www.nature.com/scientificdata

scientific data

OPEN

DATA DESCRIPTOR

PDSI_CMIP6: an ensemble CMIP6-projected self-calibrated Palmer drought severity index dataset

Jinghua Xiong & Yuting Yang

Drought, characterized by below-average water supply, profoundly affects regional water resources and various ecosystem services. The Palmer Drought Severity Index (PDSI) is a widely used metric for drought monitoring and climate change assessments but suffers from inherent climatic inconsistencies and lacks comprehensive and reliable estimates under changing climate conditions. Here we develop a monthly multi-model and multi-scenario dataset of self-calibrated PDSI for the period 1850–2094, derived from 11 climate model outputs within the Coupled Model Intercomparison Project 6 (PDSI_CMIP6). The traditional two-layer bucket model in PDSI is replaced with direct hydrological outputs from CMIP6 models, ensuring alignment with CMIP6 projections. The PDSI estimates are validated against soil moisture simulations through correlation and regression analysis. Application of the dataset reveals pronounced spatial heterogeneity in long-term drought trends across continents, with limited global-mean change but notable regional intensification under climate change. This dataset provides uncertainty-constrained quantifications of terrestrial moisture conditions in a changing climate, faithfully reflecting CMIP6-projected hydrological changes.

Background & Summary

Hydrological extremes such as droughts and pluvial floods stem from abnormal climatic variability[1–3], leading to substantial economic and ecological losses and impacting the sustainability of natural and social systems[4]. Improved understanding of hydrological extremes is thus crucial for adapting to water-related hazards and managing water resources efficiently[5]. Extensive research has detected changes in hydrological extremes as responses to global climate change[6], primarily driven by increasing CO_2 concentrations and associated consequences, including rising air temperatures and intensifying precipitation extremes[7,8]. These changes have induced variable hydrological system responses, such as increased floods from heavy rainfall and decreased floods related to snowmelt[9]. Despite an overall small global flood change, the contradictory responses were also observed in historical drought changes[10]. Furthermore, the Coupled Model Intercomparison Project Phase 6 (CMIP6) reported that a larger fraction of land areas will likely experience increased river floods and agricultural and ecological droughts than decreased[11–13]. These projections necessitate improved assessments of hydrological extremes under climate change.

Various metrics have been proposed to track the spatiotemporal evolution of hydrological extremes, including Standardized Precipitation Evapotranspiration Index (SPEI) and Standardized Precipitation Index (SPI)[14–17]. While widely used, these indices rely on simplified representations of hydro-meteorological processes, often neglecting underlying land surface conditions and limiting their ability to capture the complexity of hydrological dynamics under climate change. The Palmer Drought Severity Index (PDSI) and its variants (e.g., self-calibrated PDSI) are prominent for quantifying hydrological extremes[18], owing to their incorporation of both antecedent and current water inputs (precipitation, P) and outputs (potential evaporation, E_P) through a two-layer bucket water balance model. This structure has enabled PDSI to outperform simpler statistical metrics (e.g., SPEI) in monitoring hydrological and agricultural droughts[19]. However, recent critiques highlighted its inadequate representation of hydrological processes, pramarily the offline hydrological framework and over-simplified hydrological parameterization[20]. Specifically, the runoff generation is abstracted as a simple saturation-excess mechanism, without consideration of several crucial processes, including snow/ice dynamics, vegetation dynamics, and plant responses to elevated CO_2[21]. The model inadequacy can substantially impact the representation of

State Key Laboratory of Hydroscience and Engineering, Department of Hydraulic Engineering, Tsinghua University, Beijing, China. ✉e-mail: yuting_yang@tsinghua.edu.cn

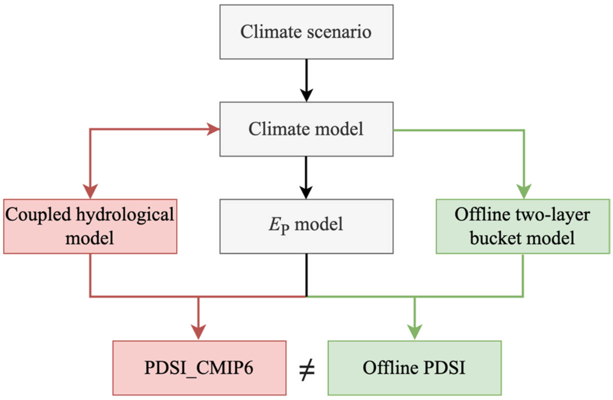

Fig. 1 Conceptual comparisons between PDSI_CMIP6 (pink elements) and offline PDSI (green elements) stuctures. The gray color means the shared components.

hydrological processes under climate change[22]. In this case, improved accuracy of PDSI has been achieved when substituting the original model with state-of-the-art simulations from global climate or hydrological models[21,23]. In addition, the PDSI estimates directly derived from inherent hydrological modules within climate models, which maintain internal consistency with climate forcing, produce more consistent projections with the underlying climate than the traditional offline estimates, as conceptualized in Fig. 1[23]. Moreover, previous studies have also indicated an upgraded performance of offline PDSI, when considering the plant physiological response to elevated atmospheric CO_2 concentration within the Penman-Monteith E_{P} algorithm[22]. Nevertheless, the influence of different E_{P} inputs on PDSI remains unknown when using advanced hydrological simulations.

Several global datasets of PDSI has been developed using the conventional offline approach to facilitate various applications[24,25]. These datasets have provided unprecedented opportunities to understand the historical changes in hydrological extremes. In addition, a few studies have calculated future PDSI with projected meteorological forcing from climate models[26,27]. However, as mentioned above, these datasets prevent the derived PDSI from faithfully capturing the hydrological extremes corresponding to the underlying climate conditions[e.g. 10,23]. Moreover, existing datasets generally overlook the influence of model uncertainties and inter-scenario variability on future drought projections. These gaps underscore the urgent need for a physically consistent, uncertainty-aware, and multi-scenario global PDSI dataset.

To this end, we present an improved monthly PDSI dataset (PDSI_CMIP6), derived directly from ensemble CMIP6 climate model simulations. This dataset explicitly couples climatic and hydrological anomalies—an aspect commonly ignored in traditional offline frameworks (Fig. 1). PDSI_CMIP6 spans the historical period from 1850 to 2014 and includes multiple radiative forcing levels of Shared Socioeconomic Pathways (SSPs) from 2015 to 2094. The dataset is available online in a common NetCDF format with a consistent spatial resolution of 1° over global land. Detailed methodologies used to produce PDSI_CMIP6 are provided in the Methods section. The Data Records section explains each data record associated with the dataset. The sensitivity analysis of PDSI_CMIP6 to different algorithms of E_{P} is presented in the Technical Validation section, along with additional comparisons with soil moisture anomalies. Illustrative applications along with important usage notes are discussed in the Usage Notes section. Overall, PDSI_CMIP6 offers a physically consistent and scenario-resolved quantification of global terrestrial drought and wetness dynamics in a changing climate.

Methods

The main procedures for producing PDSI_CMIP6 are outlined in Fig. 2. We obtained monthly meteorological and hydrological outputs from 11 climate models in CMIP6 under various emission scenarios. We utilized the self-calibrated PDSI algorithm, which presented superior performance in spatial comparabilities than conventional PDSI[28,29]. All calculations were conducted on a grid scale for each model individually, after which the results were integrated into an ensemble. The calculated PDSI values were compared with soil moisture simulations for validation. Finally, a spatiotemporal analysis of the derived PDSI under climate change was conducted as an illustrative example, revealing the hydrological response to future warming. Further details about the methods and data are provided in the following sections.

CMIP6 simulations.

Multiple variables were directly derived from CMIP6 climate model outputs to facilitate the calculations of PDSI, containing six hydrological variables and six meteorological variables. The six hydrological variables include evaporation (variable name: evspsbl), latent heat flux (variable name: hfls), sensible heat flux (variable name: hfss), runoff (variable name: mrro), soil moisture (variable name: mrso), and surface soil moisture (variable name: mrsos). Additionally, the six meteorological variables include precipitation (variable name: pr), relative humidity (variable name: hurs), air temperature (variable name: tas), wind speed (variable name: sfcWind), air pressure (variable name: ps), and extraterrestrial solar radiation (variable name: rsdt)[30]. These variables were retrieved monthly for the historical period (1850–2014) and projected future period (2015–2094) under four climate change scenarios: SSP1-2.6, SSP2-4.5, SSP3-7.0, and SSP5-8.5, representing different socio-economic pathways and emissions trajectories. Specifically, SSP1-2.6, SSP2-4.5, SSP3-7.0, and SSP5-8.5 correspond to sustainable development, middle-of-the-road development, regional rivalry, and fossil-fueled

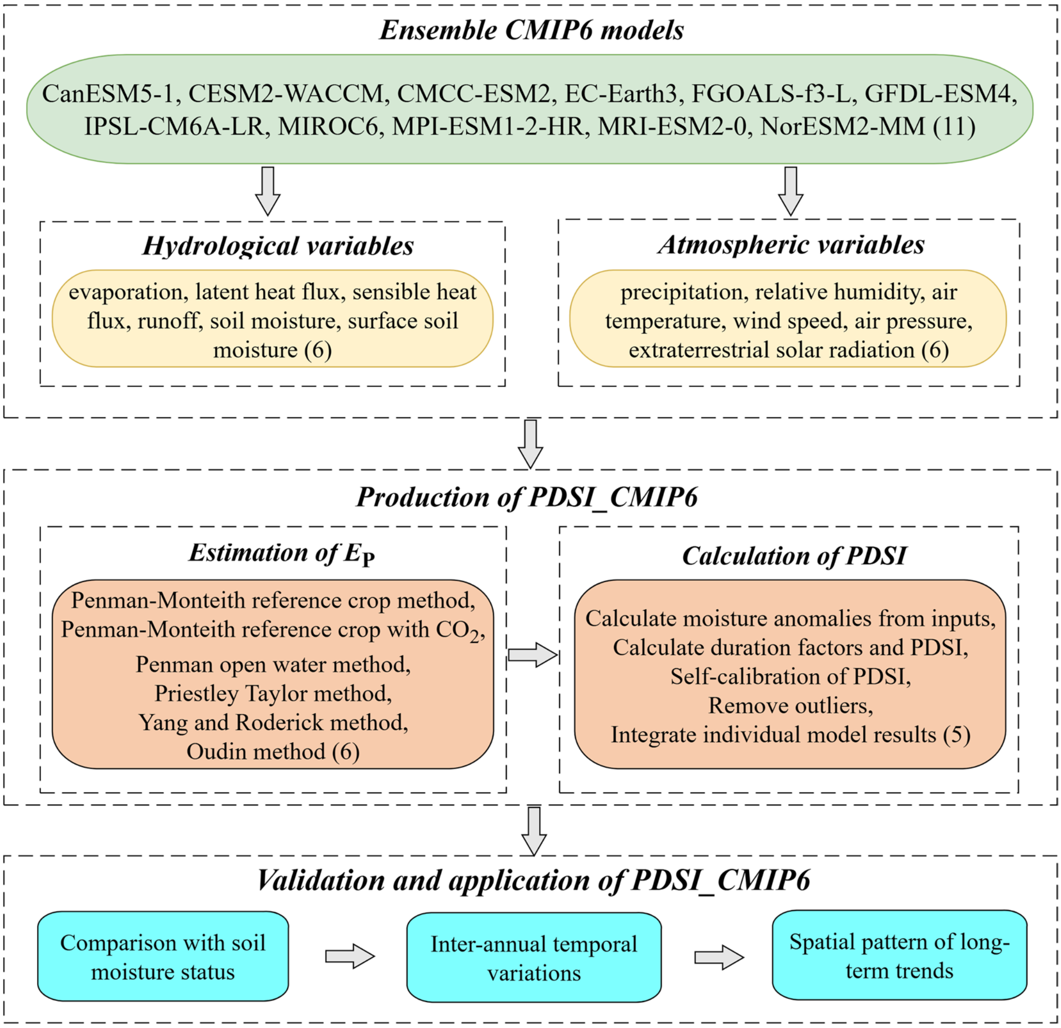

Fig. 2 Schematic of the generation of PDSI_CMIP6. Values within the brackets represent the number of datasets, variables, or methods.

development, respectively, with projected radiative forcing levels of 2.6, 4.5, 7.0, and 8.5 W/m^2 by the year 2100[31]. The selection of models, study periods, and future scenarios is intended to maximize data availability. Eleven models were selected, as they are the only ones providing the required 12 variables from 1850 to 2094 under all scenarios (Table 1). These models are managed by different institutions and have varying spatial resolutions ranging from 0.7° (EC-Earth3) to 2.8° (CanESM5-1). For models from the same institution, only the one with the highest spatial resolution (e.g., MPI-ESM1-2-HR vs. MPI-ESM1-2-LR), the most advanced representation (e.g., CMCC-ESM2 vs. CMCC-CM2-SR5), or the standard configuration (e.g., EC-Earth3 vs. EC-Earth3-Veg) is retained to avoid bias within the ensemble. The model outputs were interpolated to a consistent 1° resolution using the first-order conservative remapping method to embed the results[32].

Quantification of EP. To test how E_P estimates affect the PDSI calculation, we estimated E_P using six different approaches, including the Penman-Monteith reference crop (PM-RC) equation[33], the Penman-Monteith reference crop considering the plant response to CO_2 changes (PM-RC-CO_2)[22], the open water Penman equation (Penman-OW)[34], the Priestley-Taylor model (PT)[35], the Yang and Roderick model (YR)[36], and the Oudin model[37].

The PM-RC is the standard approach for estimating water consumption from reference crop under non-water-limited condition recommended by the Food and Agriculture Organization of the United Nations, which is expressed as:

$$E_P = \frac{\Delta \cdot (R_n - G)/\lambda + \gamma \cdot \frac{900}{T + 273} \cdot u_2 \cdot (e_s - e_a)}{\Delta + \gamma \cdot (1 + 0.34 \cdot u_2)} \tag{1}$$

where Δ is the slope of the saturation vapor pressure curve (kPa·°C^{-1}), γ is a psychrometric constant (kPa·°C^{-1}), u_2 (m·s^{-1}) is the wind speed at the 2-meter level above the ground, e_s and e_a are the saturation and actual vapor

ID	Model name	Model center	Raw resolution (Lon × Lat)
1	CanESM5-1	Canadian Centre for Climate Modelling and Analysis, Canada	128 × 64
2	CESM2-WACCM	National Center for Atmospheric Research, United States	288 × 192
3	CMCC-ESM2	Centro Euro-Mediterraneo sui Cambiamenti Climatici, Italy	288 × 192
4	EC-Earth3	European research consortium EC-Earth	512 × 256
5	FGOALS-f3-L	Institute of Atmospheric Physics, Chinese Academy of Sciences, China	288 × 180
6	GFDL-ESM4	NOAA Geophysical Fluid Dynamics Laboratory, United States	288 × 180
7	IPSL-CM6A-LR	Institut Pierre-Simon Laplace, France	144 × 143
8	MIROC6	National Institute for Environmental Studies, Japan	256 × 128
9	MPI-ESM1-2-HR	Max Planck Institute for Meteorology, Germany	384 × 192
10	MRI-ESM2-0	Meteorological Research Institute, Japan	320 × 160
11	NorESM2-MM	Norwegian Climate Centre, Norway	288 × 192

Table 1. Basic information of the climate models included in the PDSI_CMIP6.

pressure (kPa), respectively. R_n is the net radiation at the crop surface ($MJ \cdot m^{-2} \cdot d^{-1}$) and G is soil heat flux density. The difference between R_n and G (i.e., $R_n - G$) is equivalent to the sum of the latent and sensible heat fluxes. λ is the latent heat of vaporization ($MJ \cdot kg^{-1}$).

However, a recent study suggested that the neglected plant response to elevated CO_2 would cause overrated E_P from the original PM-RC algorithm, resulting in inaccurate PDSI estimation. A new variant of the PM-RC method has been proposed to account for the plant response of CO_2 changes (i.e., PM-RC-CO_2)[22], which is also applied for the E_P estimation of PDSI_CMIP6:

$$E_P = \frac{\Delta \cdot (R_n - G)/\lambda + \gamma \cdot \frac{900}{T+273} \cdot u_2 \cdot (e_s - e_a)}{\Delta + \gamma \cdot \{1 + u_2 \cdot [0.34 + 2.4 \cdot 10^{-4} \cdot ([Ca] - 300)]\}} \tag{2}$$

where [Ca] is the concentration of the atmosphere CO_2 (ppm).

The Penman-OW model estimates E_P as[34]:

$$E_P = \frac{\Delta \cdot (R_n - G) + \gamma \cdot 6.43 \cdot (1 + 0.536 \cdot u_2) \cdot (e_s - e_a)}{\lambda \cdot (\Delta + \gamma)} \tag{3}$$

Different from the Penman-type methods, the PT method does not provide an explicit expression of the aerodynamic term, and calculates E_P as[35]:

$$E_P = \alpha \cdot (R_n - G) \cdot \frac{\Delta}{\lambda \cdot (\Delta + \gamma)} \tag{4}$$

where α is the Priestley-Taylor coefficient, and is taken as a fixed value of 1.26 (Priestley-Taylor coefficient).

The YR method adpots a similar form with PT but an improved parameterization of the Bowen ratio[36]:

$$E_P = \frac{\Delta}{\lambda \cdot (\Delta + \beta \cdot \gamma)} \cdot (R_n - G) \tag{5}$$

where β is a dimensionless coefficient of 0.24.

The Oudin model is a tempearture-based E_P model and is expressed as[37]:

$$E_P = \frac{R_e}{\lambda \cdot \rho} \cdot \frac{T + 5}{100}, \text{ if } T + 5 > 0 \tag{6.1}$$

$$E_P = 0, \text{ if } T + 5 \leq 0 \tag{6.2}$$

where R_e is the extraterrestrial radiation ($MJ \cdot m^{-2} \cdot d^{-1}$), and ρ is the density of water ($kg \cdot m^{-3}$).

Estimation of PDSI. The original PDSI relies solely on precipitation and E_P to simulate regional water balance using a simple two-layer bucket water balance model, which estimates evaporation (E), recharge to soils (R), runoff (RO), and water loss to the soil layers (L)[18]. These modeled hydrological variables are then used to derive a standardized index (i.e., PDSI) that measures the departure of the current hydrological status from its climatological mean. However, the original two-layer bucket model has been criticized for being oversimplified, lacking representations of many processes such as snow/frozen soil and vegetation that affect hydrological dynamics[22]. More importantly, the estimated hydrological variables from this two-layer bucket model differ from those estimated by climate models, leading to internal inconsistencies between the projected PDSI and climate models. To address these issues, we disregarded the hydrological calculations embedded in the original PDSI model and instead relied on direct hydrological outputs from climate models for the calculation of PDSI[23].

Potential values for different water balance fluxes, including potential recharge to soils (R_{P}), potential runoff (RO_{P}), and potential water loss to the soil layers (L_{P}), are defined following E_{P} to measure the deviation from the climatology. The available water capacity (AWC), considered as the soil moisture maximas in different models, is required to describe soil dynamics. The climatically appropriate for existing conditions (CAFEC) values are subsequently defined as the product of the potential values and corresponding climatology coefficients, denoted as α:

$$\alpha_{\mathrm{X}} = \frac{\overline{X}}{\overline{X}_{\mathrm{P}}} \tag{7}$$

where $\overline{X}$ and $\overline{X}_{\mathrm{P}}$ represents the long-term average of actual and potential values of various water balance components. In particular, the CAFEC precipitation ($\hat{P}$) represents the amount of precipitation needed to maintain a normal soil moisture level:

$$\hat{P} = \alpha_{\mathrm{E}} \cdot E_{\mathrm{P}} + \alpha_{\mathrm{R}} \cdot R_{\mathrm{P}} + \alpha_{\mathrm{RO}} \cdot RO_{\mathrm{P}} - \alpha_{\mathrm{L}} \cdot L_{\mathrm{P}} \tag{8}$$

The difference between actual precipitation and $\hat{P}$ is given as the moisture departure (D):

$$D = P - \hat{P} \tag{9}$$

A monthly climatic characteristic coefficient ($K_i^{'}$) is applied to consider different meanings of D_{i} in different regions:

$$K_i = 1.5 \cdot \log_{10} \cdot \left(\frac{\frac{\overline{E_{\mathrm{P}_i}} + \overline{R}_i + \overline{RO_i}}{\overline{P}_i + \overline{L}_i} + 2.8}{\overline{D}_i} \right) + 0.5 \tag{10.1}$$

$$K_i^{'} = \frac{17.67}{\sum_{i=1}^{12} \overline{D}_i \cdot K_i} \cdot K_i \tag{10.2}$$

where the subscript i denotes the status in the month i (from 1 to 12). The PDSI index can be subsequently calculated as:

$$Z_{i,j} = D_{i,j} \cdot K_i^{'} \tag{11.1}$$

$$PDSI_{i,j} = p \cdot PDSI_{i-1,j} + q \cdot Z_{i,j} \tag{11.2}$$

where the subscript j denotes the status in the year j from 1850 to 2094, with the whole period used for coefficients calibration. p and q are the duration coefficients to determine the sensitivity of PDSI to moisture anomaly ($Z_{i,j}$) and its auto-correlation. They are considered as 0.897 (p) and 1/3 (q) based on linear slopes between the length and severity of the most extreme droughts observed in Kansas and Iowa, United States[38]. However, the application of traditional PDSI is constrained by limited spatial comparability, which has been improved via a self-calibration procedure[28,29]. It re-fits a linear regression model using the original PDSI scores and derives a new set of parameters p and q for each grid cell, considering the spatial divergence of climate background in different regions. We additionally capped the grids with values higher (lower) than 10 (-10) to reduce the outlier effects on spatial integration of PDSI. As shown in Table 2, such outliers are considered to rarely happen.

All calculations of PDSI and E_{P} were carried out on each 1° grid cell with valid AWC values. Greenland and Antarctica are excluded from the calculations due to the limited availability of hydrological simulations (e.g., soil moisture). A land mask map was generated for each climate model and the ensemble mean, using grid cells with a land percentage greater than 80% based on the variable sftlf (Fig. 3). This land mask defines the grid cells used for PDSI computation, which are then allocated for further analysis.

Validation and application of PDSI_CMIP6.

PDSI_CMIP6 was validated against soil moisture simulations using Pearson correlation and linear regression analysis to assess its internal consistency with hydrological components within CMIP6 Earth System Models—relationships not captured in traditional offline PDSI estimates[39]. To guide potential users, we also provide illustrative examples demonstrating practical applications of PDSI_CMIP6. Temporal changes in PDSI_CMIP6 across different continents and globally were investigated using the area-weighted method to extract data from the gridded product (see Figure S1 for geographical classification)[40]. Long-term linear trend was estimated using the least square method[41] and spatial distribution was presented as the ensemble mean of PDSI_CMIP6. All calculations were performed using the Climate Data Toolbox for MATLAB[42]. Note the spatiotemporal analysis was illustrated based on the SSP5-8.5 scenario, with results for other scenarios discussed in the supplementary file.

PDSI range	Class
$[4,+\infty)$	Extreme moist event
$[3,4)$	Severe moist event
$[2,3)$	Moderate moist event
$[1,2)$	Mild moist event
$[0.5,1)$	Incipient moist spell
$(-0.5,0.5)$	Near normal
$(-1,-0.5]$	Incipient dry spell
$(-2,-1]$	Mild drought event
$(-3,-2]$	Moderate drought event
$(-4,-3]$	Severe drought event
$(-\infty,-4]$	Extreme drought event

Table 2. Classification of PDSI[24].

Data Records

The PDSI_CMIP6 dataset provides NetCDF files containing the estimated PDSI from each CMIP6 model (Table 1)[43]. The monthly dataset spans from 1850 to 2014 for historical period and from 2015 to 2094 under four climate change scenarios, with a re-gridded spatial resolution of 1°. Each NetCDF file is linked with a land mask map to facilitate the extraction of specific study areas (Fig. 3). The files also contain essential metadata, including the date of generation, contact information for the authors, and variable attributes (Table 1).

Technical Validation

Sensitivity to E_P estimates. Using the outputs from the CanESM5-1 model as an example, we first tested the sensitivity of calculated PDSI_CMIP6 to different E_P estimates. We observed high agreements between PDSI_CMIP6 based on PM-RC-CO_2 E_P and the ensemble mean of various E_P estimates, with a fitted linear slope of 0.97 ($R^2 = 0.95$) under the SSP5-8.5 scenario (Fig. 4f). Notably, the highest consistency was detected with the PM-RC method, indicating limited effects of CO_2 concentration on PDSI_CMIP6 (Fig. 4a). Comparisons with the other four E_P methodologies also demonstrated the insensitivity of PDSI_CMIP6 to different E_P inputs (Fig. 4b–e). This is because E_P does not influence the estimation of actual hydrological variables, when directly using hydrological outputs from climate models. Although E_P directly determines the climatology coefficients and subsequent calculations (Eq. 7), limited impacts on PDSI_CMIP6 are observed. Similar consistencies between different E_P algorithms were found in other climate change scenarios (Figures S2–S4). In the following, the PM-RC-CO_2 algorithm is used to generate the PDSI_CMIP6 dataset.

Comparison with soil moisture. To evaluate the internal consistency of PDSI_CMIP6 with CMIP6-simulated hydrological conditions, we quantified correlations between PDSI_CMIP6 and model-derived soil moisture. This assessment addresses a key limitation of previous offline PDSI estimates, which often fail to align with the hydrological outputs of their driving models. Ensemble results indicate that 96% of the land grid cells exhibit positive correlations under the SSP5-8.5 scenario on an annual scale, with an average correlation level of 0.64 (Fig. 5l). Individual analyses of different models reveal relatively weaker correlations in CMCC-ESM2, FGOALS-f3-L, and MRI-ESM2-0, primarily due to negative correlations over extremely arid regions like the Sahara Desert and northwestern China (Figures S5–S8). These discrepancies may be attributed to inaccurate hydrological simulations or biased duration coefficients in the PDSI algorithm. Evaluations for other climate change scenarios show similar patterns to SSP5-8.5, with 97% (SSP1-2.6), 96% (SSP2-4.5), and 96% (SSP3-7.0) of the grid cells exhibiting positive correlations, with average correlation levels of 0.64 (Figures S9–S11). Overall, the strong correlations between PDSI_CMIP6 and soil moisture demonstrate the reasonable climatic-hydrological consistency of PDSI_CMIP6 under various scenarios.

Usage Notes

This section presents illustrative examples demonstrating the potential applications of PDSI_CMIP6, including analyses of spatiotemporal trends and continental-scale assessments under climate change. Dataset limitations are also discussed, along with proposed directions for future improvements..

Temporal trends of PDSI_CMIP6. Spatial integration was performed across different continents using a weighted area method to extract large-scale characteristics of PDSI_CMIP6 (Fig. 6). While there is only a very minor change on a global scale (Fig. 6g), significant trends have been identified at the continental level. Specifically, increasing PDSI_CMIP6 (wetting) trends are observed in Africa (Fig. 6a), Asia (Fig. 6b), and North America (Fig. 6e), whereas decreasing (drying) trends are evident in Australia (Fig. 6c) and South America (Fig. 6f) under the SSP5-8.5 scenario. These trends are predominantly driven by anthropogenic climate change and are consistent among various emission scenarios from SSP1-2.6 to SSP5-8.5 (Figs. 6 and S12–S14). Specifically, wetting trends are largely attributed to increases in precipitation, corroborated by corresponding changes in actual evaporation, runoff, and soil moisture anomalies (Figures S15–S18). In contrast, drying trends reflect either precipitation deficits (e.g., Australia) or persistent soil moisture depletion (e.g., South America). However, it is important to note the large discrepancies between model projections within CMIP6, particularly for Australia. Therefore, caution should be exercised when making quantitative assessments based on individual climate model outputs.

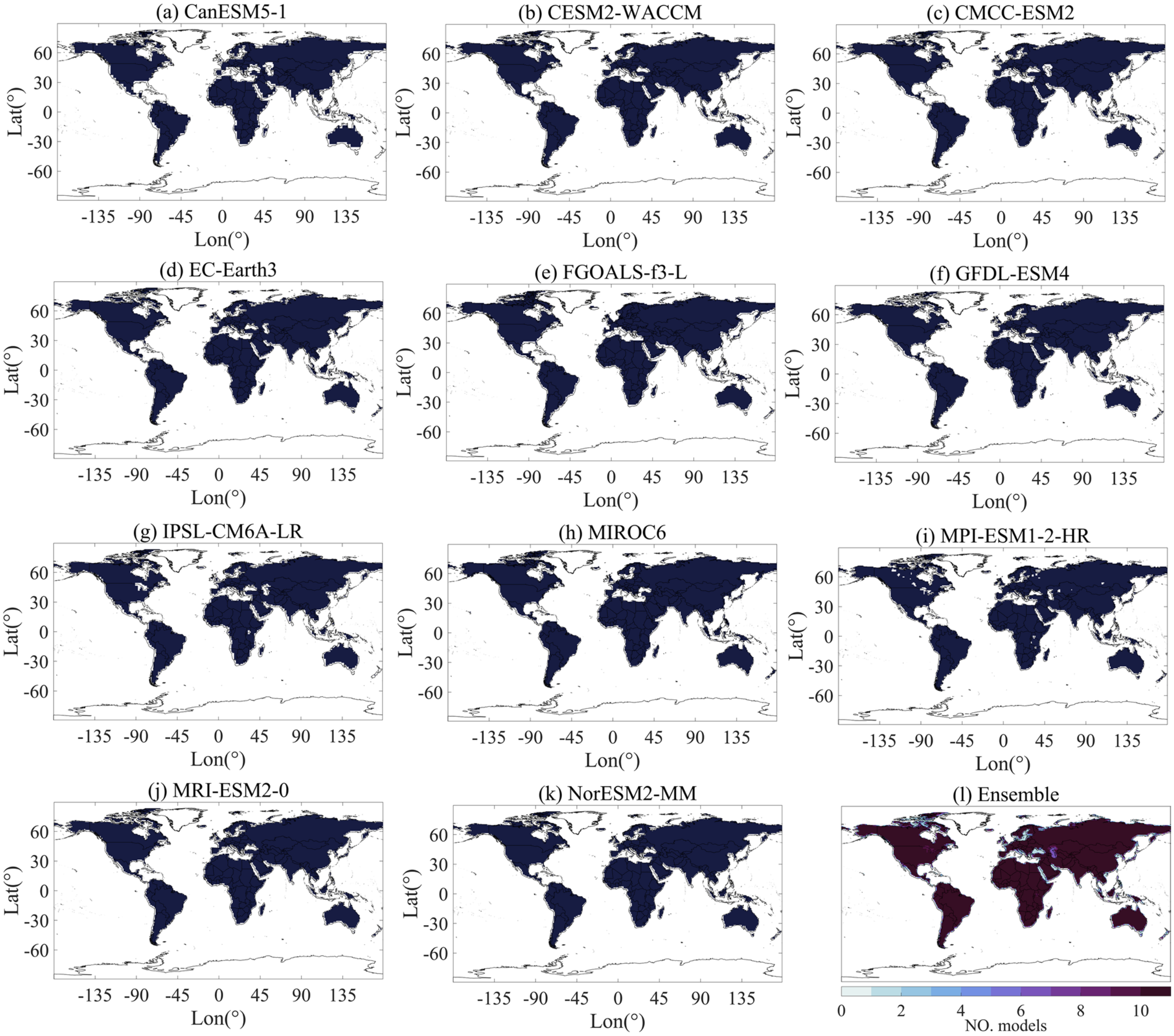

Fig. 3 Land mask of PDSI_CMIP6 for each selected climate models and their ensemble mean. Ensemble results (l) show the number of the models with available PDSI estimates.

At the local scale, significant drying trends are evident over the West United States, Mexico, Brazil and Chile, the Mediterranean region, South China, and several countries in the southern and northern of Africa (Fig. 7). On the contrary, widespread wetting is discovered in India and Southwest China, the southeastern part of South America, central Africa, and almost all the regions over 50°N. The divergent responses of PDSI_CMIP6 to climate change indicate the strong spatial variability, which persistently exists in future projections under various scenarios (Figures S19–S21). Additional comparison with long-term soil moisture anomalies reveal consistent spatial patterns acorss scenarios, indicating the dominanting role of soil moisture changes in PDSI variations (Figures S22–S25). Regression analysis further support this relationship, with moderate-to-strong agreement between PDSI and soil moisture trends, yielding R^2 values of 0.54, 0.58, 0.62, and 0.63 under SSP1-2.6, SSP2-4.5, SSP3-7.0, and SSP5-8.5 scenarios, respectively (Figures S26–S29).

Uncertainty of PDSI_CMIP6. Despite the significant potential of PDSI_CMIP6 for assessing changes in global hydrological extremes, several important considerations should be noted when using the dataset:

(i) As the first global PDSI dataset derived directly from CMIP6 simulations, PDSI_CMIP6 enables a reassessment of hydrological extremes embedded within the latest generation of climate projections. The ensemble framework provides valuable insights into inter-model variability in drought responses to global warming. However, the accuracy of PDSI_CMIP6 is inherently dependent on the quality of the underlying CMIP6 climate variables, which are subject to uncertainties arising from both model structure and emissions scenarios. Nevertheless, the dataset offers a valuable tool for benchmarking and evaluating reanalysis products and large-scale model ensembles, such as those in the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP).

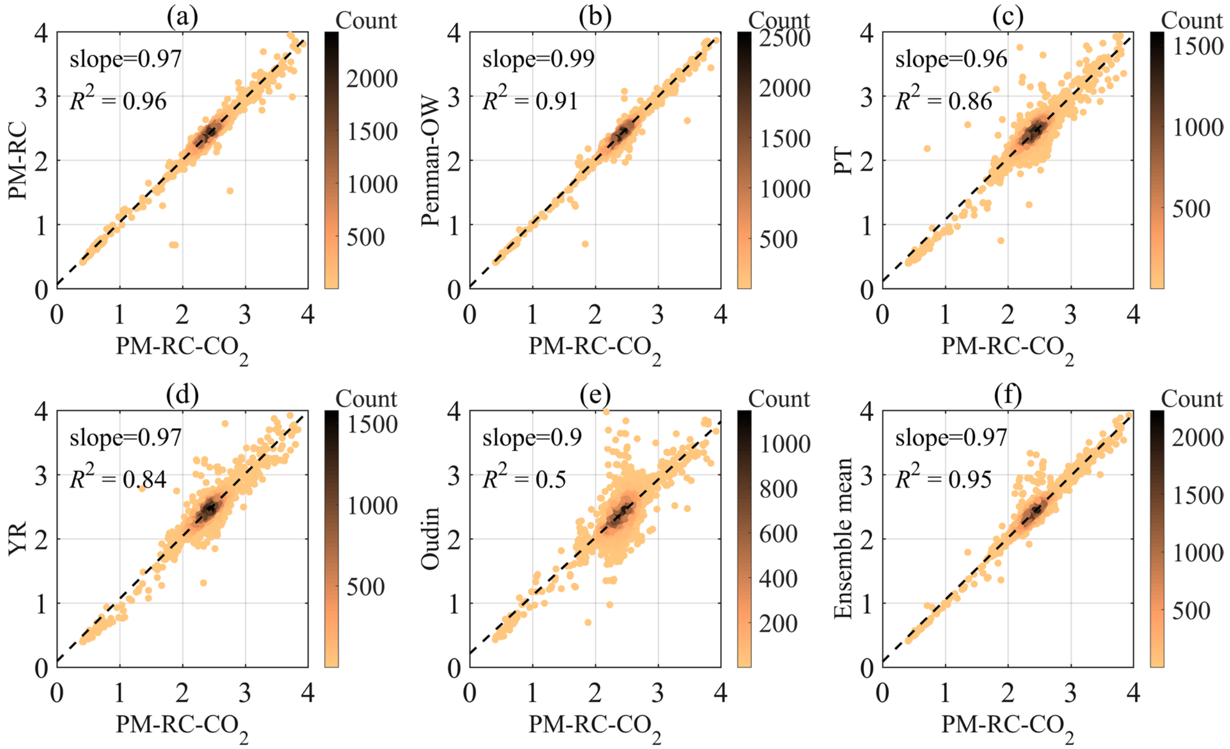

Fig. 4 Scatter plot of the standard deviation of PDSI derived from different E_P algorithms taking CanESM5-1 as an example under the SSP5-8.5 scenario. The inserted dash line represents the fitted linear regression model.

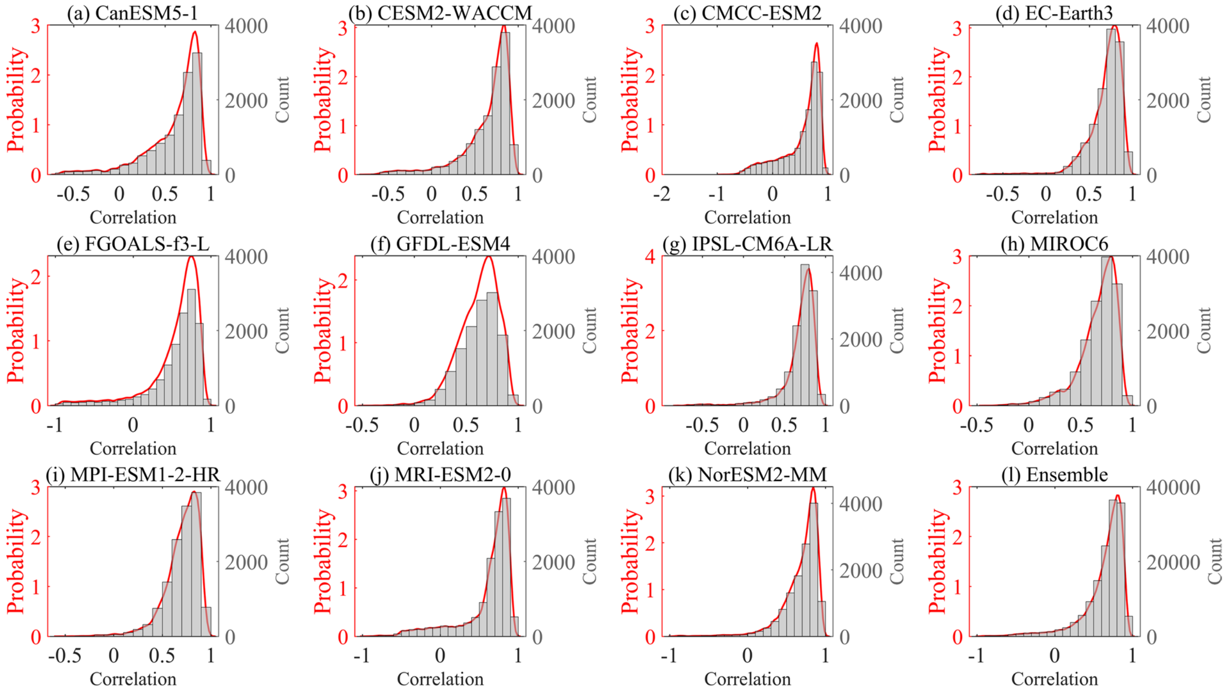

Fig. 5 Empirical distributions of the probability density (red line, left axis) and histogram (gray bars, right axis) of the correlation coefficients between PDSI_CMIP6 and soil moisture of different models under the SSP5-8.5 scenario.

(ii) While self-calibration enhances the spatial comparability of traditional PDSI estimates, it can occasionally produce extreme or unrealistic values (e.g., >10 or < -10), often resulting from imbalances in the water budget of individual models. These anomalous values have been manually removed from the dataset. As such, users are advised to focus on large-scale (continental to global) analyses when using PDSI_CMIP6,

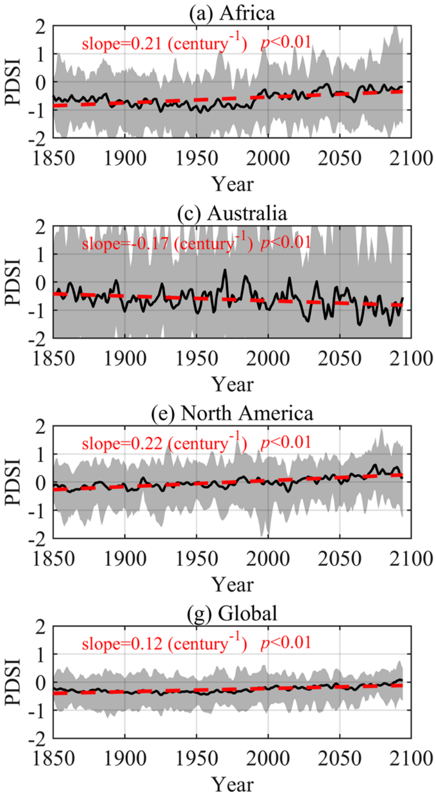

Fig. 6 Inter-annual changes of PDSI_CMIP6 averaged over different continents and the globe during 1850-2094 under the SSP5-8.5 scenario. The inserted dash lines and numbers represent the linear regression models.

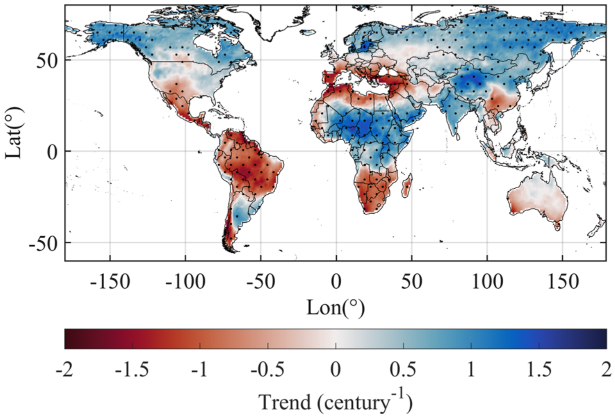

Fig. 7 Global distribution of the ensemble mean of long-term trends in PDSI_CMIP6 during 1850-2094 under the SSP5-8.5 scenario. Significant regions ($p < 0.05$, t-test) from more than 80% of the available climate models are marked with black dots.

particularly avoiding detailed regional assessments in extremely arid or cold regions where model performance is known to be less reliable.

(iii) The spatial resolution of PDSI_CMIP6 (1°) is finer than the native resolution of several CMIP6 models (e.g., CanESM5-1; see Table 1). Users should therefore exercise caution when interpreting fine-scale spatial patterns, particularly near land–ocean boundaries or in regions where the original model resolution is coarse. In such cases, excluding models with particularly low spatial resolution may help reduce uncertainty in regional analyses.

Code availability

The MATLAB code for the calculation and generation of the PDSI_CMIP6 dataset with calculations examples provided are available in the code repository here: https://zenodo.org/records/16203183.

Received: 6 August 2024; Accepted: 6 August 2025;

References

1. Merz, B. *et al*. Causes, impacts and patterns of disastrous river floods. *Nat Rev Earth Environ* **2**, 592–609 (2021).
2. Cook, B. I. *et al*. Megadroughts in the Common Era and the Anthropocene. *Nat Rev Earth Environ* **3**, 741–757 (2022).
3. Taye, M. T. & Dyer, E. Hydrologic Extremes in a Changing Climate: a Review of Extremes in East Africa. *Curr Clim Change Rep* **10**, 1–11 (2024).
4. de Brito, M. M. *et al*. Uncovering the Dynamics of Multi-Sector Impacts of Hydrological Extremes: A Methods Overview. *Earth's Future* **12**, e2023EF003906 (2024).
5. Mazzoleni, M., Odongo, V., Mondino, E. & Di Baldassarre, G. Water management, hydrological extremes, and society: modeling interactions and phenomena. *Ecology and Society* **26** (2021).
6. Yin, J. *et al*. Large increase in global storm runoff extremes driven by climate and anthropogenic changes. *Nat Commun* **9**, 4389 (2018).
7. Papalexiou, S. M. & Montanari, A. Global and Regional Increase of Precipitation Extremes Under Global Warming. *Water Resources Research* **55**, 4901–4914 (2019).
8. Xiong, J. & Yang, Y. Climate Change and Hydrological Extremes. *Curr. Clim. Change Rep.* **11**, 1 (2025).
9. Zhang, S. *et al*. Reconciling disagreement on global river flood changes in a warming climate. *Nat. Clim. Chang.* **12**, 1160–1167 (2022).
10. Sheffield, J., Wood, E. F. & Roderick, M. L. Little change in global drought over the past 60 years. *Nature* **491**, 435–438 (2012).
11. Asadieh, B. & Krakauer, N. Y. Global change in streamflow extremes under climate change over the 21st century. *Hydrology and Earth System Sciences* **21**, 5863–5874 (2017).
12. Trenberth, K. E. *et al*. Global warming and changes in drought. *Nature Clim Change* **4**, 17–22 (2014).
13. Intergovernmental Panel on Climate Change (IPCC). *Climate Change 2021 – The Physical Science Basis: Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, (2023).
14. Guttman, N. B. Accepting the Standardized Precipitation Index. *a calculation algorithm. J. Am. Water Resour. Assoc.* **35**, 311–322 (1999).
15. Vicente-Serrano, S. M., Beguería, S. & López-Moreno, J. I. A Multiscalar Drought Index Sensitive to Global Warming: The Standardized Precipitation Evapotranspiration Index. *Journal of Climate* **23**, 1696–1718 (2010).
16. Wang, T., Tu, X., Singh, V. P., Chen, X. & Lin, K. Global data assessment and analysis of drought characteristics based on CMIP6. *J. Hydrol.* **596**, 126091 (2021).
17. Cook, B. I. *et al*. Twenty-first century drought projections in the CMIP6 forcing scenarios. *Earth's Future* **8**, 2019–001461 (2020).
18. Palmer, W. C. *Meteorological Drought*. U.S. Department of Commerce, Weather Bureau, (1965).
19. Dai, A. Characteristics and trends in various forms of the Palmer Drought Severity Index during 1900–2008. *Journal of Geophysical Research: Atmospheres* **116** (2011).
20. Vicente-Serrano, S. M., Beguería, S. & López-Moreno, J. I. Comment on "Characteristics and trends in various forms of the Palmer Drought Severity Index (PDSI) during 1900–2008" by Aiguo Dai. *Journal of Geophysical Research: Atmospheres* **116** (2011).
21. Wang, Z., Yang, Y., Zhang, C., Guo, H. & Hou, Y. Historical and future Palmer Drought Severity Index with improved hydrological modeling. *Journal of Hydrology* **610**, 127941 (2022).
22. Yang, Y., Roderick, M. L., Zhang, S., McVicar, T. R. & Donohue, R. J. Hydrologic implications of vegetation response to elevated CO_2 in climate projections. *Nature Clim Change* **9**, 44–48 (2019).
23. Yang, Y. *et al*. Comparing Palmer Drought Severity Index drought assessments using the traditional offline approach with direct climate model outputs. *Hydrology and Earth System Sciences* **24**, 2921–2930 (2020).
24. van der Schrier, G., Barichivich, J., Briffa, K. R. & Jones, P. D. A scPDSI-based global data set of dry and wet spells for 1901–2009. *Journal of Geophysical Research: Atmospheres* **118**, 4025–4048 (2013).
25. Dai, A., Trenberth, K. E. & Qian, T. A Global Dataset of Palmer Drought Severity Index for 1870–2002: Relationship with Soil Moisture and Effects of Surface Warming. *Journal of Hydrometeorology* **5**, 1117–1130 (2004).
26. Zhao, T. & Dai, A. The Magnitude and Causes of Global Drought Changes in the Twenty-First Century under a Low–Moderate Emissions Scenario. *Journal of Climate* **28**, 4490–4512 (2015).
27. Zhao, T. & Dai, A. CMIP6 Model-Projected Hydroclimatic and Drought Changes and Their Causes in the Twenty-First Century. *Journal of Climate* **35**, 897–921 (2022).
28. Wells, N., Goddard, S. & Hayes, M. J. A Self-Calibrating Palmer Drought Severity Index. *Journal of Climate* **17**, 2335–2351 (2004).
29. Zhong, Z., He, B., Guo, L. & Zhang, Y. Performance of Various Forms of the Palmer Drought Severity Index in China from 1961 to 2013. *Journal of Hydrometeorology* **20**, 1867–1885 (2019).
30. Eyring, V. *et al*. Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development* **9**, 1937–1958 (2016).
31. O'Neill, B. C. *et al*. The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6. *Geoscientific Model Development* **9**, 3461–3482 (2016).
32. Jones, P. W. First- and Second-Order Conservative Remapping Schemes for Grids in Spherical Coordinates. *Monthly Weather Review* **127**, 2204–2210 (1999).
33. Allen, R. G., Pereira, L. S., Raes, D. & Smith, M. Crop evapotranspiration - Guidelines for computing crop water requirements - *FAO Irrigation and drainage paper 56* Rome (1998).
34. Penman, H. L. Natural Evaporation from Open Water, Bare Soil and Grass. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* **193**, 120–145 (1948).
35. Priestley, C. H. B. & Taylor, R. J. On the Assessment of Surface Heat Flux and Evaporation Using Large-Scale Parameters. *Monthly Weather Review* **100**, 81–92 (1972).
36. Yang, Y. & Roderick, M. L. Radiation, surface temperature and evaporation over wet surfaces. *Quarterly Journal of the Royal Meteorological Society* **145**, 1118–1129 (2019).
37. Oudin, L. *et al*. Which potential evapotranspiration input for a lumped rainfall–runoff model?: Part 2—Towards a simple and efficient potential evapotranspiration model for rainfall–runoff modelling. *Journal of Hydrology* **303**, 290–306 (2005).
38. Alley, W. M. The Palmer Drought Severity Index: Limitations and Assumptions. *Journal of Applied Meteorology and Climatology* **23**, 1100–1109 (1984).
39. Szép, I. J., Mika, J. & Dunkel, Z. Palmer drought severity index as soil moisture indicator: physical interpretation, statistical behaviour and relation to global climate. *Physics and Chemistry of the Earth, Parts A/B/C* **30**, 231–243 (2005).
40. Craigmile, P. F. & Guttorp, P. Comparing CMIP6 Climate Model Simulations of Annual Global Mean Temperatures to a New Combined Data Product. *Earth and Space Science* **10**, e2022EA002468 (2023).
41. Yang, Y. *et al*. Evapotranspiration on a greening Earth. *Nat Rev Earth Environ* **4**, 626–641 (2023).
42. Greene, C. A. *et al*. The Climate Data Toolbox for MATLAB. *Geochemistry, Geophysics, Geosystems* **20**, 3774–3781 (2019).
43. Xiong, J. PDSI_CMIP6 dataset. *Zenodo* https://doi.org/10.5281/zenodo.16131240 (2025).

Acknowledgements

This study is financially supported by the Ministry of Science and Technology of China (Grant No. 2023YFC3206603) and the Science and Technology Department of Qinghai Province (Grant No. 2024-SF-A6).

Author contributions

Jinghua Xiong and Yuting Yang designed the dataset. Jinghua Xiong wrote the first draft of the manuscript and Yuting Yang revised the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41597-025-05790-3.

Correspondence and requests for materials should be addressed to Y.Y.

Reprints and permissions information is available at www.nature.com/reprints.