [PubMed]
52. Lesk, C.; Rowhani, P.; Ramankutty, N. Influence of extreme weather disasters on global crop production. *Nature* **2016**, *529*, 84–87. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.