



OPEN

DATA DESCRIPTOR

A global dataset of terrestrial evapotranspiration and soil moisture dynamics from 1982 to 2020

Kun Zhang^{1,2,9}✉, Huiling Chen^{3,9}, Ning Ma⁴, Shasha Shang⁵, Yunquan Wang⁶, Qinglin Xu⁷ & Gaofeng Zhu⁸✉

Quantifying terrestrial evapotranspiration (ET) and soil moisture dynamics accurately is crucial for understanding the global water cycle and surface energy balance. We present a novel, long-term dataset of global ET and soil moisture derived from the newly developed Simple Terrestrial Hydrosphere model, version 2 (SiTHv2). This ecohydrological model, driven by multi-source satellite observations and hydrometeorological variables from reanalysis data, provides daily global ET-related estimates (e.g., total ET, plant transpiration, soil evaporation, intercepted evaporation) and three-layer soil moisture dynamics at a 0.1° spatial resolution. Validation with *in-situ* measurements and comparisons with mainstream global ET and soil moisture products demonstrate robust performance of SiTHv2 in both magnitude and temporal dynamics of ET and soil moisture at multiple scales. The comprehensive water path characterization in the SiTHv2 model makes this seamless dataset particularly valuable for studies requiring synchronized water budget and vegetation response to water constraints. With its long-term coverage and high spatiotemporal resolution, the SiTHv2-derived ET and soil moisture product will be suitable to support analyses related to the hydrologic cycle, drought assessment, and ecosystem health.

Background & Summary

Water and energy exchanges between the land and atmosphere play a critical role in shaping the climate^{1,2}. Evapotranspiration (ET) serves as a crucial nexus in the intricate global water and carbon cycles^{3–5}, not only maintaining water balance but also affecting surface energy balance^{6–8}. As the primary mechanism for water transfer from soil to atmosphere, ET consumes over 50% of available radiation and returns 60% of annual terrestrial precipitation^{9,10}. Accurate estimation of ET and its different components (i.e., plant transpiration, soil evaporation, and canopy interception evaporation) is thus essential for understanding global energy balance and water cycle, particularly when separating vegetation feedbacks in the context of a changing climate.

Terrestrial ET is theoretically governed by two terms: atmospheric demand, which determines the potential ability to accept water vapor¹¹, and local water supply, which regulates the actual amount of ET¹². Among various water transfer mechanisms on Earth's surface, soil moisture plays a crucial role in impacting actual ET by directly representing soil water availability^{13,14}. Consequently, accurate estimates of both ET and soil moisture mutually complement each other. Although process-based ET models have proven effective in quantifying terrestrial ET¹⁵, the majority of these models do not comprehensively integrate underlying ecohydrological processes, such as canopy interception¹⁶, groundwater table dynamics^{17,18}, and root water uptake^{19,20}.

¹School of Geospatial Engineering and Science, SunYat-Sen University, Zhuhai, China. ²School of Biological Sciences, The University of Hong Kong, Hong Kong, China. ³College of Geography and Environmental Science, Zhejiang Normal University, Jinhua, China. ⁴Key Laboratory of Water Cycle and Related Land Surface Processes, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China. ⁵Tianjin Key Laboratory of Water Resources and Environment, Tianjin Normal University, Tianjin, China. ⁶School of Environmental Studies, China University of Geosciences, Wuhan, China. ⁷The 404 Company Limited, CNNC, Lanzhou, China. ⁸College of Earth and Environmental Sciences, Lanzhou University, Lanzhou, China. ⁹These authors contributed equally: Kun Zhang, Huiling Chen. ✉e-mail: zhangkun3@mail.sysu.edu.cn; zhugf@lzu.edu.cn

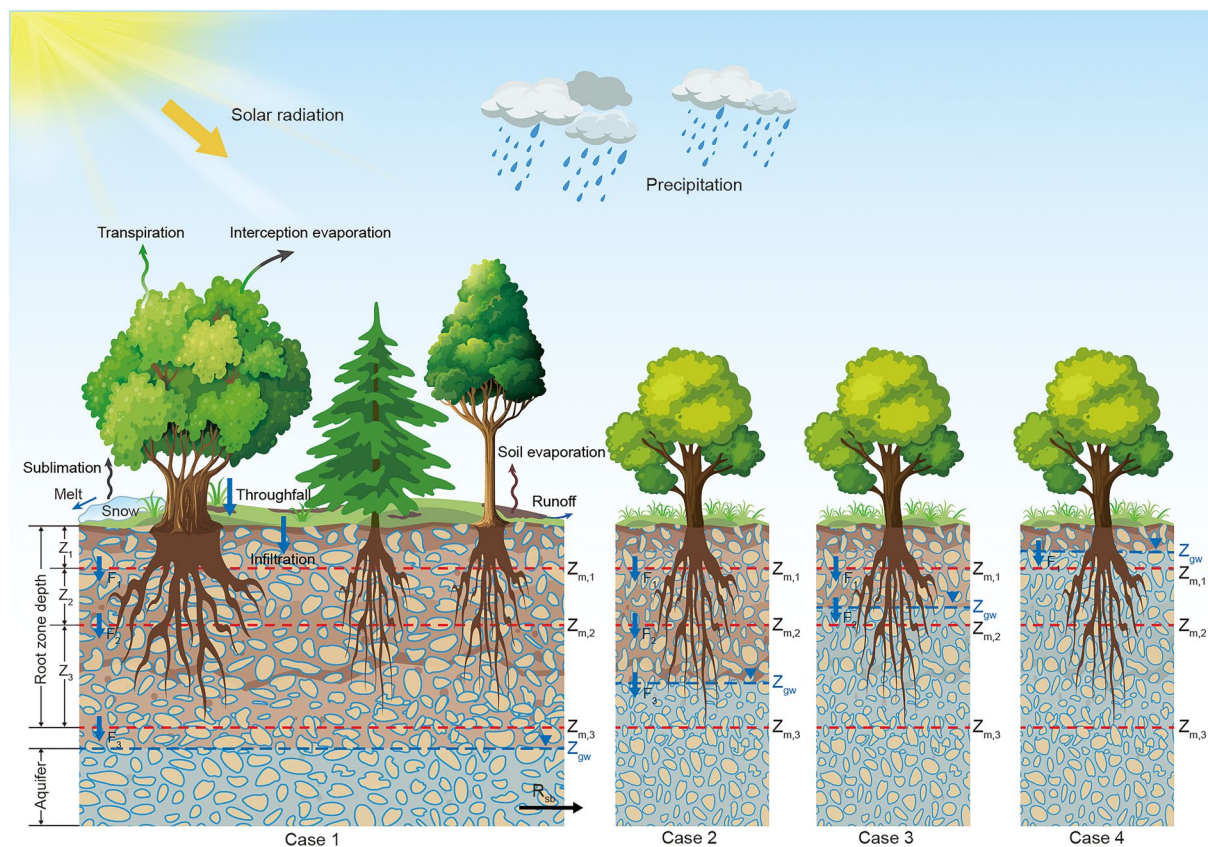


Fig. 1 The conceptual diagram of the hydrological process in the SiTHv2 model. Different cases indicate the groundwater table (z_{gw}) dynamically related with potential root zone depth.

The Simple Terrestrial Hydrosphere (SiTH) model²¹ integrates the eco-hydrological process with remotely sensed information to characterize the water pathway within the groundwater-soil-plant-atmosphere continuum (GSPAC)²², which demonstrated favorable performance in modeling ET across various ecosystems^{23–25}. Recent evidence has shown that groundwater table dynamics significantly affect root zone soil moisture^{26,27}, surface runoff production^{28,29}, land-atmosphere interactions^{30,31}, and regional climate^{32,33}. However, most research focused on regional scales, while simple and effective expressions of groundwater effects on vegetation at the global scale have been lacking. To address this, we considered potential deeper layer recharge in soil moisture based on the saturated hydraulic conductivity for different soil properties³⁴. It should be noted that root distribution is important in influencing vegetation transpiration³⁵, while existing models have predominantly focused on the vertical distribution of roots with relatively shallow burial depths. Meanwhile, the comprehensive effect of vertical root distribution and multi-layer soil water content on plant transpiration are not well represented in current models²¹. As a result, previous models with shallow root depths tend to underestimate vegetation drought resistance and overestimate the sensitivity of plant transpiration to water stress. Thus, the comprehensive effect of vertical root distribution and multi-layer soil water content are synchronously considered in the updated SiTH model (SiTHv2)²⁰. Moreover, we modified the critical parameter of soil moisture in SiTHv2 to better represent plant water stress by considering different vegetation heights and varied environmental conditions. Notably, the SiTHv2 model uses satellite-observed vegetation status, including vegetation greenness³⁶ and optical depth³⁷, to characterize plant growth and health.

In this study, we present a novel dataset of global terrestrial ET (including different components) and soil moisture generated using the SiTHv2 model. This dataset provides daily estimates at a spatial resolution of 0.1°, with seamless global land coverage. The conceptual graph of the SiTHv2 model is illustrated in Fig. 1, and the model description and data production procedure are presented schematically in the Methods section. The availability of this dataset will provide essential support and valuable insights for scientific research in the domains of land-atmosphere interactions, ecohydrological modeling, and global change studies.

Methods

Forcing data and preprocessing. *Remote sensing data.* Vegetation dynamics play a crucial role in the SiTHv2 model, particularly in the partitioning scheme for available energy. We utilized the leaf area index (LAI) from the European Geoland2/BioPar project Version 2 (GEOV2)³⁸ based on the retrievals of the Advanced Very High-Resolution Radiometer (AVHRR). This product offers global LAI observations every 10 days at a 0.05° spatial resolution and has been proven as a reliable satellite-based product for providing long-term vegetation variables with global coverage. Additionally, we employed vegetation optical depth (VOD) to characterize the

actual vegetation water content from the VOD Climate Archive (VODCA)³⁷. This product based on microwave observations in various spectral bands, provides daily VOD estimates at a 0.25° spatial resolution. We selected the Ku-band due to its extensive data availability. Furthermore, we used the Global Land Surface Satellite (GLASS) product³⁹ to obtain surface broadband albedo and emissivity observations with 8-day intervals and a spatial resolution of 0.05°. Specifically, the GLASS02B03 dataset informed the black-sky and white-sky albedo, while the GLASS03B02 dataset contributed to broadband emissivity.

Hydrometeorological data. Hydrometeorological variables serve as essential inputs for the SiTHv2 model. We acquired daily air temperature and air pressure data from the Multi-Source Weather (MSWX) product⁴⁰, a bias-corrected meteorological dataset with global coverage and a spatial resolution of 0.1° since 1979. In terms of radiation, we used downward shortwave radiation and downward longwave radiation from the land component of the fifth-generation European reanalysis data (ERA5-Land, hereafter referred as ERA5L)⁴¹. We also utilized skin temperature from ERA5L, combined with satellite-based broadband emissivity, to calculate upward longwave radiation. Moreover, to calculate the actual albedo (i.e., blue-sky albedo), we assumed the blue-sky albedo is linearly weighted between black-sky and white-sky albedo. The weight coefficient is the sky scattering ratio, derived from the mean surface direct shortwave radiation and mean surface downward shortwave radiation from ERA5⁴². Consequently, global long-term seamless surface net radiation can be calculated using the balance equation of the four radiation components. Additionally, the precipitation data used in this study was obtained from the ERA5L product. It is noteworthy that the original temporal interval of ERA5 and ERA5L are hourly, yet their spatial resolutions differ, with ERA5 at 0.25° and ERA5L at 0.1°.

Auxiliary data. Land cover/use dynamics also significantly influence the estimation of terrestrial ET, considering different biological effects (e.g., plant height and root depth) in the parameterization scheme. Current satellite-based land cover products are hard to provide temporal dynamic information over a long-term period. Thus, we obtained annual land cover dynamics from the Historic Land Dynamics Assessment+ (HILDA+) product⁴³, a long-term land cover/use product combining multiple open data streams, including remote sensing, reconstructions, and statistics. Moreover, the global soil type used in the SiTHv2 model was acquired from the Harmonized World Soil Database (HWSD) v1.2, combining available soil information from regional and national institutes worldwide and offering a soil raster database with a 30-arc-second resolution⁴⁴.

To reconcile the varying spatial resolutions and temporal intervals of the aforementioned multiple data sources in the SiTHv2 model, we uniformly resampled the spatial resolution to 0.1° using the bilinear approach and filled the temporal gap to the daily scale based on spline interpolation.

Validation and comparison data. We evaluated the accuracy of SiTHv2 estimates using both *in-situ* observations and grid-based ET and soil moisture products. Specifically, we validated the total ET estimates of SiTHv2 using latent heat flux measurements from 175 global eddy covariance (EC) stations with a daily interval^{45,46}. These EC stations encompass nine major plant functional types (PFTs) under various climate conditions, including croplands (CRO), deciduous broadleaf forests (DBF), evergreen broadleaf forests (EBF), evergreen needleleaf forests (ENF), grasslands (GRA), mixed forests (MF), open shrublands (OSH), savannas (SAV), and wetlands (WET). Moreover, we employed pre-processing method according to previous study¹⁰ and ensure the surface energy closure at least above 70%⁴⁷ (Figure S1). Five ET products were selected for comparison at both basin and global scales, including ET estimates from: (1) the Global Land Evaporation Amsterdam Model (GLEAM)⁴⁸; (2) the calibration-free complementary relationship (CR) model⁴⁹; (3) the FluxCOM product⁵⁰; (4) the Global Land Data Assimilation System (GLDAS-Noah)⁵¹; and (5) ERA5L⁴¹. Furthermore, we utilized ET simulations from 20 Earth System Models (ESMs) in the Coupled Model Intercomparison Project Phase 6 (CMIP6)⁵² to compare SiTHv2 ET estimates concerning annual variation and magnitude of total ET volume at a global level.

We generated water-balanced ET (ET_{wb}) from the residual of the basin water balance equation (precipitation minus runoff and changes in total water storage) to validate SiTHv2-based ET estimates at a basin scale. A total of 49 basins were selected that covering a broad range of climate zones globally. The precipitation data used to calculate ET_{wb} obtained from the gauge-based Global Precipitation Climatology Center (GPCC) Full Data Monthly Product⁵³. However, for the Continental United States (CONUS), we used the Parameter-Elevation Regressions on Independent Slopes Model (PRISM) precipitation data⁵⁴. Runoff observations from hydrological stations for the majority of basins, excluding those in the CONUS and China, were obtained from the Global Runoff Data Center (GRDC). Runoff data for basins within the CONUS and China were sourced from the United States Geological Survey and the China Sediment Bulletin, respectively. Basin-scale changes in total water storage were derived from Gravity Recovery and Climate Experiment (GRACE) terrestrial water storage anomaly products⁵⁵. To minimize uncertainties arising from noise terms in different solutions, we calculated the average of three GRACE products processed by the Geoforschungs Zentrum Potsdam, the Center for Space Research at the University of Texas, Austin, and the NASA Jet Propulsion Laboratory. For more details on the ET_{wb} calculation, please refer to the authors' previous studies^{56,57}.

To validate the soil moisture estimated by the SiTHv2 model, we selected 12 sites from FLUXNET offering daily *in-situ* measurements of soil water content (SWC). We also compared SiTHv2-derived soil moisture with a global soil moisture dataset, SoMo.ml⁵⁸, which is based on the Long Short Term Memory (LSTM) network and trained with abundant *in-situ* measurements from over 1,000 stations. SoMo.ml has been considered to perform well in both spatial distribution and temporal dynamics of soil moisture, and can serve as a benchmark to evaluate modelled and remotely sensed data⁵⁸. Furthermore, we utilized four types of microwave-based satellite products for soil moisture comparison, comprising: L-band (1.4 GHz) products obtained from the Soil Moisture Ocean Salinity (SMOS) at the Barcelona Expert Centre (BEC)⁵⁹ and the Soil Moisture Active Passive (SMAP) Level-3 products⁶⁰; as well as C-band (6.9 and 7.3 GHz) products from the Advanced SCATterometer

Product	Variable	Version/Type	Spatial resolution	Interval(3)	Time span
FLUXNET2015	ET & SWC	N/A	<i>In-situ</i>	Daily	Varied with sites
GLEAM	ET	v3.7a	0.25°	Monthly	1980-
CR	ET	v1.0	0.25°	Monthly	1982–2016
FluxCOM	ET	RS_METEO(1)	0.5°	Monthly	2001–2013
GLDAS-Noah	ET	v2.0	0.25°	Monthly	1982–2014
ERA5-Land	ET	N/A	0.1°	Monthly	1950-
CMIP6	ET	20 ESMs(2)	0.5°	Monthly	1979–2014
SoMo.ml	SWC	N/A	0.25°	Daily	2001–2019
SMAP	SWC	Enhanced L3	9 km	Daily	2015-
SMOS	SWC	BEC L3	0.25°	Daily	2011-
ASCAT	SWC	v7	0.25°	Daily	2007-
AMSR2	SWC	JAXA	0.25°	Daily	2012-

Table 1. Summary of the validation/comparison datasets used in this study. Note: (1) All the ensemble latent heat flux estimates of 36 members with energy balance correction; (2) Details of the 20 ESMs (historical period) are provided in Table S1 in Supplementary; (3) The time intervals listed in the table represent only the types used in this study, while some products may have a higher temporal resolution.

(ASCAT)⁶¹ and the Advanced Microwave Scanning Radiometer 2 (AMSR2) of Japan Aerospace Exploration Agency (JAXA)⁶². These products offer a spatial resolution of 0.25°, with the exception of the SMAP L3, which features an enhanced spatial resolution of 9 km.

The selected validation and comparison product offers independent reference for evaluating the ET and soil moisture estimates derived from the SiTHv2 model. A summary of all mentioned products can be found in Table 1.

SiTHv2 model description. The SiTHv2 model is an updated version of the SiTH model, which is developed for modeling the water pathway within the GSPAC (Fig. 1). Evapotranspiration (ET) and soil moisture (also known as soil water content, SWC) are two crucial outputs generated by the model. Specifically, the total ET in the SiTHv2 is the sum of soil evaporation (E_s), plant transpiration (T_r), and canopy interception evaporation (E_i). Each item is derived from different kinds of constraints from potential transpiration and evaporation for the canopy and soil surface, respectively.

$$ET = E_i + E_s + T_r \quad (1)$$

$$E_i = f_{\text{wet}} \cdot E_{\text{pc}} \quad (2)$$

$$E_s = f_{\text{sm}} \cdot E_{\text{ps}} \quad (3)$$

$$T_r = f_v \cdot f_t \cdot \left[\sum_{i=1}^n (f_{\text{smv},i} \cdot T_{\text{ps},i}) + \sum_{i=1}^n (T_{\text{pg},i}) \right] \quad (4)$$

where n is the total number of soil layers, which is set as three in SiTHv2; the f -functions are the different constraints on the potential evapotranspiration for canopy (E_{pc}) and soil surface (E_{ps}), where the f_{wet} is the relative wetness of canopy surface; the f_{sm} is the soil moisture constraint of bare soil evaporation; the $f_{\text{smv},i}$ is the soil moisture constraint of plant transpiration in the unsaturated zone at i^{th} soil layer; the f_v and f_t are the vegetation water content and temperature constraint on plant health, respectively. Notably, the characterization of potential transpiration rate (T_p) is comprised of both unsaturated and saturated layers when the groundwater table extends to the i^{th} soil layer, denoted as $T_{\text{ps},i}$ and $T_{\text{pg},i}$, respectively (see different cases in Fig. 1). The calculation of E_{pc} and E_{ps} in the SiTHv2 model are based on the Priestly-Taylor equation⁶³, while the T_p can be derived by the residual of E_{pc} after intercepted consumption.

$$E_{\text{ps}} = \alpha \cdot \frac{\Delta}{\Delta + \gamma} \cdot \frac{(R_{\text{ns}} - G)}{\lambda} \quad (5)$$

$$T_p = (1 - f_{\text{wet}}) \cdot E_{\text{pc}} = (1 - f_{\text{wet}}) \cdot \alpha \cdot \frac{\Delta}{\Delta + \gamma} \cdot \frac{R_{\text{nc}}}{\lambda} \quad (6)$$

where α is the Priestly-Taylor coefficient and set to 1.26 in this study; Δ is the slope of the saturated vapor pressure curve ($\text{kPa } ^\circ\text{C}^{-1}$); γ is the psychrometric constant and set to $0.066 (\text{kPa } ^\circ\text{C}^{-1})$; λ is the latent heat of evaporation (MJ kg^{-1}); R_{ns} and R_{nc} are the net radiation (R_n) allocated to the bare soil and canopy surface, respectively. R_{ns} is complementary with R_{nc} , which can be derived by the Beer law:

$$R_{ns} = e^{(-k_{Rn} \cdot LAI)} \cdot R_n \tag{7}$$

where k_{Rn} is the extinction coefficient and is set to 0.6 in this study; LAI is the leaf area index. Thus, the potential transpiration for different soil layers ($T_{p,i}$) can be derived from the T_p by considering the vertical distributions of plant roots and the soil water content at different depths:

$$T_{p,i} = \frac{r_i \cdot (\bar{\theta}_i/\theta_s)^b}{\sum_{i=1}^n [r_i \cdot (\bar{\theta}_i/\theta_s)^b]} \cdot T_p \tag{8}$$

where θ_s is the saturated soil water content; $\bar{\theta}_i$ is the layer mean water content depend on the fraction of unsaturated zone for each soil layer; b is a soil parameter (dimensionless) depends on soil properties²¹; r_i presents the vertical root density, which is described by the linear dose response model⁶⁴:

$$r_i = \frac{1}{1 + (z_{m,i}/D_{50})^c} - \frac{1}{1 + (z_{m,i-1}/D_{50})^c} \tag{9}$$

where the depth at which 50% of the root mass is located is referred to as D_{50} ; $z_{m,i}$ is the bottom depth of the i^{th} layer; c is the shape parameter derived from the logarithmic relation between D_{50} and D_{95} , which is the depth at which 95% of the root mass is located. Thus, the variables $T_{ps,i}$ and $T_{pg,i}$ can be further obtained from the $T_{p,i}$ under different conditions of groundwater depths (z_{gw}) and soil water content for different layers.

$$T_{ps,i} = \begin{cases} 0 & z_{gw} \leq z_{m,i-1} \\ \frac{(z_{gw} - z_{m,i-1}) \cdot \theta_i}{(z_{gw} - z_{m,i-1}) \cdot \theta_i + (z_{m,i} - z_{gw}) \cdot \theta_s} \cdot T_{p,i} & z_{m,i-1} < z_{gw} < z_{m,i} \\ T_{p,i} & z_{gw} \geq z_{m,i} \end{cases} \tag{10}$$

$$T_{pg,i} = \begin{cases} T_{p,i} & z_{gw} \leq z_{m,i-1} \\ \frac{(z_{m,i} - z_{gw}) \cdot \theta_s}{(z_{gw} - z_{m,i-1}) \cdot \theta_i + (z_{m,i} - z_{gw}) \cdot \theta_s} \cdot T_{p,i} & z_{m,i-1} < z_{gw} < z_{m,i} \\ 0 & z_{gw} \geq z_{m,i} \end{cases} \tag{11}$$

In addition, the f -functions for different constraints in SiTHv2 can be calculated as follows:

$$f_{wet} = \min \left\{ \chi \cdot \frac{S_c}{T_p}, 1 \right\} \tag{12}$$

$$f_{sm} = \begin{cases} 0 & \theta_i \leq \theta_{wp} \\ \frac{\theta_i - \theta_{wp}}{\theta_{fc} - \theta_{wp}} & \theta_{wp} \leq \theta_i \leq \theta_{fc} \\ 1 & \theta_i \geq \theta_{fc} \end{cases} \tag{13}$$

$$f_t = e^{-[(T_a - T_{opt})/T_{opt}]^2} \tag{14}$$

$$f_v = \sqrt{VOD/VOD_{max}} \tag{15}$$

where χ is fractional interception occurring during daytime, which is set to 0.7; S_c is the water storage capacity of canopy, which is defined as the product of precipitation and LAI ; θ_{fc} is the soil field capacity, and θ_{wp} is the soil moisture at the wilting point; θ_i is the actual soil moisture at the i^{th} soil layer; T_a is the air temperature (S_c); T_{opt} is the optimum plant growth temperature (S_c), which can be derived from the air temperature when the multiply of LAI , R_n and T_a are high; VOD_{max} is the maximum value for the annual time series of VOD retrievals.

In SiTHv2, the critical threshold of soil moisture (θ_c), which describes the point of the plant under water stress, is dynamically dependent on the plant traits and environmental conditions, rather than a static constant. Hence, the θ_c can be derived as follows^{65,66}:

$$\theta_c = (1 - p)(\theta_{fc} - \theta_{wp}^h) + \theta_{wp}^h \tag{16}$$

$$p = \frac{1}{1 + E_p} - \frac{w}{1 + H_c} \tag{17}$$

$$\theta_{wp}^h = \theta_{wp}/k \quad (18)$$

where θ_{fc} is the soil field capacity; θ_{wp}^h is the adjusted wilting point soil moisture by considering the canopy height (H_c); p is a parameter for regulating θ_c between θ_{fc} and θ_{wp}^h , which can be determined by potential ET (E_p) and H_c ; w is set to 0.1 denotes the weight of H_c on the θ_c ; k is a sensitivity index of soil water content, which can be determined as the square root of H_c . Thus, the soil moisture constraint function for potential transpiration at the i^{th} layer ($f_{smv,i}$) can be calculated as follows:

$$f_{smv,i} = \begin{cases} 0 & \theta_i \leq \theta_{wp}^h \\ 1 - \left(\frac{\theta_c - \theta_i}{\theta_c - \theta_{wp}^h} \right)^k & \theta_{wp}^h \leq \theta_i \leq \theta_c \\ 1 & \theta_i \geq \theta_c \end{cases} \quad (19)$$

Collectively, in the SiTHv2 model, all the water and energy components are theoretically balanced. That is, all the consumption terms (e.g., ET, deep percolation, etc.) and input terms (e.g., precipitation, etc.) are dynamically balanced on the time series. Notably, the dynamic changes in soil moisture at different layers are both intermediate results of the previous time step and important variables regulating the dynamic changes in soil water stress in the next stage. For more detailed model structure, please refer to the authors' previous studies^{20,21}.

Global data generation. The SiTHv2 model was employed to estimate daily terrestrial ET and its various components at a spatial resolution of 0.1°, as well as the soil moisture dynamics across three soil layers. This model was implemented for each grid cell in their respective time series, driven by forcing data such as air temperature, pressure, net radiation, precipitation, and satellite-based observations of vegetation dynamics and surface albedo. To enhance computational efficiency, parallel computation was implemented at each grid cell. Importantly, a 100-year spin-up period was applied to achieve equilibrium states for each intermediate variable during the simulation. Finally, global seamless estimates of total terrestrial ET, plant transpiration, soil evaporation, canopy intercepted evaporation, and three layers soil moisture were obtained by running the SiTHv2 model from 1982 to 2020. In addition, we also aggregated the daily estimates to provide the monthly and yearly dataset.

Data Records

The dataset comprises multi-type of outputs, including the total terrestrial ET, plant transpiration (T_i), bare soil evaporation (E_s), canopy interception evaporation (E_i), and the soil moisture dynamics across three soil layers at varying depths. Generated dataset is available at three temporal intervals (i.e., daily, monthly, and annual) and has a global spatial resolution of 0.1°. All data is stored in Network Common Data Form (NetCDF) files, accessible via the data repository of National Tibetan Plateau Data Center (TPDC)⁶⁷. Files with different temporal resolutions are organized in separate directories at TPDC. Filenames follow the structure “SiTH.v2.<VV>.<XXXX>.<YYYY>.nc”, where “VV” represents the variable name (e.g., ET – Evapotranspiration), “XXXX” indicates the temporal resolution (e.g., Daily, Monthly, Annual), and “YYYY” denotes the four-digit year. Further details are provided in the global attributes of each NetCDF file.

Technical Validation

ET validation using *in-situ* observations. At the site scale, we employed observed latent heat flux (LE) from the FLUXNET2015 dataset to validate the SiTHv2-derived ET estimates. *In-situ* observed LE flux was converted into ET values by dividing by the latent heat of vaporization. All site observations were categorized according to their plant functional types (PFTs). Figure 2 demonstrates that SiTHv2-derived ET estimates exhibit strong agreement with flux observations, achieving correlation coefficients (R) of at least 0.76 across different PFTs. The highest R -value, approximately 0.85, is observed in grasslands (GRA), wetlands (WET), and deciduous broadleaf forests (DBF). In terms of root mean square error (RMSE), the largest value occurred in evergreen broadleaf forests (EBF) at 1.06 mm day⁻¹, representing the sole situation where RMSE exceeded 1 mm day⁻¹ among all PFTs. It is worth noting that RMSE differences are also associated with varying ET magnitudes in different PFTs. Conversely, the dimensionless Nash-Sutcliffe Efficiency (NSE) effectively eliminates this effect. We found that the NSE of SiTHv2-derived ET estimates of EBF is the lowest at 0.22, while it remained greater than 0.5 in other PFTs, even reaching 0.7 in GRA. Additionally, linear slopes between SiTHv2 and *in-situ* observations are closely aligned with the 1:1 line in most PFTs, particularly in DBF (0.99), EBF (1.02), MF (0.96), and WET (0.93). Overall, the SiTHv2 model is capable of providing reliable ET estimates at the *in-situ* level.

Comparison of ET with independent water balanced ET at basin level. We compared the ET estimates derived from SiTHv2 with independent water balance results (ET_{wb}) across 49 major global basins. This comparison aimed to address uncertainties arising from scale mismatches between *in-situ* measurements and grid-based ET estimations at a spatial resolution of 0.1°. As illustrated in Fig. 3, SiTHv2-derived ET estimates demonstrated a strong agreement with ET_{wb} , exhibiting a R -value of 0.96 and a NSE value of 0.90. Although this NSE value is slightly lower than that of the CR model (NSE = 0.93), both SiTHv2 and CR exhibits identical linear regression slopes of 1.01. Notably, SiTHv2 displayed a superior intercept compared to CR, indicating a closer estimates to ET_{wb} in the basins with low ET volumes. While the GLEAM and ERA5L product also showed a close slope (0.99), both of them consistently overestimated ET in comparison to ET_{wb} due to their elevated intercepts.

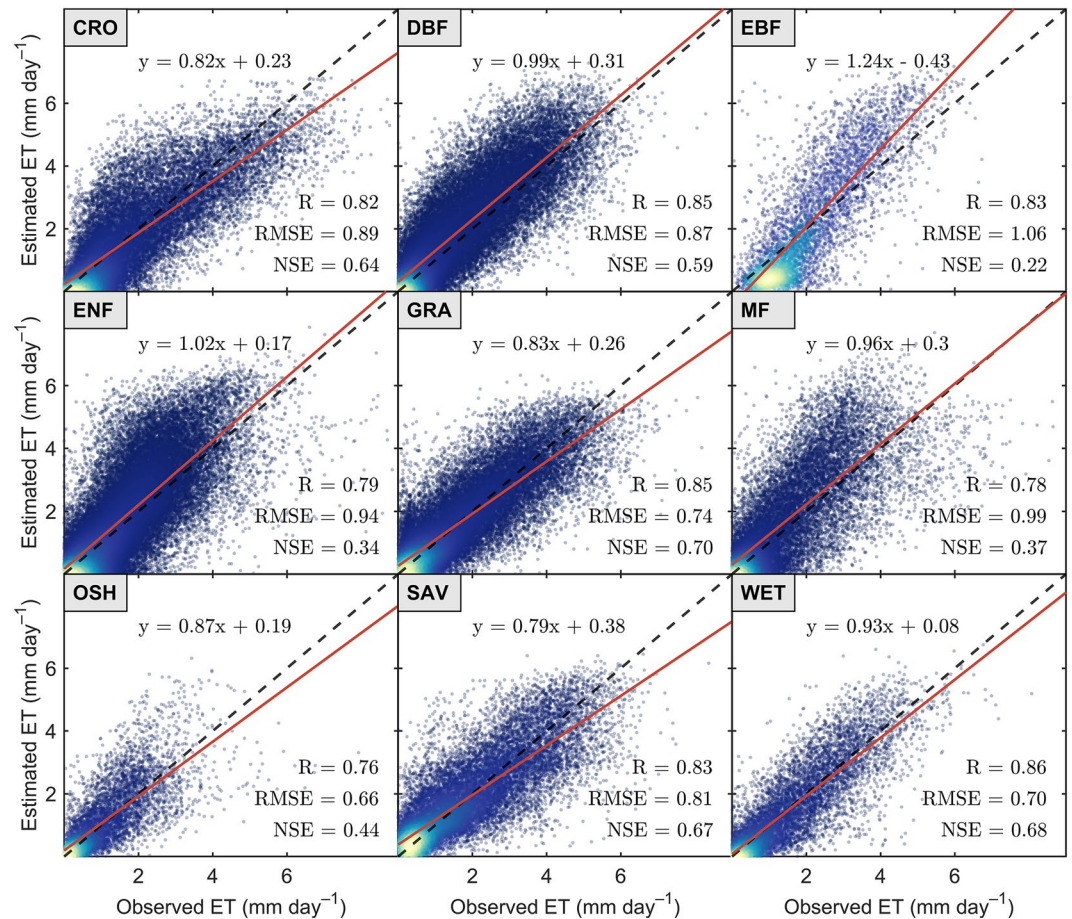


Fig. 2 Validation of evapotranspiration (ET) between SiTHv2-derived estimates and *in-situ* observations based on eddy covariance for different plant functional types (PFTs). Details of selected sites are summarized in an Excel file stored in the repository, and the link is available in the Code Availability section.

In contrast, FluxCOM exhibited the same R -value as SiTHv2, however, SiTHv2's RMSE of 95.68 and NSE of 0.90 surpassed the corresponding FluxCOM statistics, highlighting its better performance at these basins.

The accuracy of estimated ET volume was further assessed by comparing the ratio of mean annual ET to the ET_{wb} across different basins. From Fig. 4a, the majority of basins exhibit ratio values near 1, indicating that the SiTHv2-derived ET is closely aligned with ET_{wb} . Moreover, a comparison with other mainstream global ET products reveals that the ratio values of SiTHv2 across 49 basins (with a mean value of 0.98) demonstrate superior performance (Fig. 4b). A relatively high ratio value is observed in a few individual basins located at the high latitudes of the Northern Hemisphere, where a similar pattern also found in other ET products. Such discrepancy might be attributed to the inadequate characterization of the soil freezing and thawing process^{68,69} in current ET models.

Comparison of ET with global ET product. Using the long-term global ET estimates from SiTHv2, we generated a global distribution of mean annual ET (Fig. 5). Spatially, tropical regions exhibit the highest ET, while drylands show the lowest ET. Such pattern is typically determined by the combined effect of atmospheric evaporative demand and local water availability. The zonally mean profile of SiTHv2-derived ET is also similar to other global ET products, particularly the GLEAM and CR. We further evaluated the annual variation of global terrestrial ET. SiTHv2-derived ET displays a variation comparable to most ET products and fall within the interquartile range (IQR) of CMIP6 results (Fig. 6a). Regarding the annual linear trend, most global ET products (except for ERA5L) show a significant increase over the past few decades. SiTHv2 has the highest increasing trend, with a value of $0.53 \text{ mm year}^{-1}$ from 1982 to 2020, followed by GLDAS-Noah ($0.42 \text{ mm year}^{-1}$), and median of 20 ESMs in CMIP6 ($0.36 \text{ mm year}^{-1}$). CR and GLEAM has the same increasing rate of $0.31 \text{ mm year}^{-1}$. It is worth noting that the ET estimates from FluxCOM exhibit an insignificant trend, which due to its shorter time span with covering only the most recent 14 years from 2001. In terms of total volume of global terrestrial ET (Fig. 6b), the mean value of SiTHv2-derived ET is $71.5 \times 10^3 \text{ km}^3$, closely aligned with the mean value of other ET products ($71.8 \times 10^3 \text{ km}^3$). Comparatively, the 20 ensemble ESMs of CMIP6 present a broader range of global terrestrial ET volume, with a mean value of $78.5 \times 10^3 \text{ km}^3$. Previous studies^{1,50,70–73} have reported global terrestrial ET volumes primarily within the range of 65 to $75 \times 10^3 \text{ km}^3$. In summary, the global spatial pattern, annual variation, and total magnitude of ET estimated by SiTHv2 are reliable based on current understandings of global water cycles.

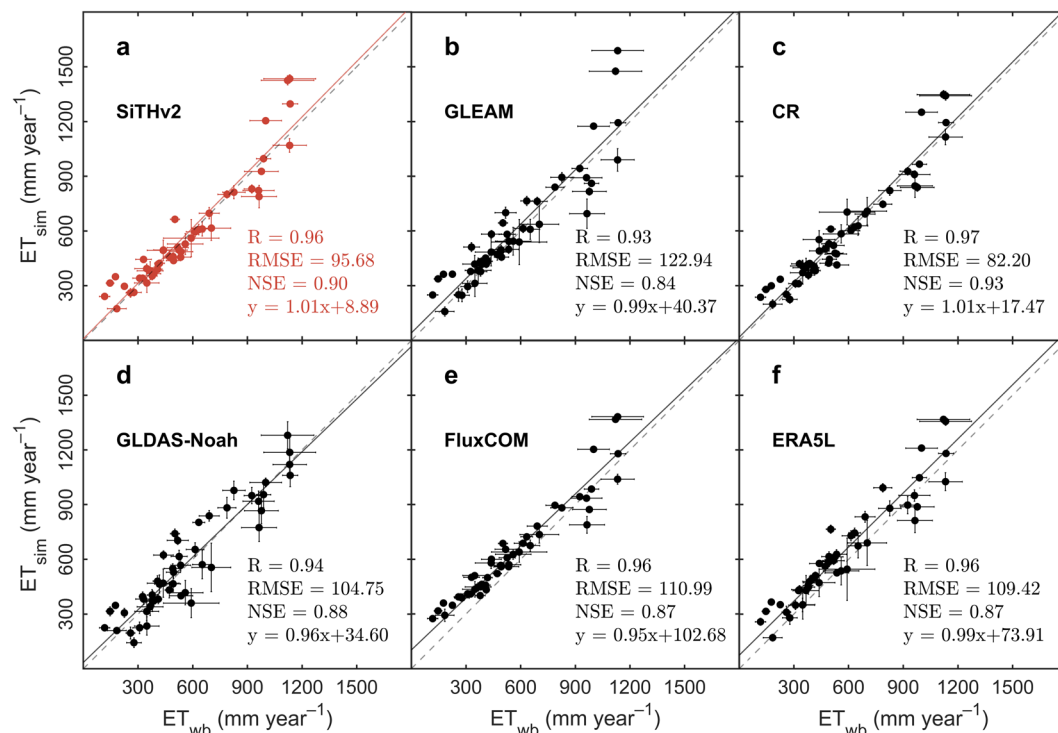


Fig. 3 Basin-level comparison between different ET estimates and the independent ET derived from basin water balance across 49 basins. Different sub-panel respectively represented the comparison results from (a) the SiTHv2 model; (b) the GLEAM model; (c) the CR model; (d) the GLDAS-Noah; (e) the FluxCOM product; (f) the ERA5L. The dash line in each sub-figure denotes the 1:1 line. The error bar for each point indicates the annual variation for simulated ET and ET_{wb}, respectively.

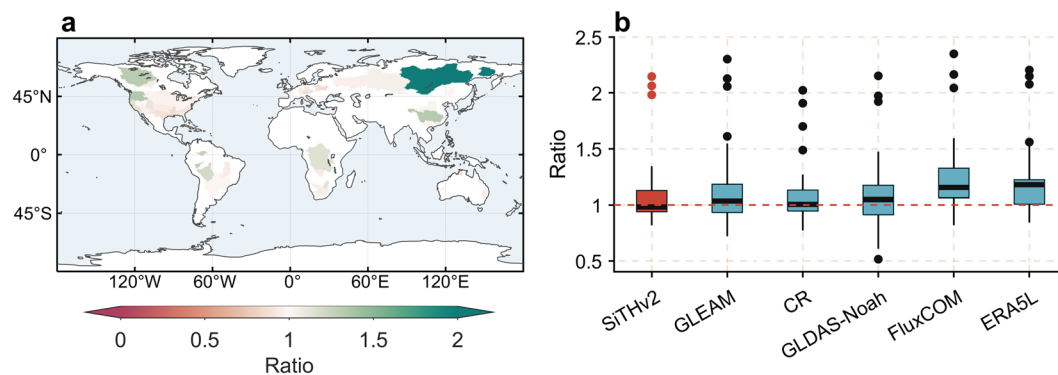


Fig. 4 The ratio of multi-year mean ET estimates from different ET datasets relative to the ET_{wb} across the 49 basins. (a) represents the spatial distribution of the ratio for the SiTHv2; (b) is the intercomparison of the ratios for different models/products.

Comparison of ET partitioning ratio with independent T/ET dataset. The SiTHv2 model enables estimation of different components of total ET, making the ratio between transpiration and ET (T/ET) a crucial metric for evaluating accuracy of ET partitioning⁷⁴. Although *in-situ* measurements can provide the T/ET benchmark based on sap flow and eddy covariance techniques, their relative scarcity and short duration limit the provision of grid-based reference applied globally⁷⁵. Moreover, isotope-based methods, which also only cover short measurement periods, tend to overestimate T/ET^{27,76,77}. By contrast, Model-data fusion (MDF) method can be employed to incorporate available observations into a process-based framework. A recent study published a new T/ET dataset over China spanning from 1981 to 2015, based on multivariate observations, demonstrating good agreement with independent site measurements⁷⁸. Hence, we extracted the grid-based T/ET ratios from SiTHv2 across China and compared with the MDF product (Fig. 7). The results indicate that the SiTHv2-derived T/ET ratios exhibit not only high spatial agreement with MDF, with a R -value of 0.85, RMSE is 0.14, and regression slope is 1.03 with intercept of 0.05 (Fig. 7b), but also a strong correlation for long-term variations ($R=0.90$).

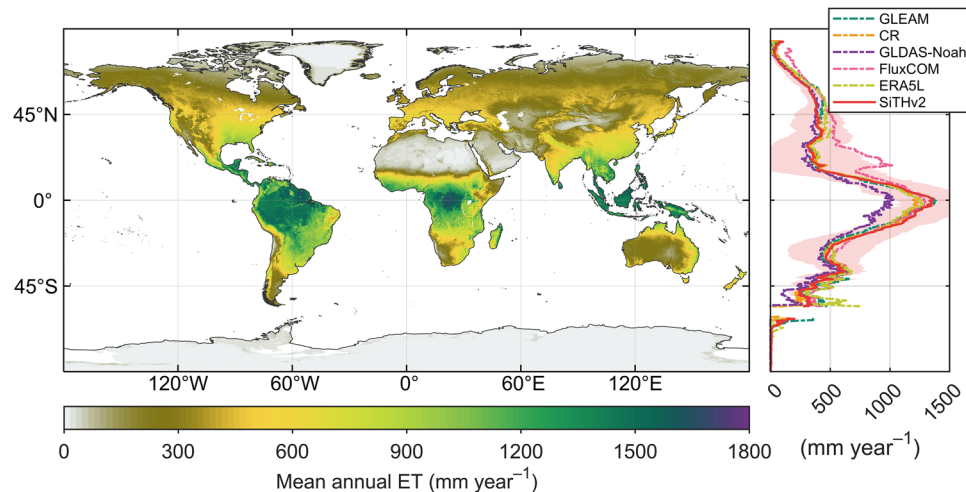


Fig. 5 Global distribution of mean annual terrestrial ET derived by the SiTHv2 model, and the comparison of zonally mean profile with the other ET products. The red shadow in the right panel represents the standard deviation for the zonally mean of SiTHv2.

Additionally, the linear trend for T/ET estimates from SiTHv2 and MDF are both significant ($p < 0.01$), with a value of $1 \times 10^{-3} \text{ year}^{-1}$ for SiTHv2 and $2 \times 10^{-3} \text{ year}^{-1}$ for MDF.

Soil moisture validation using *in-situ* observations. Soil moisture, also known as soil water content (SWC), is a key variable in the SiTHv2 model. Due to the limited availability of SWC measurements at most flux stations, we selected 12 sites with observed SWC to validate the performance of the SWC estimates in SiTHv2 model at the *in-situ* level. These selected flux sites can cover the majority of various PFTs under different climate conditions. As depicted in Fig. 8, the SiTHv2 model can generally capture the dynamic changes in SWC across most sites. Specifically, the correlation coefficient reaches a mean value of 0.75, with the highest R -value of 0.81 observed at the AU-Stp (GRA) while the lowest R -value of 0.55 at the AU-Tum (EBF). The metrics of RMSE for all selected sites are relatively small, with a mean value of $0.07 \text{ m}^3 \text{ m}^{-3}$. The site AU-ASM (SAV) has the smallest RMSE value of $0.04 \text{ m}^3 \text{ m}^{-3}$, while four sites (i.e., ES-Lju, US-Me2, US-Ton, and US-Var) exhibit a slightly higher RMSE value of $0.09 \text{ m}^3 \text{ m}^{-3}$. In addition to the internal error of model performance, it should be noted that the differences in SWC depth between the SiTHv2 model (first layer at 5 cm) and site measurements (depth varying across sites) may also contribute to uncertainties in such comparisons. Nonetheless, the *in-situ* comparisons indicate that the SWC estimates from the SiTHv2 model are reliable and successfully characterize soil moisture dynamics across different sites globally.

Comparison of soil moisture with data-driven product. To evaluate the global spatial pattern and magnitude of SiTHv2-derived SWC estimates, we employed the recently published global soil moisture product, SoMo.ml⁵⁸, which is built upon deep learning techniques (i.e., LSTM network) and numerous *in-situ* observations. Figure 9a illustrates the mean annual surface SWC spanning from 1982 to 2020, with tropical areas exhibiting the highest SWC and the high latitudes of Northern Hemisphere following closely. Excluding polar regions, zones surrounding 20°N and 20°S displays the lowest zonally mean SWC, attributed to the vast drylands and deserts in these areas. Furthermore, the surface SWC estimated by SiTHv2 shows good spatial agreement with SoMo.ml (Fig. 9b), with a R -value of 0.81 and an overall RMSE of $0.08 \text{ m}^3 \text{ m}^{-3}$. A high NSE value (0.45) also suggests high reliability when compared to SoMo.ml. Meanwhile, we analyzed the correlation between SiTHv2- and SoMo.ml-derived annual SWC at each grid cell (Fig. 9c), revealing that a large portion of the globe presents a high R -value, particularly exceeding 0.60 for the areas between 45°N and 45°S. Additionally, the annual variations of SWC in these two datasets are quite similar, achieving a R -value of 0.73. It is important to note that the comparison was conducted from 2000 to 2019, constrained by the temporal range of SoMo.ml. Overall, SiTHv2-based SWC estimates not only exhibit reasonable spatial patterns but also supply reliable annual dynamics across the majority of grid cells on a global scale.

Comparison of soil moisture with satellite product. In recent years, microwave remote sensing technology has become crucial for monitoring large-scale surface soil moisture via satellites, resulting in various satellite-based soil moisture products. Despite most products covering only the past dozen years due to the satellite launch, microwave remote sensing, based on the physical mechanism of dielectric constant, is currently the most direct and effective way for obtaining large-scale surface soil water conditions. It also serves as an important benchmark for evaluating the SWC estimates at a large scale. We compared surface SWC from SiTHv2 with four microwave-based SWC products: SMAP, SMOS, ASCAT, and AMSR2. SiTHv2-SWC shows high consistency with satellite-SWC products globally, particularly in arid and semi-arid regions, where the mean correlation coefficient (R) can reach 0.8 or higher (Fig. 10a). Notably, lower R -values are observed in tropical rainforests (e.g., Amazon) and high-latitude regions (above 60°N). This does not imply significant errors in SiTHv2-derived SWC in these

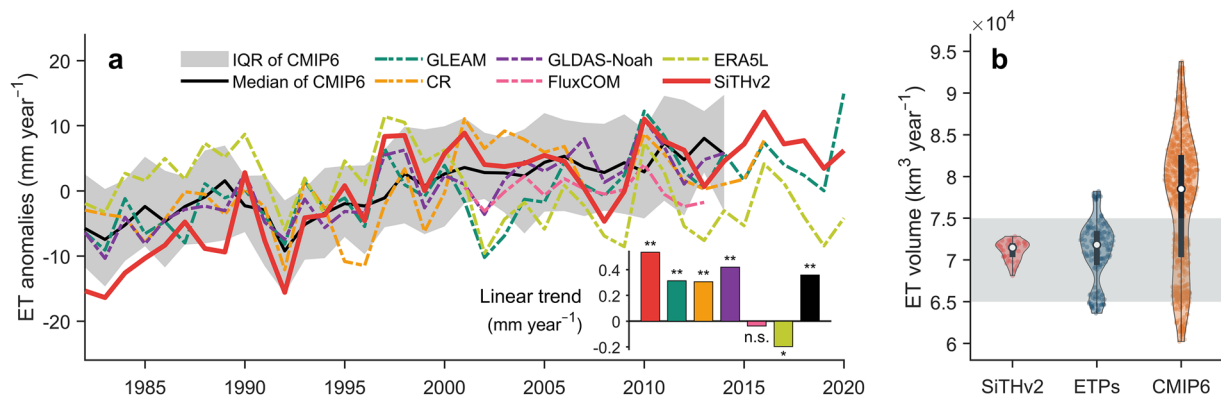


Fig. 6 Comparison of annual terrestrial ET from different data sets. **(a)** represents the annual variation and linear trend from 1982 to 2020, where the asterisks denote that the trend passed the Mann-Kendall test with a p -value < 0.01 ; **(b)** shows the annual total terrestrial ET volume from SiTHv2, mainstream global ET products (ETPs), and the CMIP6 simulations. A total of 20 ensemble ESMs from CMIP6 were used in this study; details can be found in Table S1.

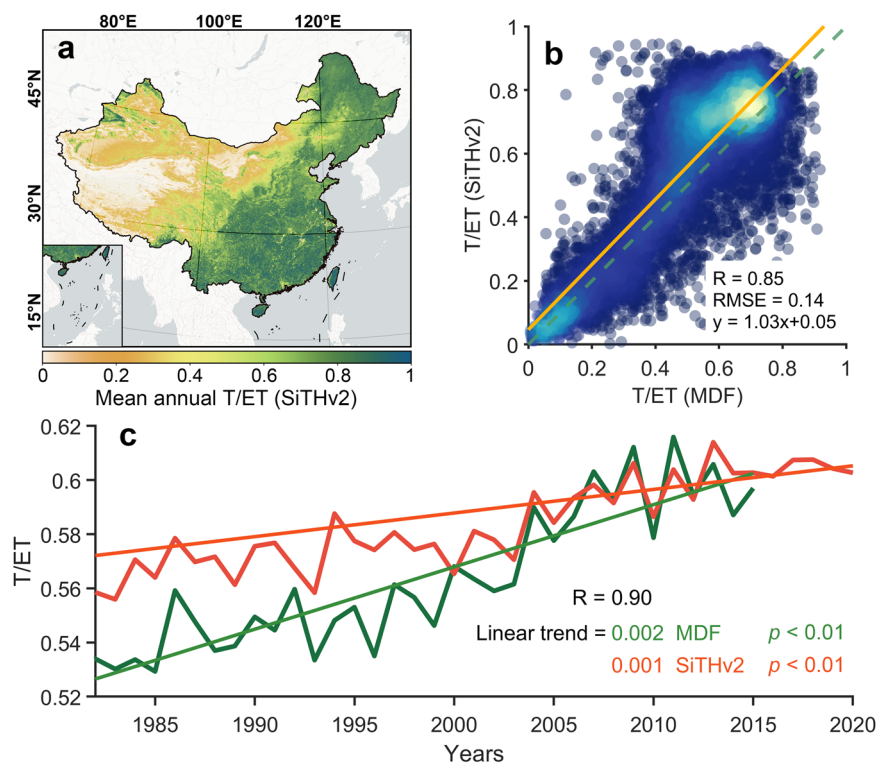


Fig. 7 Comparison of spatial patterns and annual variations between SiTHv2- and MDF-derived T/ET ratios over China. **(a)** depicts the spatial distribution of mean annual T/ET across China; **(b)** represents an intensity scatter plot with linear regression based on the grid-wise T/ET from SiTHv2 and MDF; **(c)** displays the annual variation of T/ET in China from SiTHv2 and MDF, respectively. The significance of the linear trend is based on the Mann-Kendall test with a p -value < 0.01 .

areas, as microwave retrievals of surface soil moisture are affected by factors such as dense vegetation and soil freeze-thaw processes, causing signal interference and larger errors in satellite-SWC accuracy in regions with forest and snow cover⁷⁹. Among the four satellite-based SWC products, SiTHv2 exhibits the highest correlation with SMAP, with a median R -value over 0.5 based on grid-by-grid statistics (Fig. 10b). Furthermore, the range of SWC values for SiTHv2 in terms of the 25th to 75th percentile is closer to SMAP and SMOS products, while AMSR2 produces the lowest results (Fig. 10c).

In addition, we calculated the interannual trend of surface SWC from 1982 to 2020 using the SWC estimates from SiTHv2. The global spatial distribution of the annual trend is displayed in Fig. 11, revealing considerable spatial differences in SWC changes. For example, soil moisture has notably increased in the Indian subcontinent

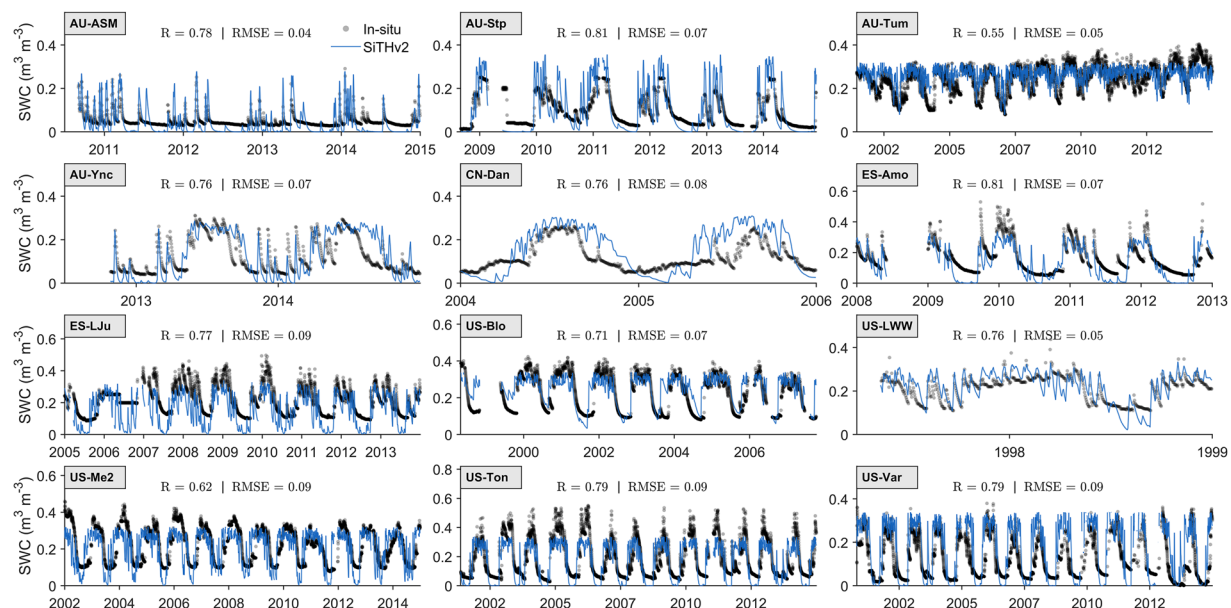


Fig. 8 Validation of soil moisture estimates (layer 1) obtained from the SiTHv2 model based on daily-scale surface soil moisture observations at 12 stations. The time span for different sites is adjusted according to the observation duration at each site.

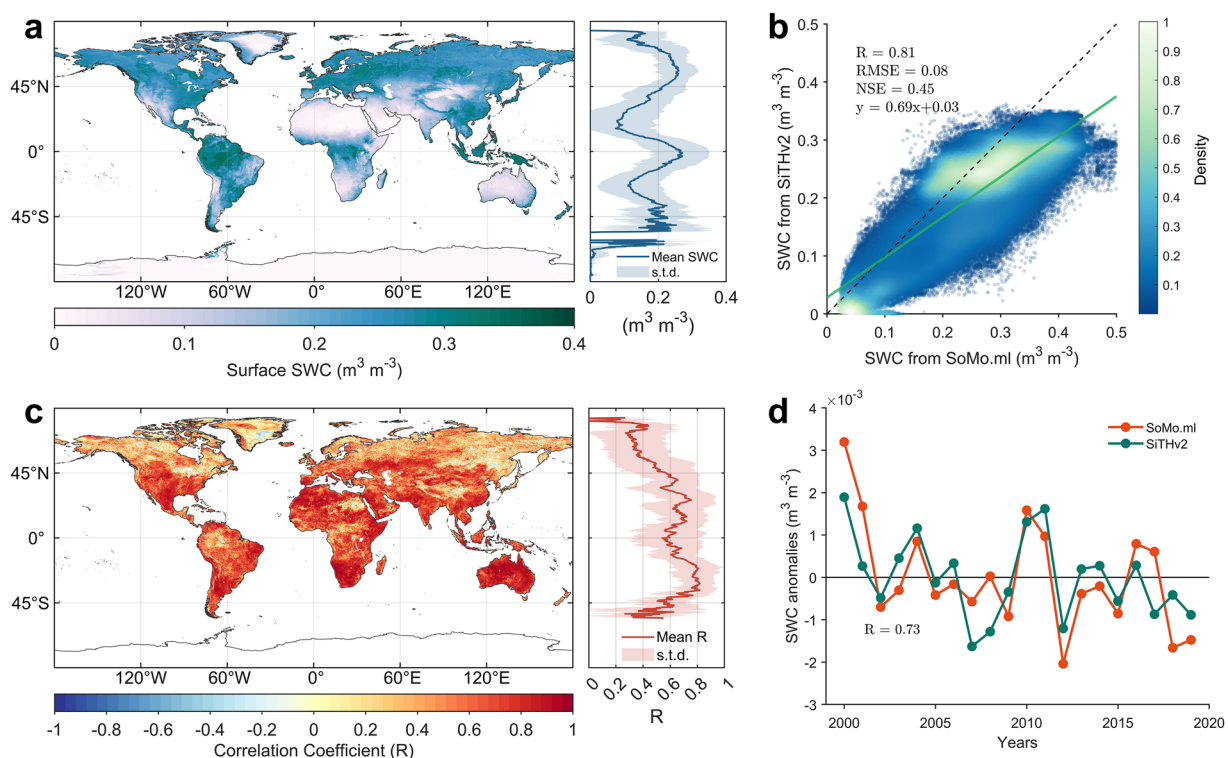


Fig. 9 Soil moisture estimate comparison between SiTHv2 and SoMo.ml. (a) Mean annual surface soil water content (SWC) from SiTHv2 spanning 1982 to 2020; (b) Mean annual spatial pattern regression between SiTHv2 and SoMo.ml during the overlapping period of 2000 to 2019; (c) Global correlation between SiTHv2 and SoMo.ml at each grid cell; (d) Annual variation of global SWC.

and the Tibetan Plateau, while arid and semi-arid regions like the western United States and southern Sahara have become increasingly dry. We selected eight typical regions with significant SWC increases or decreases over the past 39 years (1982–2020) for a daily-scale comparison of SiTHv2-derived SWC and other satellite

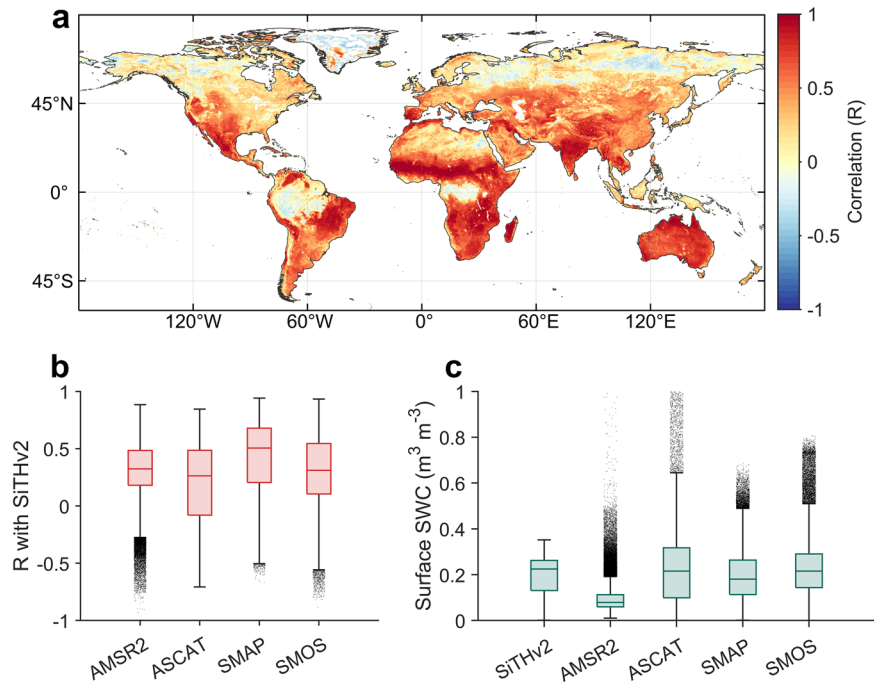


Fig. 10 Overall comparison of SWC derived from SiTHv2 with satellite-based products. **(a)** Spatial distribution of the mean correlation coefficient (R) between annual SWC from SiTHv2 and four satellite-based products (AMSR2, ASCAT, SMAP, and SMOS); **(b)** Boxplot shows the global statistics of R values for each combination between SiTHv2-SWC and different satellite-based SWC products; **(c)** The range distribution of SWC magnitudes from SiTHv2 and other satellite-based products.

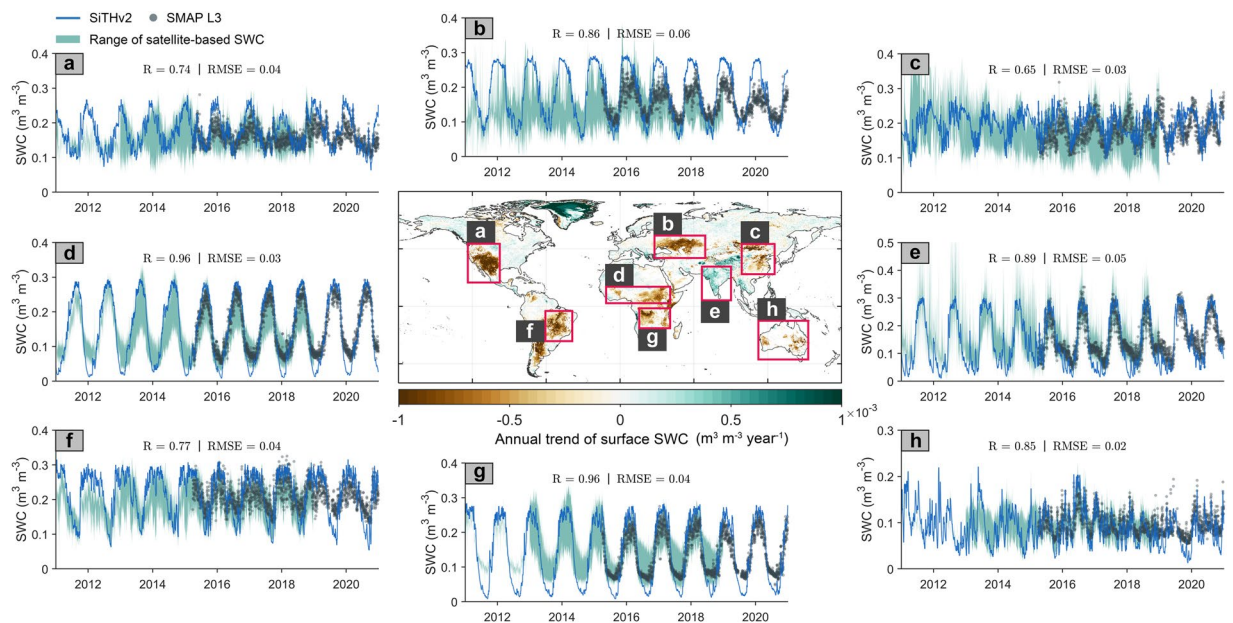


Fig. 11 Global pattern of the annual trend in SWC derived from SiTHv2 from 1982 to 2020; The subplots depict temporal comparisons (daily) between SiTHv2- and satellite-based SWC products in selected hotspot regions, include: **(a)** western United States; **(b)** Central Asia; **(c)** North China; **(d)** southern Sahara; **(e)** Indian subcontinent; **(f)** Brazilian Plateau; **(g)** southern Africa; and **(h)** Australia. The temporal variation of satellite-based SWC products starts from 2011, subject to the data availability (e.g., SMOS). The SMAP L3 data is highlighted with black dots, while the remaining satellite products are presented as a range of ensemble mean.

products. Statistically, SiTHv2 effectively captures SWC seasonality in these key regions, with R -value of at least 0.65 (Fig. 11c) and up to 0.96 (Fig. 11d,g), while RMSE values range from a maximum of 0.06 m³ m⁻³ (Fig. 11b)

to a minimum of $0.02 \text{ m}^3 \text{ m}^{-3}$ (Fig. 11h). Overall, the SiTHv2-derived SWC estimates closely align with microwave-based satellite observations, particularly showing higher consistency with the SMAP L3 product. Thus, the comparison with satellite products, combined with previous SoMo.ml statistics, collectively suggest that the global SWC estimates from SiTHv2 are reliable and can provide robust data support for hydrology- and climate-related applications.

Usage Notes

The original datatype running in the SiTHv2 model is the 32-bit double float, while the outputs are stored as 16-bit short integers with two decimal places to conserve storage space and facilitate transfer and usage. Consequently, the released dataset must be rescaled to obtain corrected values for each variable. Specifically, high-level software (e.g., Matlab, ArcGIS, etc.) or programming languages (e.g., Python, R, NCL, etc.) can be utilized to load data from the NetCDF file. Subsequently, the “scale_factor” attribute for each variable should be applied to retrieve the corrected values as follows:

$$Var_{\text{corrected}} = \text{double_float}(Var_{\text{original}}) \times \text{scale_factor}^{-1} \quad (20)$$

Although the newly SiTHv2-derived ET and soil moisture product perform well compared to similar data products, there are still limitations and uncertainties remain due to errors in forcing data and internal simulation processes. For instance, precipitation serves as a crucial water input in the SiTHv2 model, significantly influencing soil moisture and water supply for ET. However, current precipitation datasets, including the ERA5L precipitation used in this study, exhibit high uncertainties stemming from a scarcity of ground-based observations in certain regions⁸⁰. Additionally, the model does not emphasize the freeze-thaw process of soil at low temperatures in high latitudes, which may introduce errors in soil moisture (ice and water) during boreal winters^{68,69}, although the absolute ET amount does not undergo major changes at such low temperatures. Nonetheless, based on multi-scale validation and comparison, the SiTHv2-derived product is capable of offering reliable ET and soil moisture estimates. By featuring long-term time span and global seamless coverage with high spatio-temporal resolution, this dataset will provide a dependable solution and data support for global water cycle analysis, drought assessment, energy budget, and climate change related studies.

Code availability

The codes to process the data and generate the figures, and the details of selected flux sites are available at <https://github.com/kunlz/Codes.longterm.SiTHv2.product>. The model codes of SiTHv2 are available at <https://github.com/kunlz/SiTHv2>.

Received: 13 September 2023; Accepted: 16 April 2024;

Published online: 03 May 2024

References

1. Oki, T. & Kanae, S. Global hydrological cycles and world water resources. *Science* **313**, 1068, <https://doi.org/10.1126/science.1128845> (2006).
2. Trenberth, K. E., Fasullo, J. T. & Kiehl, J. Earth's global energy budget. *Bulletin of the American Meteorological Society* **90**, 311–324, <https://doi.org/10.1175/2008BAMS2634.1> (2009).
3. Fisher, J. B. *et al.* The future of evapotranspiration: Global requirements for ecosystem functioning, carbon and climate feedbacks, agricultural management, and water resources. *Water Resources Research* **53**, 2618–2626, <https://doi.org/10.1002/2016WR020175> (2017).
4. Yang, Y. *et al.* Evapotranspiration on a greening earth. *Nature Reviews Earth & Environment* **4**, 626–641, <https://doi.org/10.1038/s43017-023-00464-3> (2023).
5. Katul, G. G., Oren, R., Manzoni, S., Higgins, C. & Parlange, M. B. Evapotranspiration: A process driving mass transport and energy exchange in the soil-plant-atmosphere-climate system. *Reviews of Geophysics* **50**, <https://doi.org/10.1029/2011RG000366> (2012).
6. Fisher, J. B. *et al.* Ecotress: Nasa's next generation mission to measure evapotranspiration from the international space station. *Water Resources Research* **56**, e2019WR026058, <https://doi.org/10.1029/2019WR026058> (2020).
7. Zeng, Z. *et al.* Impact of earth greening on the terrestrial water cycle. *Journal of Climate* **31**, 2633–2650, <https://doi.org/10.1175/JCLI-D-17-0236.1> (2018).
8. Jung, M. *et al.* Recent decline in the global land evapotranspiration trend due to limited moisture supply. *Nature* **467**, 951–954, <https://doi.org/10.1038/nature09396> (2010).
9. Zhang, K. *et al.* Parameter sensitivity analysis and optimization for a satellite-based evapotranspiration model across multiple sites using moderate resolution imaging spectroradiometer and flux data. *Journal of Geophysical Research: Atmospheres* **122**, 230–245, <https://doi.org/10.1002/2016JD025768> (2017).
10. Zhang, K. *et al.* Parameter analysis and estimates for the modis evapotranspiration algorithm and multiscale verification. *Water Resources Research* **55**, 2211–2231, <https://doi.org/10.1029/2018WR023485> (2019).
11. Peng, L., Li, D. & Sheffield, J. Drivers of variability in atmospheric evaporative demand: Multiscale spectral analysis based on observations and physically based modeling. *Water Resources Research* **54**, 3510–3529, <https://doi.org/10.1029/2017WR022104> (2018).
12. Wang, K. & Dickinson, R. E. A review of global terrestrial evapotranspiration: Observation, modeling, climatology, and climatic variability. *Reviews of Geophysics* **50**, <https://doi.org/10.1029/2011RG000373> (2012).
13. Senéviratne, S. I. *et al.* Investigating soil moisture-climate interactions in a changing climate: A review. *Earth-Science Reviews* **99**, 125–161, <https://doi.org/10.1016/j.earscirev.2010.02.004> (2010).
14. Zhu, G. *et al.* Evaluating the complementary relationship for estimating evapotranspiration using the multi-site data across north china. *Agricultural and Forest Meteorology* **230–231**, 33–44, <https://doi.org/10.1016/j.agrformet.2016.06.006> (2016).
15. Zhang, K., Kimball, J. S. & Running, S. W. A review of remote sensing based actual evapotranspiration estimation. *Wiley Interdisciplinary Reviews: Water* **3**, 834–853, <https://doi.org/10.1002/wat2.1168> (2016).
16. Herbst, M., Rosier, P. T., McNeil, D. D., Harding, R. J. & Gowing, D. J. Seasonal variability of interception evaporation from the canopy of a mixed deciduous forest. *Agricultural and Forest Meteorology* **148**, 1655–1667, <https://doi.org/10.1016/j.agrformet.2008.05.011> (2008).

17. Fan, Y., Li, H. & Miguez-Macho, G. Global patterns of groundwater table depth. *Science* **339**, 940–943, <https://doi.org/10.1126/science.1229881> (2013).
18. Evaristo, J., Jasechko, S. & McDonnell, J. J. Global separation of plant transpiration from groundwater and streamflow. *Nature* **525**, 91–94, <https://doi.org/10.1038/nature14983> (2015).
19. Fan, Y., Miguez-Macho, G., Jobbágy, E. G., Jackson, R. B. & Otero-Casal, C. Hydrologic regulation of plant rooting depth. *Proceedings of the National Academy of Sciences* **114**, 10572–10577, [10/gb26pv](https://doi.org/10.1073/pnas.1616661114) (2017).
20. Zhang, K., Zhu, G., Ma, N., Chen, H. & Shang, S. Improvement of evapotranspiration simulation in a physically based ecohydrological model for the groundwater–soil–plant–atmosphere continuum. *Journal of Hydrology* **613**, 128440, <https://doi.org/10.1016/j.jhydrol.2022.128440> (2022).
21. Zhu, G. *et al.* Development and evaluation of a simple hydrologically based model for terrestrial evapotranspiration simulations. *Journal of Hydrology* **577**, 123928, <https://doi.org/10.1016/j.jhydrol.2019.123928> (2019).
22. Scanlon, T. M. & Kustas, W. P. Partitioning evapotranspiration using an eddy covariance-based technique: Improved assessment of soil moisture and land–atmosphere exchange dynamics. *Vadose Zone Journal* **11**, vzj2012.0025, <https://doi.org/10.2136/vzj2012.0025> (2012).
23. Chen, H. *et al.* Evaluation of evapotranspiration models using different lai and meteorological forcing data from 1982 to 2017. *Remote Sensing* **12**, 2473, <https://doi.org/10.3390/rs12152473> (2020).
24. Chen, H. *et al.* Uncertainties in partitioning evapotranspiration by two remote sensing-based models. *Journal of Hydrology* **604**, 127223, <https://doi.org/10.1016/j.jhydrol.2021.127223> (2022).
25. Zou, M. *et al.* Integrating eco-evolutionary optimality principle and land processes for evapotranspiration estimation. *Journal of Hydrology* **616**, 128855, <https://doi.org/10.1016/j.jhydrol.2022.128855> (2023).
26. Fan, Y. Groundwater in the earth's critical zone: Relevance to large-scale patterns and processes. *Water Resources Research* **51**, 3052–3069, <https://doi.org/10.1002/2015WR017037> (2015).
27. Maxwell, R. M. & Condon, L. E. Connections between groundwater flow and transpiration partitioning. *Science* **353**, 377–380, <https://doi.org/10.1126/science.aaf7891> (2016).
28. Clark, M. P. *et al.* Improving the representation of hydrologic processes in earth system models. *Water Resources Research* **51**, 5929–5956, <https://doi.org/10.1002/2015WR017096> (2015).
29. Liang, X. & Xie, Z. Important factors in land–atmosphere interactions: Surface runoff generations and interactions between surface and groundwater. *Global and Planetary Change* **38**, 101–114, [https://doi.org/10.1016/S0921-8181\(03\)00012-2](https://doi.org/10.1016/S0921-8181(03)00012-2) (2003).
30. Martínez-de la Torre, A. & Miguez-Macho, G. Groundwater influence on soil moisture memory and land–atmosphere fluxes in the Iberian peninsula. *Hydrology and Earth System Sciences* **23**, 4909–4932, <https://doi.org/10.5194/hess-23-4909-2019> (2019).
31. Barlage, M., Chen, F., Rasmussen, R., Zhang, Z. & Miguez-Macho, G. The importance of scale-dependent groundwater processes in land–atmosphere interactions over the central United States. *Geophysical Research Letters* **48**, e2020GL092171, <https://doi.org/10.1029/2020GL092171> (2021).
32. Wu, W.-Y. *et al.* Divergent effects of climate change on future groundwater availability in key mid-latitude aquifers. *Nature Communications* **11**, 3710, <https://doi.org/10.1038/s41467-020-17581-y> (2020).
33. Famiglietti, J. S. The global groundwater crisis. *Nature Climate Change* **4**, 945–948, <https://doi.org/10.1038/nclimate2425> (2014).
34. Wang, Y., Merlin, O., Zhu, G. & Zhang, K. A physically based method for soil evaporation estimation by revisiting the soil drying process. *Water Resources Research* **55**, 9092–9110, <https://doi.org/10.1029/2019WR025003> (2019).
35. Jackson, R. B. *et al.* A global analysis of root distributions for terrestrial biomes. *Oecologia* **108**, 389–411, <https://doi.org/10.1007/BF00333714> (1996).
36. Fang, H., Baret, F., Plummer, S. & Schaepman-Strub, G. An overview of global leaf area index (lai): Methods, products, validation, and applications. *Reviews of Geophysics* **57**, 739–799, <https://doi.org/10.1029/2018RG000608> (2019).
37. Moesinger, L. *et al.* The global long-term microwave vegetation optical depth climate archive (vodca). *Earth System Science Data* **12**, 177–196, <https://doi.org/10.5194/essd-12-177-2020> (2020).
38. Baret, F. *et al.* Geov1: Lai and fapar essential climate variables and fcover global time series capitalizing over existing products. part1: Principles of development and production. *Remote Sensing of Environment* **137**, 299–309, <https://doi.org/10.1016/j.rse.2012.12.027> (2013).
39. Liang, S. *et al.* The global land surface satellite (glass) product suite. *Bulletin of the American Meteorological Society* **102**, E323–E337, <https://doi.org/10.1175/BAMS-D-18-0341.1> (2021).
40. Beck, H. E. *et al.* Mswx: Global 3-hourly 0.1 bias-corrected meteorological data including near-real-time updates and forecast ensembles. *Bulletin of the American Meteorological Society* **103**, E710–E732, <https://doi.org/10.1175/BAMS-D-21-0145.1> (2022).
41. Muñoz-Sabater, J. *et al.* Era5-land: A state-of-the-art global reanalysis dataset for land applications. *Earth System Science Data* **13**, 4349–4383, <https://doi.org/10.5194/essd-13-4349-2021> (2021).
42. Hersbach, H. *et al.* The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society* **146**, 1999–2049, <https://doi.org/10.1002/qj.3803> (2020).
43. Winkler, K., Fuchs, R., Rounsevell, M. & Herold, M. Global land use changes are four times greater than previously estimated. *Nature Communications* **12**, 2501, <https://doi.org/10.1038/s41467-021-22702-2> (2021).
44. Wieder, W. R., Boehnert, J., Bonan, G. B. & Langseth, M. RegridDED harmonized world soil database v1.2. ORNL DAAC <https://doi.org/10.3334/ORNLDAAC/1247> (2014).
45. Baldocchi, D. *et al.* Fluxnet: A new tool to study the temporal and spatial variability of ecosystem–scale carbon dioxide, water vapor, and energy flux densities. *Bulletin of the American Meteorological Society* **82**, 2415–2434, [10.1175/1520-0477\(2001\)082<2415:FANTTS>2.3.CO;2](https://doi.org/10.1175/1520-0477(2001)082<2415:FANTTS>2.3.CO;2) (2001).
46. Pastorello, G. *et al.* The fluxnet2015 dataset and the oneflux processing pipeline for eddy covariance data. *Scientific Data* **7**, 225, <https://doi.org/10.1038/s41597-020-0534-3> (2020).
47. Foken, T. The energy balance closure problem: An overview. *Ecological Applications* **18**, 1351–1367 (2008).
48. Martens, B. *et al.* Glead v3: Satellite-based land evaporation and root-zone soil moisture. *Geoscientific Model Development* **10**, 1903–1925, <https://doi.org/10.5194/gmd-10-1903-2017> (2017).
49. Ma, N. & Szilagyi, J. The cr of evaporation: A calibration-free diagnostic and benchmarking tool for large-scale terrestrial evapotranspiration modeling. *Water Resources Research* **55**, 7246–7274, <https://doi.org/10.1029/2019WR024867> (2019).
50. Jung, M. *et al.* The fluxcom ensemble of global land–atmosphere energy fluxes. *Scientific Data* **6**, 74, <https://doi.org/10.1038/s41597-019-0076-8> (2019).
51. Rodell, M. *et al.* The global land data assimilation system. *Bulletin of the American Meteorological Society* **85**, 381–394, <https://doi.org/10.1175/BAMS-85-3-381> (2004).
52. Eyring, V. *et al.* Overview of the coupled model intercomparison project phase 6 (cmip6) experimental design and organization. *Geoscientific Model Development* **9**, 1937–1958, <https://doi.org/10.5194/gmd-9-1937-2016> (2016).
53. Schneider, U. *et al.* Evaluating the hydrological cycle over land using the newly-corrected precipitation climatology from the global precipitation climatology centre (gpcc). *Atmosphere* **8**, 52, <https://doi.org/10.3390/atmos8030052> (2017).
54. Daly, C. *et al.* Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *International Journal of Climatology* **28**, 2031–2064, <https://doi.org/10.1002/joc.1688> (2008).
55. Tapley, B. D. *et al.* Contributions of grace to understanding climate change. *Nature Climate Change* **9**, 358–369, <https://doi.org/10.1038/s41558-019-0456-2> (2019).

56. Ma, N., Szilagyi, J. & Zhang, Y. Calibration-free complementary relationship estimates terrestrial evapotranspiration globally. *Water Resources Research* **57**, e2021WR029691, <https://doi.org/10.1029/2021WR029691> (2021).
57. Ma, N., Zhang, Y. & Szilagyi, J. Water-balance-based evapotranspiration for 56 large river basins: A benchmarking dataset for global terrestrial evapotranspiration modeling. *Journal of Hydrology* **630**, 130607, <https://doi.org/10.1016/j.jhydrol.2024.130607> (2024).
58. O. S. & Orth, R. Global soil moisture data derived through machine learning trained with *in-situ* measurements. *Scientific Data* **8**, 170, <https://doi.org/10.1038/s41597-021-00964-1> (2021).
59. Pablos, M. *et al.* Influence of quality filtering approaches in bec smos l3 soil moisture products. In *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, 6941–6944, <https://doi.org/10.1109/IGARSS.2019.8900273> (2019).
60. Chan, S. *et al.* Development and assessment of the smap enhanced passive soil moisture product. *Remote Sensing of Environment* **204**, 931–941, <https://doi.org/10.1016/j.rse.2017.08.025> (2018).
61. Wagner, W. *et al.* The ascats soil moisture product: A review of its specifications, validation results, and emerging applications. *Meteorologische Zeitschrift* **22**, 5–33 (2013).
62. Kim, S., Parinussa, R. M., Liu, Y. Y., Johnson, F. M. & Sharma, A. A framework for combining multiple soil moisture retrievals based on maximizing temporal correlation. *Geophysical Research Letters* **42**, 6662–6670, <https://doi.org/10.1002/2015GL064981> (2015).
63. 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, [10.1175/1520-0493\(1972\)100<0081:OTAOSH>2.3.CO;2](https://doi.org/10.1175/1520-0493(1972)100<0081:OTAOSH>2.3.CO;2) (1972).
64. Schenk, H. J. & Jackson, R. B. Mapping the global distribution of deep roots in relation to climate and soil characteristics. *Geoderma* **126**, 129–140, <https://doi.org/10.1016/j.geoderma.2004.11.018> (2005).
65. Purdy, A. J. *et al.* Smap soil moisture improves global evapotranspiration. *Remote Sensing of Environment* **219**, 1–14, <https://doi.org/10.1016/j.rse.2018.09.023> (2018).
66. van Diepen, C., Wolf, J., van Keulen, H. & Rappoldt, C. Wofost: A simulation model of crop production. *Soil Use and Management* **5**, 16–24, <https://doi.org/10.1111/j.1475-2743.1989.tb00755.x> (1989).
67. Zhang, K. & Zhu, G. A global dataset of terrestrial evapotranspiration and soil moisture from 1982 to 2020. *National Tibetan Plateau Data Center* <https://doi.org/10.11888/terre.tpcd.300751> (2023).
68. Cao, B., Gruber, S., Zheng, D. & Li, X. The era5-land soil temperature bias in permafrost regions. *The Cryosphere* **14**, 2581–2595, <https://doi.org/10.5194/tc-14-2581-2020> (2020).
69. Cao, B., Wang, S., Hao, J., Sun, W. & Zhang, K. Inconsistency and correction of manually observed ground surface temperatures over snow-covered regions. *Agricultural and Forest Meteorology* **338**, 109518, <https://doi.org/10.1016/j.agrformet.2023.109518> (2023).
70. Miralles, D. G., De Jeu, R. A. M., Gash, J. H., Holmes, T. R. H. & Dolman, A. J. Magnitude and variability of land evaporation and its components at the global scale. *Hydrology and Earth System Sciences* **15**, 967–981, <https://doi.org/10.5194/hess-15-967-2011> (2011).
71. Zeng, Z. *et al.* A worldwide analysis of spatiotemporal changes in water balance-based evapotranspiration from 1982 to 2009. *Journal of Geophysical Research: Atmospheres* **119**, 1186–1202, <https://doi.org/10.1002/2013JD020941> (2014).
72. Zhang, K. *et al.* Vegetation greening and climate change promote multidecadal rises of global land evapotranspiration. *Scientific Reports* **5**, 75–77, <https://doi.org/10.1038/srep15956> (2015).
73. Zhang, Y. *et al.* Coupled estimation of 500 m and 8-day resolution global evapotranspiration and gross primary production in 2002–2017. *Remote Sensing of Environment* **222**, 165–182, <https://doi.org/10.1016/j.rse.2018.12.031> (2019).
74. Gu, C. *et al.* Partitioning evapotranspiration using an optimized satellite-based et model across biomes. *Agricultural and Forest Meteorology* **259**, 355–363, <https://doi.org/10.1016/j.agrformet.2018.05.023> (2018).
75. Lian, X. *et al.* Partitioning global land evapotranspiration using cmip5 models constrained by observations. *Nature Climate Change* **8**, 640, <https://doi.org/10.1038/s41558-018-0207-9> (2018).
76. Jasechko, S. *et al.* Terrestrial water fluxes dominated by transpiration. *Nature* **496**, 347–350, <https://doi.org/10.1038/nature11983> (2013).
77. Schlaepfer, D. R. *et al.* Terrestrial water fluxes dominated by transpiration: Comment. *Ecosphere* **5**, art61, <https://doi.org/10.1890/ES13-00391.1> (2014).
78. Niu, Z. *et al.* A spatial-temporal continuous dataset of the transpiration to evapotranspiration ratio in china from 1981–2015. *Scientific Data* **7**, 369, <https://doi.org/10.1038/s41597-020-00693-x> (2020).
79. Gruber, A., Lannoy, G. D., Albergel, C., Al-Yaari, A. & Wagner, W. Validation practices for satellite soil moisture retrievals: What are (the) errors? *Remote Sensing of Environment* **244**, 111806, <https://doi.org/10.1016/j.rse.2020.111806> (2020).
80. Jiang, Y. *et al.* Tphpr: A long-term (1979–2020) high-accuracy precipitation dataset (1/30, daily) for the third pole region based on high-resolution atmospheric modeling and dense observations. *Earth System Science Data* **15**, 621–638, <https://doi.org/10.5194/essd-15-621-2023> (2023).

Acknowledgements

This study was supported by the National Natural Science Foundation of China (grant nos. 42171019, 41871078, 42271029, and 42071045), and the IGSNRR Kezhen-Bingwei Youth Talents Program (2022RC003). We are also grateful to the principal investigators and their teams for all the forcing and comparison datasets used in this study.

Author contributions

K.Z. and G.Z. designed the study. K.Z. generated the dataset. K.Z., H.C. and N.M. analyzed the data. K.Z. wrote the paper and all authors reviewed and edited 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/s41597-024-03271-7>.

Correspondence and requests for materials should be addressed to K.Z. or G.Z.

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

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024