# A global terrestrial evapotranspiration product based on the three-temperature model with fewer input parameters and no calibration requirement

Leiyu Yu[1], Guo Yu Qiu[1], Chunhua Yan[1], Wenli Zhao[2, 3], Zhendong Zou[4], Jinshan Ding[1], Longjun Qin[1], Yujiu Xiong[5, 6]

[1]Shenzhen Engineering Laboratory for Water Desalinization with Renewable Energy, School of Environment and Energy, Peking University, Shenzhen, 518055, PR China
[2]Department of Earth and Environmental Sciences, Data Science Institute, Columbia University, New York, NY 10027, USA
[3]Max Planck Institute for Biogeochemistry, Hans-Knöll-Str. 10, Jena, 07745, Germany
[4]Shenzhen Investment Holdings Co., LTD, Shenzhen, 518048, PR China
[5]School of Civil Engineering, Sun Yat-Sen University, Guangzhou, 510275, PR China
[6]Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Zhuhai, 519082, PR China

*Correspondence to*: Yujiu Xiong (xiongyuj@mail.sysu.edu.cn)

**Abstract.** Accurate global terrestrial evapotranspiration (ET) estimation is essential to better understand Earth's energy and water cycles. Although several global ET products exist, recent studies indicate that ET estimates exhibit high uncertainty. With the increasing trend of extreme climate hazards (e.g., droughts and heat waves), accurate ET estimation under extreme conditions remains challenging. To overcome these challenges, we used 3-hour and 0.25° Global Land Data Assimilation System (GLDAS) datasets (net radiation, land surface temperature (LST), and air temperature) and a three-temperature (3T) model, without resistance and parameter calibration, in global terrestrial ET product development. The results demonstrated that the 3T model-based ET product agreed well with both global eddy covariance (EC) observations at daily (root mean square error (RMSE) = 1.1 mm day$^{-1}$, N=294058) and monthly scales (RMSE = 24.9 mm month$^{-1}$, N=9632) and basin-scale water balance observations (RMSE =116.0 mm yr$^{-1}$, N=34). The 3T model-based global terrestrial ET product was comparable to other common ET products, i.e., MOD16, P-LSH, PML, GLEAM, GLDAS, and Fluxcom, retrieved from various models, but the 3T model performed better under extreme weather conditions in croplands than did the GLDAS, attaining 9.0–20% RMSE reduction. The proposed daily and 0.25° ET product covering the period of 2001-2020, could provide periodic and large-scale information to support water cycle-related studies. The dataset is freely available at the Science Data Bank (http://doi.org/10.57760/sciencedb.o00014.00001, Xiong et al., 2022).

## 1 Introduction

Evapotranspiration (ET), the second-largest component of the global hydrological cycle (Trenberth et al., 2007), plays an important role in linking global energy and water cycles (Trenberth et al., 2009). ET is usually observed via techniques such as those involving evaporation pans, sap flowmeters, weighing lysimeters, stable isotopes, Bowen ratio systems, eddy

covariance (EC) systems, and scintillometers (Liu et al., 2022). However, these methods can only reflect ET representing the flux footprint of a given instrument (normally smaller than 1 km$^2$), which cannot provide spatial ET data for large-scale (e.g., basin and continental) studies. With the advancement of remote sensing (RS) technology, which can provide multiple information in regard to the land surface and atmosphere, remote estimation remains the most feasible and economic way to obtain continuous spatial ET data across field to global scales (Han et al., 2021; Zhang et al., 2016). Several global ET estimates have been developed over the past two decades based on various theories, including 1) surface energy balance residual methods, e.g., the ET product based on the Surface Energy Balance System (SEBS) (EB) (Chen et al., 2021); 2) Penman–Monteith (PM) and Priestley–Taylor (PT) equation-based methods, e.g., MOD16 (Mu et al., 2011), P-LSH (Zhang et al., 2015), PML (Zhang et al., 2019), and GLEAM (Martens et al., 2017; Miralles et al., 2011); 3) land surface models, e.g., the Global Land Data Assimilation System (GLDAS) (Rodell et al., 2004); 4) multimodel ensemble approach, e.g., GLASS (Yao et al., 2014), Hi-GLASS (Yao et al., 2017), and a synthesized ET product (Elnashar et al., 2021); and 5) empirical methods, e.g., Fluxcom (Jung et al., 2019). Although these ET products have been rigorously evaluated and widely applied, notable disagreement exists among these ET products. For example, Mueller et al. (2013) reported that the multi-year mean ET value retrieved from 40 ET products ranged from 423 to 563 mm yr$^{-1}$. In addition, while the interannual variation in some ET products exhibited similar change trends, inconsistent or even contrasting trends occurred among these ET products (Kim et al., 2021). The abovementioned phenomena indicate that high uncertainties remain in ET estimates and products (Fisher et al., 2017).

The uncertainty in ET estimates mainly originates from the quality of model input data, model (or algorithm) assumptions, and variable parameterization (Badgley et al., 2015; Cao et al., 2021; Khan et al., 2018; Vinukollu et al., 2011). In terms of model input datasets, meteorological data (i.e., relative humidity, RH, and wind speed, WS) are essential for most models. However, gridded meteorological data are generally produced via the data assimilation method based on limited ground observations, but simulation results may not necessarily capture real conditions, which could undoubtedly affect ET estimates (model output). For example, RH, directly affecting the vapour pressure deficit (VPD), retrieved from three meteorological reanalysis products exhibited a low correlation with in situ EC tower observations, with the coefficient of determination ranging from 0.005 to 0.09 (Cao et al., 2021). A similar problem exists between simulated and observed WS datasets (Vinukollu et al., 2011). In terms of model (or algorithm) assumptions, different descriptions of the ET process within the soil-plant-atmosphere continuum could yield single-layer versus multilayer models and incorrect but useful paradigms (Bonan et al., 2021; Raupach and Finnigan, 1988). Even though big-leaf models simplify the land surface as a homogeneous single layer, which is physically incorrect, they are recognized as highly computable and applicable models (Cheng et al., 2021). In contrast, multilayer models can more reasonable represent vertical vegetation and soil structures, but these models require more computational resources and additional hypotheses must be introduced to determine the model input or solve the model. This increase in model structure complexity and parameterization can increase the risk of error propagation or uncertainty in ET estimates, as revealed in the literature, e.g., Ershadi et al. (2015) and Zhao et al. (2020). For instance, under varying model assumptions and data availability levels, the surface resistance can be parameterized in different ways. In parameterization, several empirical coefficients and biophysical values required for resistance estimation must be calibrated. The error in ET

estimates based on the PM method and the difference between surface resistance values with and without calibration can range from 12% to 53% in terms of the mean absolute percentage error (MAPE) (Zhao et al., 2020). To reduce the above uncertainty in ET estimates, the PM equation was simplified as the PT model, by replacing the resistance terms with an empirical coefficient ($\alpha$) (Priestley & Taylor, 1972). Eventually, the combined uncertainty due to the model input data quality, model (or algorithm) assumptions, and variable parameterization schemes could lead to propagation errors in ET simulation results (Bengtsson and Shukla, 1988; Rienecker et al., 2011). Therefore, simpler algorithms without resistance parameterization (Yao et al., 2013; Yao et al., 2015) and variable calibration (Ma et al., 2021) requirements are necessary to reduce the uncertainty in ET estimates.

The three-temperature (3T) model, without calibration and resistance parameterization requirements, was proposed to reduce the uncertainty in ET estimates (Qiu, 1996). Based on the surface energy balance residual method, the inputs of the 3T model mainly comprise variables that can be directly measured or easily determined via RS, such as net radiation, surface temperature, and air temperature. The 3T model has been evaluated with an acceptable accuracy considering various land cover types on different spatial scales (Qiu et al., 1999; Wang et al., 2016; Xiong et al., 2019; Qiu et al., 2020; Zhao et al., 2020). Specifically, this model typically performs well in ET rate estimation in water-limited arid regions (Tian et al., 2013; Xiong et al., 2019), where surface and aerodynamic resistance values are very difficult to accurately estimate. Consequently, ET in these arid regions has usually been assumed as zero in certain ET products (Mu et al., 2011; Jung et al., 2019). In addition, the 3T model is sensitive to the temperature, and the model could potentially be suitable for ET estimation under notable temperature fluctuations (i.e., extreme heat or drought conditions). As such, the 3T model may provide an accurate dataset to support the attainment of Sustainable Development Goals (SDGs) (Guo et al., 2021) under an increasing frequency and intensity of extreme events (IPCC, 2022).

The objectives of this study were to 1) propose a global ET product with a low uncertainty based on the 3T model, 2) evaluate the product performance with global EC network and catchment water budget methods, 3) compare the established product to available mainstream ET products, and 4) explore the product suitability under extreme weather conditions, such as extreme heat and drought.

## 2 Materials and methods

### 2.1 Estimation of transpiration, evaporation, and evapotranspiration with the three-temperature model

The 3T model, proposed by Qiu (1996), comprises two equations for vegetation transpiration ($E_v$) and soil evaporation ($E_s$) calculation. This model mainly utilizes net solar radiation, surface temperature, and air temperature as model inputs. In this model, the resistance terms in the energy balance equation are eliminated via the introduction of a dry surface without evaporation or transpiration, as detailed in Qiu et al., (1999). In RS-based applications in which most pixels cannot represent pure vegetation or soil conditions, ET calculation depends on the fractional vegetation cover $f$, as follows (Xiong & Qiu, 2011):

$$LE_v = R_{n,c} - R_{n,cr} \frac{T_c - T_a}{T_{cr} - T_a} \qquad\qquad f = 1 , \qquad\qquad (1)$$

$$LE_s = R_{n,s} - G_s - \left(R_{n,sr} - G_{sr}\right)\frac{T_s - T_a}{T_{sr} - T_a} \qquad\qquad f = 0 , \qquad\qquad (2)$$

$$L(ET) = LE_v + LE_s \qquad\qquad 0 < f < 1 , \qquad\qquad (3)$$

where $LE$ (units: W m$^{-2}$) is the latent heat flux, $L$ (J kg$^{-1}$) is the latent heat of ET, $R_{n,c}$ and $R_{n,s}$ are the vegetation and soil net radiation components (W m$^{-2}$), respectively, $T_c$ and $T_s$ are the vegetation and soil surface temperatures (K), respectively, $T_a$ is the air temperature (K), and $G$ is the ground heat flux (W m$^{-2}$). The subscript r denotes the reference vegetation or soil.

## 2.2 Parameterization and datasets

The variables of the 3T model (Eqs. 1 to 3) can be parameterized as follows:

The net radiation ($R_n$) can be calculated by summing $R_{ns}$ and $R_{nl}$ (Eq. (4)), and the canopy and soil components, $R_{n,c}$ and $R_{n,s}$, respectively, can be calculated by partitioning $R_n$ based on the fractional vegetation cover via Eqs. (5) to (6) (Mu et al., 2007):

$$R_n = R_{ns} + R_{nl} , \qquad\qquad (4)$$

$$R_{n,c} = R_n \times f , \qquad\qquad (5)$$

$$R_{n,s} = R_n \times (1 - f) , \qquad\qquad (6)$$

The fractional vegetation cover, $f$, can be calculated according to the normalized difference vegetation index (NDVI) with Eq. (7) (Cleugh et al., 2007):

$$f = \frac{NDVI - NDVI_{min}}{NDVI_{max} - NDVI_{min}} , \qquad\qquad (7)$$

where $NDVI_{max}$ and $NDVI_{min}$ are threshold values, defined as the mean values of the lower and upper 5% positive terrestrial NDVI values, respectively.

The ground heat flux, G, can be directly extracted from net radiation $R_n$, according to Su (2002). Vegetation and soil component temperatures, $T_c$ and $T_s$, respectively, can be derived from the land surface temperature (LST) according to Lhomme et al. (1994), as described in Xiong et al. (2015).

In this study, $R_{ns}$, $R_{nl}$, $T_a$, and LST datasets were derived from the GLDAS (https://ldas.gsfc.nasa.gov/gldas/) with spatial and temporal scales of 0.25° and 3 hours (GLDAS_NOAH025_3H_2.1), respectively (Beaudoing and Rodell, 2020; Rodell et al., 2004). A monthly NDVI dataset with a spatial resolution of 0.05° was obtained from MOD13C2 (version 6) (Didan et al.,

2015) and resampled to 0.25° via the nearest neighbour method with the HEG tool (HDF-EOS to GeoTIFF Conversion Tool; https://lpdaac.usgs.gov/tools/heg/). Each dataset covered the 2001–2020 period (Table 1).

To remotely estimate ET at the watershed scale, Xiong and Qiu (2014) proposed a simple method to determine the reference temperature. Specifically, a pixel with the maximum temperature within a given watershed can be defined as the reference pixel. Once the reference pixel has been determined, the reference vegetation temperature, $T_{cr}$ (or reference soil temperature, $T_{sr}$), can be obtained with Eq. (8) (or Eq. (9)). In the global-scale application in this study, 31 terrestrial climate regions based on the Köppen–Geiger climate classification system (Kottek et al., 2006) were first divided into subregions via the principal component analysis (PCA) and K-means clustering methods, aiming to maintain relatively equivalent climate conditions within each subregion. Specifically, PCA was used to select major variables to describe regional characteristics from GLDAS meteorological factors (i.e., net radiation, air temperature, humidity, wind speed, precipitation, and air pressure) and land surface conditions (i.e., albedo, land surface temperature, NDVI, soil moisture, and soil temperature). Thereafter, these variables were used to classify the 31 climate regions through the K-means clustering method. Because of meteorology and land surface variation, the subregions varied from 90 to 110 in different months and a reference pixel could be determined in each subregion for applying the 3T model.

$$T_{cr} = T_{c,max} = \max(T_{c1}, T_{c2}, ..., T_{ci}) \ , \tag{8}$$

$$T_{sr} = T_{s,max} = \max(T_{s1}, T_{s2}, ..., T_{si}) \ , \tag{9}$$

where $T_{ci}$ and $T_{si}$ denote the vegetation surface and soil temperatures, respectively, in pixel $i$ ($i$=1, 2, 3…) within each subregion. The reference net radiation values of the soil and vegetation components, $R_{n,sr}$ and $R_{n,cr}$, respectively, were assumed as mean $R_{n,c}$ and $R_{n,s}$ values, respectively, within the same subregion corresponding to pixels of the upper 5% $T_s$ and $T_c$ values, respectively.

$$R_{n,sr} = \overline{R_{n,s}\left[T_{s,upper5\%}\right]} = \mathrm{mean}\left(R_{n,s1}, R_{n,s2}, ..., R_{n,sj}\right) \ , \tag{10}$$

$$R_{n,cr} = \overline{R_{n,c}\left[T_{c,upper5\%}\right]} = \mathrm{mean}\left(R_{n,c1}, R_{n,c2}, ..., R_{n,cj}\right) \ , \tag{11}$$

where $R_{n,sj}$ and $R_{n,cj}$ denote the soil and vegetation net radiation values, respectively, corresponding to pixel $j$ ($j$=1, 2, 3…) of the upper 5% $T_s$ and $T_c$ values, respectively, within the same subregion.

The daytime ET was considered in this study. In global-scale applications, the daytime can be defined based on 3-hourly GLDAS net radiation values higher than 100 W m$^{-2}$. Then, all the 3-hour LE (or ET) estimates can be arithmetically averaged (or summed) into daily, monthly, and annual values.

**2.3 Evaluation of the performance of the 3T model**

ET values estimated with the 3T model were assessed on three scales due to challenges in the validation of RS-based ET estimates (Vinukollu et al., 2011; Miralles et al., 2016; Liu et al., 2016). First, ET estimates at daily and monthly scales were validated against in situ observations retrieved from global EC flux towers covering various land cover types, as widely applied in other studies, e.g., Chen et al. (2016) and Ma et al. (2021). Due to a mismatch between the flux tower footprint and pixel resolution (0.25° in this study), mean ET values in different watersheds were compared to those obtained from the water balance equation on a yearly scale. EC-based and basin-scale water budget-based validation methods are considered the most reliable and commonly used methods. Finally, ET estimates were compared to several gridded ET products on a multi-year average scale. Statistical analysis, including Pearson's correlation coefficient (r), relative bias (RB), and root mean square error (RMSE), was employed in assessment.

**2.3.1 Evaluation via the global eddy covariance network**

ET observations of 126 flux towers within the FLUXNET network (https://fluxnet.org/) were selected (Fig. 1a), and the selection process was conducted according to the following criteria: 1) a given flux tower should exhibit stable operation conditions for at least 2 consecutive years since 2001; 2) the latent heat flux (LE) was subjected to energy closure correction, and the percentage of good-quality measurement and gap fill data should be higher than 0.7; 3) the land cover within each 0.25° grid pixel containing a tower should be as homogeneous as possible (Zhang et al., 2019; Ma et al., 2021). The selected 126 EC towers were located at 26 evergreen needle leaf forest (ENF), 25 grassland (GRA), 15 cropland (CRO), 15 wetland (WET), 13 deciduous broadleaf forest (DBF), 8 evergreen broadleaf forest (EBF), 7 open/closed shrubland (OSH/CSH), 6 mixed forest (MF), 6 woody savanna (WSA), and 5 savanna (SAV) sites globally. Pixel-scale ET estimates based on the EC tower location were compared to EC tower observations.

**2.3.2 Evaluation considering the water budget in global main catchments**

The catchment ET ($ET_{wb}$), based on the water balance equation, has been recognized as a highly robust and credible method, particularly in relatively large catchments, on a multi-year (more than 10 years) scale (Liu et al., 2016). Hence, 34 catchments (Fig. 1b) were selected based on the following 2 criteria: 1) the basin area should be larger than 100000 km$^2$ to minimize uncertainties in the measurement of water balance equation components in relatively small basins; 2) the available basin data should cover more than ten years since 2001. $ET_{wb}$ can be calculated with Eq. (12):

$$ET_{wb} = P - R - \Delta S \ , \tag{12}$$

where $P$, $R$ and $\Delta S$ are the precipitation (mm yr$^{-1}$), runoff (mm yr$^{-1}$), and terrestrial water storage change (mm yr$^{-1}$), respectively, in a given catchment. Annual $\Delta$S can be calculated as the terrestrial water storage anomaly (TWSA) difference between Decembers of the target year and its previous year. Monthly 0.25°-resolution $P$ data (full monthly data version 2020) were

downloaded from the Global Precipitation Climatology Center (GPCC, http://gpcc.dwd.de/) (Schneider et al., 2020). *R* data were acquired from the Global Runoff Data Center (GRDC, https://portal.grdc.bafg.de/). Monthly 0.5°-resolution TWSA data were obtained from the JPL Mascon RL06 version 2.0 GRACE dataset (Watkins et al., 2015).

### 2.3.3 Evaluation via comparison to other commonly used global ET products

At the global scale, 6 commonly used ET products retrieved from different methods were selected for inter-comparison. Among the selected ET products, three products were based on the PM model with varying resistance parameterization schemes, i.e., MOD16 (version 6, Mu et al., 2011), P-LSH (Zhang et al., 2015), and PML (version 2, Zhang et al., 2019), while the remaining three products were based on the PT model (GLEAM version 3.5a; Miralles et al., 2011; Martens et al., 2017), land surface models (GLDAS version 2.1, Beaudoing and Rodell, 2020; Rodell et al., 2004), and machine learning (Fluxcom; Jung et al., 2019). All products were first resampled to 0.25° via the nearest neighbour method before comparison. Datasets covering the 2003–2013 period were used to maintain the above ET products. In the comparison process, non-vegetated areas (please refer to the Fluxcom product) were excluded due to the absence of ET data in certain products, such as the Fluxcom and MOD16 products.

## 3 Results

### 3.1 Performance of the 3T product versus the global EC network

At the daily scale, the 3T model-based ET estimates agreed well with the observation (N=294058), with RMSE of 32 W m$^{-2}$ (or 1.1 mm day$^{-1}$) (Fig. 2a), which was comparable to other ET products, such as GLDAS (RMSE: 32 W m$^{-2}$ or 1.1 mm day$^{-1}$) (Fig. S1), PML (RMSE: 0.7 mm day$^{-1}$) (Zhang et al., 2019), and SEBS (RMSE: 1.6 mm day$^{-1}$) (Chen et al., 2021). Moreover, the 3T model could capture the change trend of daily ET because comparison results at 10 EC sites covering various biomes in both the Southern (Figs. 2b-2c) and Northern Hemispheres (Figs. 2d-2k) indicate that interannual variability of the estimates were close to that of the observed ET. A comparison at an instantaneous 3-hour scale was also performed to test the ET estimates. EC observations across the world for the 15th day of each month in 2011 (N=6278) were compared because the data are too large to perform an entire comparison at a global scale. Although the RMSE (74 W m$^{-2}$) was a slightly greater than that at daily scale (Fig. S2a), the 3T model-based ET estimates at the 3-hour scale agree well with the GLDAS ET, with an r of 0.89 and RMSE of 21 W m$^{-2}$ (Fig. S2b). The explanation is likely that high temporal data may encounter missing values, which complicates the comparison.

At the monthly scale, the paired ET values between the 3T model and EC observations were generally distributed on both sides of the 1:1 line, revealing relatively large differences at a few points for ET values higher than 100 mm month$^{-1}$ and resulting in regression line slope and r values of 0.75 and 0.80, respectively (Fig. 3a). The RMSE and RB values between the ET estimates and EC-based observations reached 22.85 mm month$^{-1}$ and -1.2%, respectively. If monthly data were compared, similar results could be obtained, with an RMSE value of 24.90 mm month$^{-1}$, an RB value of 0.7%, and regression line slope

and r values of 0.75 and 0.78, respectively (Fig. 3b). The errors in the 3T model-based ET estimates were comparable to those in the other ET products (please refer to section 3.3 for details). For example, compared to EC observations, the RMSE and r values of an ET product retrieved from the process-based Breathing Earth System Simulator (BESS) model reached 23.4 mm month[-1] and 0.79, respectively (Jiang and Ryu, 2016). These results indicate that the ET product developed based on the 3T model agreed well with global EC observations at multi-temporal scales.

The performance of the 3T model in the different biomes was further analysed (Figs. 3c-3l). Due to data point separation in Fig. 3b, the results shown in Figs. 3c-3l are similar to those shown in Fig. 3b, with slight differences among the various biomes. The 3T model performed the best at forest sites because the paired data points were more closely distributed along the 1:1 line, with slope values ranging from 0.81 to 1.05, whereas the r values ranged from 0.75 to 0.85 (Fig. 3e-3h)). Among the different forest cover types, the ET estimates at the MF and ENF sites exhibited a lower uncertainty, with RMSE and RB values of 20.3 mm month[-1] and 13.3%, respectively, at the former sites and values of 22.8 mm month[-1] and 10.4%, respectively, at the latter sites, followed by DBF (RMSE=24.2 mm month[-1] and RB=25.8%) and EBF sites (RMSE=29.7 mm month[-1] and RB=4.6%). The 3T model performance at the shrubland sites was similar to that at the MF sites (RMSE=20.6 mm month[-1] and RB=9.2%) but with lower slope and r values of 0.53 and 0.60, respectively (Fig. 3i). At the sites of the remaining land use/land cover (LULC) types, the 3T model yielded lower ET estimates than the EC observations as the RB value ranged from -6.1% to -33.8% (Figs. 3c, 3d, and 3j to 3l). Among these sites, the 3T model exhibited the lowest bias at the GRA sites, with slope, r, and RMSE values of 0.71, 0.82, and 21.4 mm month[-1], respectively (Fig. 3j), followed by SAV (slope=0.67, r=0.83, and RMSE=27.7 mm month[-1]) (Fig. 3k), CRO (slope=0.61, r=0.78, and RMSE=27.9 mm month[-1]) (Fig. 3c), WET (slope=0.65, r=0.74, and RMSE=28.3 mm month[-1]) (Fig. 3d), and WSA sites (slope=0.49, r=0.70, and RMSE=31.7 mm month[-1]) (Fig. 3l). The 3T model performance among the different biomes, with a maximum RMSE value of 31.7 mm month[-1], was comparable to that of the other methods based on the above comparison to EC observations, with RMSE values ranging from 30 to 42.9 mm month[-1], as reported by Carter and Liang (2018), Zhang et al. (2019), and Peng et al. (2021). These results suggest that the 3T model performed with an acceptable accuracy across the various biomes.

## 3.2 Performance of the 3T product versus the water budget in global catchments

Multi-year (2003-2013) average ET values for 34 relatively large watersheds were obtained with the 3T model and compared to water balance ET ($ET_{wb}$) data. The estimated mean ET value was 514.5 mm yr[-1], with a standard deviation of 211 mm yr[-1], whereas the mean $ET_{wb}$ value reached 476.5 ± 280 mm yr[-1]. The mean ET difference reached only 38 mm yr[-1], indicating that the ET estimates obtained with the 3T model were similar to the $ET_{wb}$ values. The scatter plots shown in Figs. 4a and 4b at multi-year and annual scales, respectively, also confirmed that these two types of ET values agreed well, with r values of 0.94 and 0.91, respectively. The regression line slope at the multi-year scale was 0.71, with RMSE and RB values of 116 mm yr[-1] and 8.0%, respectively (Fig. 4a), whereas the values reached 0.69, 128 mm yr[-1], and 9.1%, respectively, at the annual scale (Fig. 4b). Fig. 4c shows the 3T model performance in each watershed in terms of RB. The RB values in nearly 70% watersheds were relatively low, within ±25%, indicating satisfactory performance of the 3T model in these watersheds. However, the 3T

model overestimated ET in approximately 21% of all watersheds with RB values greater than 60% (the red colour in Fig. 4c). These river basins were mainly located at high latitudes (approximately 60° North) with relatively low $ET_{wb}$ values (133 ± 50 mm yr$^{-1}$). ET overestimation in these regions was not only observed in this study but also observed in other ET comparison-based studies, such as Ma et al. (2021). A possible reason for the higher uncertainty may be that a higher bias occurs in the hydrological (e.g., runoff) and gridded meteorological (e.g., precipitation) data employed in the water balance equation due to the scarcity of in situ observational stations in these regions (Ma et al., 2021). Nonetheless, the above results generally suggest that the 3T model performance was comparable to that of the water balance equation.

**3.3 Comparison of the 3T product to other global ET products**

To further assess the performance of the 3T model across the various terrestrial land types, 3T model-based ET estimates were cross validated against six global ET products during the 2003–2013 period.

When EC observation data were adopted as a reference, 3T model-based ET estimates were comparable to GLDAS, GLEAM, and MOD16 data in terms of r and RMSE, with values of 0.8 and 22 mm month$^{-1}$, respectively (Fig. 3a and Figs. 5a to 5c). Although the slope of the regression line (0.75, as shown in Fig. 3a) between the 3T model-based ET estimates and observations was slightly lower than that between the ET estimates and GLDAS (0.83, as shown in Fig. 5a) and GLEAM data (0.79, as shown in Fig. 5b) and slightly higher than that between the ET estimates and MOD16 data (0.73, as shown in Fig. 5c), the absolute RB value of the 3T model was lower than that of the GLDAS (1.9%, as shown in Fig. 5a), GLEAM (2.7%, as shown in Fig. 5b), and MOD16 products (-4.7%, as shown in Fig. 5c). The remaining three products, i.e., Fluxcom, PML, and P-LSH, exhibited limited comparative advantages, with an r value of 0.9, a slope higher than 0.8 (0.92, 0.84, and 0.83, respectively), and an RMSE value ranging from 16.9 to 18.6 mm month$^{-1}$.

When $ET_{wb}$ values were adopted as a reference, although the 3T model performance was slightly lower than that of the PML, GLEAM, and P-LSH products in terms of RMSE, with a value of 116 mm month$^{-1}$ versus values of 96, 111, and 115 mm month$^{-1}$, respectively (Fig. 4a and Figs. 6a to 6c), the 3T model performed better than did the GLDAS (RMSE=120 mm month$^{-1}$), Fluxcom (RMSE=149 mm month$^{-1}$), and MOD16 products (RMSE=182 mm month$^{-1}$) (Figs. 6d to 6f). In terms of the regression line between the ET estimates and $ET_{wb}$, except for the relatively low performance of MOD16, for slope and r values of 0.58 and 0.77, respectively, r values of the other ET products were greater than 0.94 and exhibited a slight difference, with a maximum difference of 0.03, but the slope (0.71, as shown in Fig. 4a) of the regression line between the $ET_{wb}$ values and 3T model-based ET estimates was lower than that between the $ET_{wb}$ values and Fluxcom (1.13), P-LSH (0.92), PML (0.83), GLDAS (0.80), and GLEAM data (0.76) (Fig. 6). However, the absolute RB, with a value of 8% (Fig. 4a), of the 3T model was the smallest, while the absolute RB values of the other six products were greater than 8%, ranging from 8.2% to 21.8% (Fig. 6).

Via comparison of the terrestrial (excluding Antarctica) ET values retrieved from the various ET products, the mean ET value of the 3T model reached 546 mm yr$^{-1}$ during the 2003–2013 period, whereas the mean ET values obtained with the MOD16, PML, GLEAM, Fluxcom, GLDAS, and P-LSH products reached 468, 542, 544, 549, 551 and 551 mm yr$^{-1}$, respectively.

275    In terms of interannual variation (excluding Antarctica, Greenland, and desert areas according to Jung et al. (2019)), the 3T model-based estimates were similar to the other six ET products with an increasing trend from January (approximately 40 mm month$^{-1}$) to July (approximately 65 mm month$^{-1}$) and then a decreasing trend through the following months (Fig. 7a). The latitudinal distribution of the values obtained with each ET product was also determined, and the changing trend of the 3T model-based ET values was similar to that of the values obtained with the six considered ET products (Fig. 7b). Specifically,

280    the highest terrestrial ET values occurred at the equator, with values ranging from 1251 to 1390 mm yr$^{-1}$ (1293 mm yr$^{-1}$ for the 3T model), and the ET value decreased towards the North and South Poles. In the Northern Hemisphere, ET attained a second peak at approximately 20º, with values ranging from 934 to 1111 mm yr$^{-1}$ (950 mm yr$^{-1}$ for the 3T model), whereas a third peak occurred from 37º to 45º (the third peak varied among the different ET products) in the Southern Hemisphere, with values ranging from 562 to 706 mm yr$^{-1}$ (690 mm yr$^{-1}$ for the 3T model). This ET peak distribution trend was correlated with

285    the global vegetation distribution. However, it should be noted that the ET values obtained with the 3T model were generally lower than those obtained with the ET products between approximately 30º south and 45º north (except MOD16), and a large discrepancy in ET estimates occurred, particularly between approximately 17º south and 17º north, where the difference could exceed 350 mm yr$^{-1}$. These results suggest that even though the ET products are similar, the ET estimates in certain areas may differ, and uncertainty may exist in ET estimates in these regions.

290    Pixel-by-pixel comparison of the various ET products was also conducted. To overcome the influence of the resampling method on the obtained ET values, only GLDAS and GLEAM data, sharing the same spatial resolution with the 3T model-based estimates (0.25°), are shown in Fig. 8. The left panel shows the global land ET distribution, and the 3T model-based ET values generally exhibited a similar distribution to that of the ET values obtained with the two ET products, as shown in Fig. 8. However, obvious differences existed, especially in arid regions such as Sahara, Middle East, Mongolia, and southeast of

295    the Qinghai-Tibet Plateau, where the 3T model-based ET estimates were higher than the values obtained with the two ET products. The scatter plots in the right panel of Fig. 8 reveal that the 3T model-based ET estimates were very similar to the GLDAS-based ET values. Moreover, the slope of the regression line between the 3T model- and GLDAS-based ET values was 0.93, with r and RMSE values of 0.95 and 114.6 mm yr$^{-1}$, respectively (Fig. 8d), whereas the values reached 0.89, 0.94 and 130.6 mm yr$^{-1}$, respectively, between the 3T model and GLEAM (Fig. 8e). RMSE and r statistics between the 3T model-

300    based ET estimates and the values obtained with each ET product were visualized in a heatmap (Fig. 9), in which the darker the blue colour is, the higher the r value and the lower the RMSE value. Overall, the 3T model-based ET product is consistent with the other 6 products with r ranging from 0.89 (compared to MOD16) to 0.96 (compared to GLDAS) and RMSE ranging from 108.5 (compared to GLDAS) to 177.7 (compared to MOD16) mm yr$^{-1}$. Interestingly, it was obvious that the 3T model-based, GLDAS, and PML products with the same model inputs were highly consistent according to the higher r and lower

305    RMSE values (the corresponding blue cubes are in the bottom left of Fig. 9), while the ET products calibrated or upscaled based on EC towers, i.e., PML, P-LSH, and Fluxcom, were highly consistent (the corresponding blue cubes are in the top right of Fig. 9).

The abovementioned results indicate that the 3T model-based ET estimates were comparable to the data obtained with the commonly used global ET products.

## 4 Discussion

### 4.1 Characteristics of the global terrestrial ET product based on the 3T model

As indicated in section 3, the 3T model-based global terrestrial ET product agreed well with ground observations and was comparable to other commonly used ET products. Particularly, the determined global terrestrial (excluding Antarctica) ET volume (in units of $10^3$ km$^3$ yr$^{-1}$) based on 3T model-based estimates reached 73.8 from 2003-2013, which is not only consistent with that determined based on the other ET products (excluding MOD16), as indicated in section 3.3, ranging from 73.2 to 74.5 (Table 3), but also consistent with values reported in other studies, e.g., 72.3±0.9, as obtained with a complementary relationship-based ET product from 1982–2016 (Ma et al., 2021), and 71.1, as determined with a water balance and machine learning-based ET product from 1982–2009 (Zeng et al., 2014).

It should be noted that the 3T model differed from the methods used to estimate ET in the adopted ET products, as listed in Table 2. In particular, the 3T model excludes resistance and requires no parameter calibration. Resistance terms are unavoidable in PM models, which could lead to a high uncertainty in ET estimates (Zhao et al., 2020; Cao et al., 2021). As described in section 3.3, the MOD16, P-LSH, and PML products, based on the PM equation with varied resistance parameterization methods, exhibited obvious differences (e.g., 300 mm yr$^{-1}$ at a few locations, as shown in Fig. 7b) and performed differently via a comparison to ground observations. This occurred because the canopy resistance is difficult to estimate, in addition to empirical relationship adopted in the estimation process. These empirical equations are site- and biome-specific equations and normally require calibration (Mu et al., 2011; Zhang et al., 2015; Zhang et al., 2019). Because a large number of EC towers sites are used for calibration in resistance estimation in P-LSH and PML, these two products performed better than MOD16 using only 46 sites (Figs. 5 and 6). Nonetheless, calibration typically requires observational data, while limited in situ observations restrict accurate calibration of biome-specific coefficients on a global scale. A recent study confirmed that models requiring no calibration could decrease the uncertainty in global ET estimates (Ma et al., 2021). The obtained results indicate that the 3T model-based ET product achieved a lower uncertainty than that achieved by MOD16 retrieved from the PM equation with a complex resistance parameterization scheme and limited calibration and that the 3T model-based ET product was comparable to P-LSH and PML developed from the PM equation with adequate calibration during resistance parameterization.

Although the 3T model-based ET estimates suffer from the domain size when determining the reference site, the uncertainty may be limited. We tested the difference between the 3T model-based ET values estimated using two different regimes with different subregions and sizes. Specifically, Köppen-Geiger climate regimes with 31 subregions and detailed subregions with numbers of 90-110 were used. In general, the two groups of daily ET estimates in 2011 showed little difference, with mean ET values of 47 and 42 W m$^{-2}$, respectively and were close to the EC observation, with RMSE values of 32 and 33 W m$^{-2}$

(Figs. 10a and 10b, respectively). At a yearly scale, however, the 3T model-based ET estimates from 90-110 subregions (Fig. 10d) were much closer to the water balance ET than those estimates using 31 sub-regions (Fig. 10c). The results indicate that the smaller the domain size of a region where the reference parameters were obtained, the more accurate the 3T model, which is consist with our previous findings (Xiong et al., 2015, 2019).

In addition, the 3T model-based global terrestrial ET product required less data in terms of model inputs than those required by the adopted ET products, as listed in Table 2. Specifically, the 3T model requires net radiation, soil heat flux, air temperature, LST, and vegetation index (i.e., NDVI) data to decompose the radiation (or LST) components of vegetation and soil. For example, PM-based ET estimation requires wind speed and VPD data, in addition to net radiation, soil heat flux, and air temperature data. However, wind speed and VPD data, especially the former, exhibit high heterogeneity in space, and current commonly used reanalysis datasets contain high uncertainty, e.g., the difference in wind speed can exceed 5 m s$^{-1}$ among several products (Yang et al., 2019), thus increasing bias in global ET products. While a model with a higher complexity may better describe the ET process, a satisfactory model performance normally depends on abundant data, not only regarding model inputs but also regarding model (or parameter) calibration (Medici et al., 2012; Wu et al., 2020). Otherwise, a relatively simple model with fewer input datasets could be more reasonable, e.g., the GLEAM product based on the PT method, a simplified version of the PM equation, outperformed the PM-based MOD16 product in this and other studies (e.g., Cao et al., 2021). Although the performance of the 3T model-based ET product was similar to that of the GLEAM product, an empirical parameter, namely, the PT coefficient, is required in the PT-based GLEAM product. In estimation with the GLEAM product, the PT coefficient was set to 0.8 for tall canopies and 1.26 for short vegetation and bare soil, respectively (Miralles et al., 2011), but the value varies among the different biomes (e.g., Komatsu 2005), especially on a short time scale (daily) (Guo et al., 2015). In fact, the input datasets of the 3T model are commonly available with an adequate credibility (Bao & Zhang, 2013; Cao et al., 2022; Fu & Wang, 2014; Ji et al., 2015; Peng et al., 2019; Xu et al., 2019; Zhang et al., 2016; Zhou et al., 2017), resulting in easy model application.

## 4.2 The 3T model-based terrestrial ET product for extreme weather condition monitoring

Although the 3T model is the most sensitive to the LST among the various model inputs (Xiong and Qiu, 2011), this characteristic may result in a suitable model ability in capturing ET variation during extreme-temperature events such as heatwaves and flash droughts. Since both the frequency and damage extent of heatwaves and flash droughts are increasing, these extreme weather conditions have attracted extensive attention worldwide (IPCC, 2022). For instance, Senay et al. (2020) applied a temperature-sensitive model, i.e., the Operational Simplified Surface Energy Balance (SSEBop) model, to estimate ET and used its anomalies to successfully detect the 2011/2012 drought in southcentral United States and the 2005 drought in Australia. However, this study mainly used relatively low ET values to qualitative describe droughts, while few studies focused on the accuracy of ET estimates under similar extreme conditions (i.e., heat and drought conditions). Hence, this section further examines the 3T model-based terrestrial ET product for extreme weather condition monitoring by validating its performance against EC flux tower observations under extreme heat, extreme atmospheric drought, and extreme soil drought conditions.

These three types of extreme hazards were defined according to daily observations from 2001–2020 retrieved from FLUXNET and GLDAS reanalysis data (version 2.1) based on corresponding site locations: 1) extreme heat conditions occur when the daily $T_a$ at a given EC tower site is higher than the 95th percentile of the daily $T_a$ in GLDAS data; 2) extreme atmospheric drought conditions occur when the daily VPD at a given EC tower location is higher than the 95th percentile of the daily VPD in GLDAS data; 3) extreme soil drought conditions occur when the daily soil moisture matches the 5th percentile of EC tower data. It should be mentioned that some data points indicated both extreme heat and extreme atmospheric drought conditions. These points were designated as extreme heat conditions instead of extreme atmospheric drought conditions. Finally, there remained 11213 data points across 80 sites, 19687 data points across 112 sites, and 12338 data points across 95 sites representing extreme heat, extreme atmospheric drought, and extreme soil drought conditions, respectively. GLDAS estimates, with a high temporal resolution in the monitoring of extreme events (Liu et al., 2019) and the same spatiotemporal input datasets as those employed for the 3T model, were also used in the analysis.

Under extreme heat conditions (Fig. 11), although the 3T model-based ET product exhibited various performance levels in the different biomes, the product generally yielded results closely agreeing with observations. In terms of the mean ET value, extreme heat conditions at the DBF, WET, OSH, MF, CRO, and ENF sites were the best captured with the 3T model-based ET product (Fig. 11a), with a maximum difference of 11.9 W m$^{-2}$ from EC observations, followed by the GRA, WSA, SAV, and EBF sites with difference values ranging from 24.0 to 51.1 W m$^{-2}$. The GLDAS performed similarly to the 3T model-based ET product, but with a notably higher bias than that of EC observations. The RMSE violin plots shown in Fig. 11b further verify the above statement because the RMSE values obtained with the 3T model-based ET product, with median values of 23.6, 29.0, 15.3, 31.2, and 24.4 W m$^{-2}$ at the OSH, ENF, WET, CRO, and MF sites, respectively, were much smaller than those obtained with the GLDAS (37.9, 37.4, 19.9, 35.2, and 28.2 W m$^{-2}$, respectively). The maximum RMSE values obtained with the 3T model-based ET product were also smaller than those obtained with the GLDAS, 48.3, 20.6, 20.2, 15.6, and 14.1 W m$^{-2}$ lower at the CRO, DBF, OSH, WET, and MF sites, respectively. These results indicate that the 3T model-based ET product could accurately capture the low ET values under extreme heat conditions in most biomes and performed better than did the GLDAS.

Under extreme atmospheric drought condition (Fig. 12), in terms of the mean ET value, the 3T model-based ET product suitably captured extreme atmospheric drought conditions at the OSH, ENF, MF, DBF, GRA, WET, CRO, and WSA sites (Fig. 12a), with a maximum difference of 14.5 W m$^{-2}$ from EC observations, followed by the SAV and EBF sites with difference values ranging from 18.2 to 29.7 W m$^{-2}$. The GLDAS also performed similarly to the 3T model-based ET product, but with a higher bias over EC observations, which was further confirmed by the RMSE violin plots shown in Fig. 12b. The median RMSE values obtained with the 3T model-based ET product (the white points in Fig. 12b) reached 30.7, 22.2, 26.2, 21.0, and 12.7 W m$^{-2}$ at the CRO, MF, ENF, SAV, and WET sites, respectively, while the values obtained with the GLDAS reached 38.4, 27.4, 29.0, 22.9, and 14.4, respectively. The above results indicate that the 3T model-based ET product could accurately capture the low ET values under extreme atmospheric drought conditions at the CRO, MF, ENF, and WET sites and performed better than did the GLDAS.

Under extreme soil drought condition (Fig. 13), in terms of the mean ET value, the 3T model-based ET product suitably captured extreme soil drought conditions at the GRA, OSH, MF, WSA, WET, CRO, DBF, and ENF sites (Fig. 13a), with a maximum difference of 7.2 W m$^{-2}$ from EC observations, followed by the SAV, and EBF sites with difference values ranging
410 from 13.2 to 25.5 W m$^{-2}$. The median RMSE values obtained with the 3T model-based ET product (the white points in Fig. 13b) at the SAV, EBF, CRO, OSH, and ENF sites reached 18.9, 23.2, 18.2, 5.4, and 15.3 W m$^{-2}$, respectively, while the values obtained with the GLDAS reached 23.8, 26.4, 20.0, 6.9, and 15.8, respectively. In addition, compared to the GLDAS, the maximum RMSE values obtained with the 3T model-based ET product at the ENF, EBF, CRO, WET, and SAV sites was reduced by 55.6, 18.5, 14.0, 9.3, and 3.0 W m$^{-2}$, respectively. The acquired results indicate that the 3T model-based ET product
415 could accurately capture the low ET values under extreme atmospheric drought conditions at the CRO, EBF, and ENF and performed better than did the GLDAS.

It should be noted that the 3T model-based ET product exhibited a good performance in crop ET estimation under these three types of extreme conditions. Compared to the GLDAS, the 3T model-based ET estimates were closer to the considered EC observations and exhibited smaller errors, as described in the previous discussion. Considering that CRO areas are important
420 for human society but highly sensitive to extreme events (Xia et al., 2021) and crop ET estimation suffers more challenges than those encountered in the other natural biomes (He et al., 2019; Melton et al., 2021), the sensitivity of the 3T model to the temperature ensures that the method could provide very high potential ability for crop ET estimation, especially under extreme temperature conditions.

## 5 Data availability

425 The daily and monthly ET dataset presented and analysed in this article has been released and is available for free download from the Science Data Bank (http://doi.org/10.57760/sciencedb.o00014.00001, Xiong et al., 2022). The dataset is published under the Creative Commons Attribution 4.0 International (CC BY 4.0) license.

## 6 Conclusions

A global ET product, derived from reanalysis and RS data based on the 3T model, was provided with daily and 0.25°
430 resolutions from 2001 to 2020. The product was thoroughly assessed via direct evaluation against FLUXNET EC tower data at the daily and monthly scales and water balance-based catchment ET data at the annual scale, in addition to cross validation against six commonly used global ET products. The 3T model-based ET estimates generally agreed well with the above observations. Furthermore, the 3T model exhibited a very high potential for accurate ET estimation under extreme weather conditions. Since the 3T model requires only a few input parameters (i.e., $R_n$, LST, and $T_a$) without the need for parameter
435 calibration, it could be concluded that the model is easy and simple to apply and the proposed ET product could provide reasonable information to support water cycle-related studies.

**Author contributions**

YU LY was responsible for the writing – original draft preparation, data processing and presentation, programming and visualization. QIU GY contributed to the methodology, supervision, project administration, funding acquisition and writing – review & editing. YAN CH, ZHAO WL, ZOU ZD, QIN LJ was responsible for the writing – review & editing. DING JS was responsible for the programming and visualization. XIONG YJ was responsible for the conceptualization, formal analysis, funding acquisition,  writing – review & editing and supervision.

**Competing interests**

The authors declare that they have no conflict of interest.

**Acknowledgements**

The GLDAS data were obtained from the Goddard Earth Sciences Data and Information Services Center (https://disc.gsfc.nasa.gov/, last access: 12 March 2022). MOD13C2 data were obtained from the NASA Land Processes Distributed Active Archive Center (https://lpdaac.usgs.gov/, last access: 17 January 2022). FLUXNET data were obtained from (https://fluxnet.org/, last access: 5 September 2021). The Koppen-Geiger climate classification data were obtained from (http://koeppen-geiger.vu-wien.ac.at/present.htm, last access: 16 April 2021). The Fluxcom data were obtained from (http://www.fluxcom.org/, last access: 3 October 2020). The GLEAM data were obtained from (https://www.gleam.eu/, last access: 17 June 2021). The MOD16 data were obtained from the Numerical Terradynamic Simulation Group (http://files.ntsg.umt.edu/, last access: 25 November 2020). The P-LSH data were obtained from the Numerical Terradynamic Simulation Group (http://files.ntsg.umt.edu/, last access: 25 November 2020). The PML data were obtained from the National Tibetan Plateau Data Center (http://data.tpdc.ac.cn/zh-hans/data/48c16a8d-d307-4973-abab-972e9449627c/, last access: 24 November 2020). The authors would like to thank the above organizations for providing datasets.

**Financial support**

This research was financially supported by the Shenzhen Science and Technology Innovation Program (Grant No. GXWD20201231165807007-20200827105738001), the National Natural Science Foundation of China (Grant No. 42071395 and 42001022) and the Sichuan Province Science and Technology Support Program (Grant No. 2021YFH0082 and 5132202020000046).

## References

Badgley, G., Fisher, J. B., Jiménez, C., Tu, K. P., and Vinukollu, R.: On uncertainty in global terrestrial evapotranspiration estimates from choice of input forcing datasets, J. Hydrometeorol., 16, 1449-1455, https://doi.org/10.1175/JHM-D-14-0040.1, 2015.

Bao, X. and Zhang, F.: Evaluation of NCEP–CFSR, NCEP–NCAR, ERA-Interim, and ERA-40 reanalysis datasets against independent sounding observations over the Tibetan Plateau, J. Clim., 26, 206-214, https://doi.org/10.1175/JCLI-D-12-00056.1, 2013.

Beaudoing, H. and Rodell M.: GLDAS Noah Land Surface Model L4 3 hourly 0.25 x 0.25 degree V2.1, Goddard Earth Sciences Data and Information Services Center (GES DISC) [data set], https://doi.org/10.5067/E7TYRXPJKWOQ, 2020.

Bengtsson, L. and Shukla, J.: Integration of space and in situ observations to study global climate change, Bull. Am. Meteorol. Soc., 69, 1130-1143, https://doi.org/10.1175/1520-0477(1988)069<1130:IOSAIS>2.0.CO;2, 1988.

Bonan, G. B., Patton, E. G., Finnigan, J. J., Baldocchi, D. D., and Harman, I. N.: Moving beyond the incorrect but useful paradigm: reevaluating big-leaf and multilayer plant canopies to model biosphere-atmosphere fluxes–a review, Agric. For. Meteorol., 306, 108435, https://doi.org/10.1017/aju.2017.9, 2021.

Cao, M., Wang, W., Xing, W., Wei, J., Chen, X., Li, J., and Shao, Q.: Multiple sources of uncertainties in satellite retrieval of terrestrial actual evapotranspiration, J. Hydrol., 601, 126642, https://doi.org/10.1016/j.jhydrol.2021.126642, 2021.

Cao, Q., Liu, Y., Sun, X., and Yang, L.: Country-level evaluation of solar radiation data sets using ground measurements in China, Energy, 241, 122938, https://doi.org/10.1016/j.energy.2021.122938, 2022.

Carter, C. and Liang, S.: Comprehensive evaluation of empirical algorithms for estimating land surface evapotranspiration, Agric. For. Meteorol., 256, 334-345, https://doi.org/10.1016/j.agrformet.2018.03.027, 2018.

Chen, M., Senay, G. B., Singh, R. K., and Verdin, J. P.: Uncertainty analysis of the Operational Simplified Surface Energy Balance (SSEBop) model at multiple flux tower sites, J. Hydrol., 536, 384-399, https://doi.org/10.1016/j.jhydrol.2016.02.026, 2016.

Chen, X., Su, Z., Ma, Y., Trigo, I., and Gentine, P.: Remote Sensing of Global Daily Evapotranspiration based on a Surface Energy Balance Method and Reanalysis Data, J. Geophys. Res.: Atmos., 126, e2020JD032873, https://doi.org/10.1029/2020JD032873, 2021.

Cheng, M., Jiao, X., Li, B., Yu, X., Shao, M., and Jin, X.: Long time series of daily evapotranspiration in China based on the SEBAL model and multisource images and validation, Earth Syst. Sci. Data, 13, 3995-4017, https://doi.org/10.5194/essd-13-3995-2021, 2021.

Cleugh, H. A., Leuning, R., Mu, Q., and Running, S. W.: Regional evaporation estimates from flux tower and MODIS satellite data, Remote Sens. Environ., 106, 285-304, https://doi.org/10.1016/j.rse.2006.07.007, 2007.

Didan, K.: MOD13C2 MODIS/Terra Vegetation Indices Monthly L3 Global 0.05Deg CMG V006. NASA EOSDIS Land Processes DAAC [Data set], https://doi.org/10.5067/MODIS/MOD13C2.006, 2015.

495 Elnashar, A., Wang, L., Wu, B., Zhu, W., and Zeng, H.: Synthesis of global actual evapotranspiration from 1982 to 2019, Earth Syst. Sci. Data, 13, 447-480, https://doi.org/10.5194/essd-13-447-2021, 2021.

Ershadi, A., McCabe, M., Evans, J., and Wood, E. F.: Impact of model structure and parameterization on Penman–Monteith type evaporation models, J. Hydrol., 525, 521-535, https://doi.org/10.1016/j.jhydrol.2015.04.008, 2015.

Fisher, J. B., Melton, F., Middleton, E., Hain, C., Anderson, M., Allen, R., McCabe, M. F., Hook, S., Baldocchi, D., and
500 Townsend, P. A.: The future of evapotranspiration: Global requirements for ecosystem functioning, carbon and climate feedbacks, agricultural management, and water resources, Water Resour. Res., 53, 2618-2626, https://doi.org/10.1002/2016WR020175, 2017.

Fu, X. and Wang, B.: Reliability evaluation of soil moisture and land surface temperature simulated by Global Land Data Assimilation System (GLDAS) using AMSR-E data, Remote Sensing and Modeling of the Atmosphere, Oceans, and
505 Interactions V, 92650O, https://doi.org/10.1117/12.2074566, 2014.

Guo, H., Liang, D., Chen, F., and Shirazi, Z.: Innovative approaches to the Sustainable Development Goals using Big Earth Data, Big Earth Data, 5, 263-276, https://doi.org/10.1080/20964471.2021.1939989, 2021.

Guo, X., Liu, H., and Yang, K.: On the application of the Priestley–Taylor relation on sub-daily time scales, Boundary Layer Meteorol., 156, 489-499, https://doi.org/10.1007/s10546-015-0031-y, 2015.

510 Han, C., Ma, Y., Wang, B., Zhong, L., Ma, W., Chen, X., and Su, Z.: Long-term variations in actual evapotranspiration over the Tibetan Plateau, Earth Syst. Sci. Data, 13, 3513-3524, https://doi.org/10.5194/essd-13-3513-2021, 2021.

He, M., Kimball, J. S., Yi, Y., Running, S. W., Guan, K., Moreno, A., Wu, X., and Maneta, M.: Satellite data-driven modeling of field scale evapotranspiration in croplands using the MOD16 algorithm framework, Remote Sens. Environ., 230, 111201, https://doi.org/10.1016/j.rse.2019.05.020, 2019.

515 IPCC: Summary for Policymakers, in: Climate Change 2022: Impacts, Adaptation, and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: H.-O. Pörtner, D.C. Roberts, M. Tignor, E.S. Poloczanska, K. Mintenbeck, A. Alegría, M. Craig, S. Langsdorf, S. Löschke, V. Möller, A. Okem, and B. Rama. Cambridge University Press. In Press, 2022.

Ji, L., Senay, G. B., and Verdin, J. P.: Evaluation of the Global Land Data Assimilation System (GLDAS) air temperature data
520 products, J. Hydrometeorol., 16, 2463-2480, https://doi.org/10.1175/JHM-D-14-0230.1, 2015.

Jiang, C. and Ryu, Y.: Multi-scale evaluation of global gross primary productivity and evapotranspiration products derived from Breathing Earth System Simulator (BESS), Remote Sens. Environ., 186, 528-547, https://doi.org/10.1016/j.rse.2016.08.030, 2016.

Jung, M., Koirala, S., Weber, U., Ichii, K., Gans, F., Camps-Valls, G., Papale, D., Schwalm, C., Tramontana, G., and Reichstein,
525 M.: The FLUXCOM ensemble of global land-atmosphere energy fluxes, Sci. Data, 6, 1-14, https://doi.org/10.1038/s41597-019-0076-8, 2019.

Khan, M. S., Liaqat, U. W., Baik, J., and Choi, M.: Stand-alone uncertainty characterization of GLEAM, GLDAS and MOD16 evapotranspiration products using an extended triple collocation approach, Agric. For. Meteorol., 252, 256-268, https://doi.org/10.1016/j.agrformet.2018.01.022, 2018.

530 Kim, S., Anabalón, A., and Sharma, A.: An assessment of concurrency in evapotranspiration trends across multiple global datasets, J. Hydrometeorol., 22, 231-244, https://doi.org/10.1175/JHM-D-20-0059.1, 2021.

Komatsu, H.: Forest categorization according to dry-canopy evaporation rates in the growing season: comparison of the Priestley–Taylor coefficient values from various observation sites, Hydrol. Processes, 19, 3873-3896, https://doi.org/10.1002/hyp.5987, 2005.

535 Kottek, M., Grieser, J., Beck, C., Rudolf, B., and Rubel, F.: World map of the Köppen-Geiger climate classification updated, Meteorol. Z., 15, 259-263, https://doi.org/10.1127/0941-2948/2006/0130, 2006.

Lhomme, J.-P., Monteny, B., and Amadou, M.: Estimating sensible heat flux from radiometric temperature over sparse millet, Agric. For. Meteorol., 68, 77-91, https://doi.org/10.1016/0168-1923(94)90070-1, 1994.

Liu, W., Wang, L., Zhou, J., Li, Y., Sun, F., Fu, G., Li, X., and Sang, Y.-F.: A worldwide evaluation of basin-scale 540 evapotranspiration estimates against the water balance method, J. Hydrol., 538, 82-95, https://doi.org/10.1016/j.jhydrol.2016.04.006, 2016.

Liu, Y., Liu, Y., and Wang, W.: Inter-comparison of satellite-retrieved and Global Land Data Assimilation System-simulated soil moisture datasets for global drought analysis, Remote Sens. Environ., 220, 1-18, https://doi.org/10.1016/j.rse.2018.10.026, 2019.

545 Liu, Y., Qiu, G. Y., Zhang, H., Yang, Y., Zhang, Y., Wang, Q., Zhao, W., Jia, L., Ji, X., Xiong, Y., Yan, C., Ma, N., Han, S., and Cui, Y.: Shifting from homogeneous to heterogeneous surfaces in estimating terrestrial evapotranspiration: Review and perspectives, Sci. China Earth Sci., 65, 197-214, https://doi.org/10.1007/s11430-020-9834-y, 2022.

Ma, N., Szilagyi, J., and Zhang, Y.: Calibration-free complementary relationship estimates terrestrial evapotranspiration globally, Water Resour. Res., 57, e2021WR029691, https://doi.org/10.1029/2021WR029691, 2021.

550 Martens, B., Miralles, D. G., Lievens, H., Van Der Schalie, R., De Jeu, R. A., Fernández-Prieto, D., Beck, H. E., Dorigo, W. A., and Verhoest, N. E.: GLEAM v3: Satellite-based land evaporation and root-zone soil moisture, Geosci. Model Dev., 10, 1903-1925, https://doi.org/10.5194/gmd-10-1903-2017, 2017.

Medici, C., Wade, A. J., and Frances, F.: Does increased hydrochemical model complexity decrease robustness?, J. Hydrol., 440, 1-13, https://doi.org/10.1016/j.jhydrol.2012.02.047, 2012.

555 Melton, F. S., Huntington, J., Grimm, R., Herring, J., Hall, M., Rollison, D., Erickson, T., Allen, R., Anderson, M., and Fisher, J. B.: OpenET: Filling a critical data gap in water management for the western united states, J. Am. Water Resour. Assoc., https://doi.org/10.1111/1752-1688.12956, 2021.

Miralles, D. G., Holmes, T., De Jeu, R., Gash, J., Meesters, A., and Dolman, A.: Global land-surface evaporation estimated from satellite-based observations, Hydrol. Earth Syst. Sci., 15, 453-469, https://doi.org/10.5194/hess-15-453-2011, 2011.

560    Miralles, D. G., Jiménez, C., Jung, M., Michel, D., Ershadi, A., McCabe, M., Hirschi, M., Martens, B., Dolman, A. J., and Fisher, J. B.: The WACMOS-ET project–Part 2: Evaluation of global terrestrial evaporation data sets, Hydrol. Earth Syst. Sci., 20, 823-842, https://doi.org/10.5194/hess-20-823-2016, 2016.

Mu, Q., Zhao, M., and Running, S. W.: Improvements to a MODIS global terrestrial evapotranspiration algorithm, Remote Sens. Environ., 115, 1781-1800, https://doi.org/10.1016/j.rse.2011.02.019, 2011.

565    Mu, Q., Heinsch, F. A., Zhao, M., and Running, S. W.: Development of a global evapotranspiration algorithm based on MODIS and global meteorology data, Remote Sens. Environ., 111, 519-536, https://doi.org/10.1016/j.rse.2007.04.015, 2007.

Mueller, B., Hirschi, M., Jimenez, C., Ciais, P., Dirmeyer, P. A., Dolman, A. J., Fisher, J. B., Jung, M., Ludwig, F., and Maignan, F.: Benchmark products for land evapotranspiration: LandFlux-EVAL multi-data set synthesis, Hydrol. Earth Syst. Sci., 17, 3707-3720, https://doi.org/10.5194/hess-17-3707-2013, 2013.

570    Peng, X., She, J., Zhang, S., Tan, J., and Li, Y.: Evaluation of multi-reanalysis solar radiation products using global surface observations, Atmosphere, 10, 42, https://doi.org/10.3390/atmos10020042, 2019.

Peng, Z., Tang, R., Jiang, Y., and Liu, M.: Global Daily 500-M Evapotranspiration Estimation Over Vegetated Areas Using Rnadom Forest from MODIS Data, 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, 3733-3736, https://doi.org/10.1109/IGARSS47720.2021.9554153, 2021.

575    Priestley, C. H. B. and Taylor, R. J.: On the assessment of surface heat flux and evaporation using large-scale parameters, Mon. Weather Rev., 100, 81-92, https://doi.org/10.1175/1520-0493(1972)100<0081:OTAOSH>2.3.CO;2 1972.

Qiu, G. Y.: A new method for estimation of evapotranspiration, Ph.D. thesis, The united graduate school of agricultural sciences, Tottori University, Japan, 1996.

Qiu, G. Y., Ben-Asher, J., Yano, T., and Momii, K.: Estimation of soil evaporation using the differential temperature method, 580    Soil Sci. Soc. Am. J., 63, 1608-1614, https://doi.org/10.2136/sssaj1999.6361608x, 1999.

Qiu, G. Y., Yu, X., Wen, H., and Yan, C.: An advanced approach for measuring the transpiration rate of individual urban trees by the 3D three-temperature model and thermal infrared remote sensing, J. Hydrol., 587, 125034, https://doi.org/10.1016/j.jhydrol.2020.125034, 2020.

Raupach, M. R. and Finnigan, J.: 'Single-layer models of evaporation from plant canopies are incorrect but useful, whereas 585    multilayer models are correct but useless': discuss, Aust. J. of Plant Physiol., 15, 705-716, https://doi.org/10.1071/PP9880705, 1988.

Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M. G., Schubert, S. D., Takacs, L., and Kim, G.-K.: MERRA: NASA's modern-era retrospective analysis for research and applications, J. Clim., 24, 3624-3648, https://doi.org/10.1175/JCLI-D-11-00015.1, 2011.

590    Rodell, M., Houser, P., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C.-J., Arsenault, K., Cosgrove, B., Radakovich, J., and Bosilovich, M.: The global land data assimilation system, Bull. Am. Meteorol. Soc., 85, 381-394, https://doi.org/10.1175/BAMS-85-3-381, 2004.

Schneider, U., Becker, A., Finger, P., Rustemeier, E., and Ziese, M.: GPCC Full Data Monthly Product Version 2020 at 0.25°: Monthly Land-Surface Precipitation from Rain-Gauges Built on GTS-Based and Historical Data, Global Precipitation Climatology Centre (GPCC) at Deutscher Wetterdienst [data set], https://doi.org/10.5676/DWD_GPCC/FD_M_V2020_025, 2020.

Senay, G. B., Kagone, S., and Velpuri, N. M.: Operational global actual evapotranspiration: development, evaluation, and dissemination, Sensors, 20, 1915, https://doi.org/10.3390/s20071915, 2020.

Su, Z.: The surface energy balance system (SEBS) for estimation of the turbulent heat fluxes, Hydrol. Earth Syst. Sci., 6 (1), 85–99, https://doi.org/ 10.5194/hess-6-85-2002, 2002.

Tian, F., Qiu, G. Y., Yang, Y., Lü, Y., and Xiong, Y.: Estimation of evapotranspiration and its partition based on an extended three-temperature model and MODIS products, J. Hydrol., 498, 210-220, https://doi.org/10.1016/j.jhydrol.2013.06.038, 2013.

Trenberth, K. E., Fasullo, J. T., and Kiehl, J.: Earth's global energy budget, Bull. Am. Meteorol. Soc., 90, 311-324, https://doi.org/10.1175/2008BAMS2634.1, 2009.

Trenberth, K. E., Smith, L., Qian, T., Dai, A., and Fasullo, J.: Estimates of the global water budget and its annual cycle using observational and model data, J. Hydrometeorol., 8, 758-769, https://doi.org/10.1175/JHM600.1, 2007.

Vinukollu, R. K., Wood, E. F., Ferguson, C. R., and Fisher, J. B.: Global estimates of evapotranspiration for climate studies using multi-sensor remote sensing data: Evaluation of three process-based approaches, Remote Sens. Environ., 115, 801-823, https://doi.org/10.1016/j.rse.2010.11.006, 2011.

Wang, Y., Xiong, Y., Qiu, G. Y., and Zhang, Q.: Is scale really a challenge in evapotranspiration estimation? A multi-scale study in the Heihe oasis using thermal remote sensing and the three-temperature model, Agric. For. Meteorol., 230, 128-141, https://doi.org/10.1016/j.agrformet.2016.03.012, 2016.

Watkins, M. M., Wiese, D. N., Yuan, D. N., Boening, C., and Landerer, F. W.: Improved methods for observing Earth's time variable mass distribution with GRACE using spherical cap mascons, J. Geophys. Res.: Solid Earth, 120, 2648-2671, https://doi.org/10.1002/2014JB011547, 2015.

Wu, G., Cai, X., Keenan, T. F., Li, S., Luo, X., Fisher, J. B., Cao, R., Li, F., Purdy, A. J., and Zhao, W.: Evaluating three evapotranspiration estimates from model of different complexity over China using the ILAMB benchmarking system, J. Hydrol., 590, 125553, https://doi.org/10.1016/j.jhydrol.2020.125553, 2020.

Xia, Y., Fu, C., Wu, H., Wu, H., Zhang, H., Cao, Y., and Zhu, Z.: Influences of Extreme Events on Water and Carbon Cycles of Cropland Ecosystems: A Comprehensive Exploration Combining Site and Global Modeling, Water Resour. Res., 57, e2021WR029884, https://doi.org/10.1029/2021WR029884, 2021.

Xiong, Y. and Qiu, G. Y.: Estimation of evapotranspiration using remotely sensed land surface temperature and the revised three-temperature model, Int. J. Remote Sens., 32, 5853-5874, https://doi.org/10.1080/01431161.2010.507791, 2011.

Xiong, Y. and Qiu, G. Y.: Simplifying the revised three-temperature model for remotely estimating regional evapotranspiration and its application to a semi-arid steppe, Int. J. Remote Sens., 35, 2003-2027, https://doi.org/10.1080/01431161.2014.885149, 2014.

Xiong, Y., Yu L., Qiu G. Y., Yan C., Zhao W., Zou Z., Ding J., and Qin L. (2022).: A global terrestrial evapotranspiration product based on three-temperature model from 2001 to 2020 (Version V2), Science Data Bank [Data set], https://doi.org/10.57760/sciencedb.o00014.00001, 2022.

630 Xiong, Y., Zhao, S., Tian, F., and Qiu, G. Y.: An evapotranspiration product for arid regions based on the three-temperature model and thermal remote sensing, J. Hydrol., 530, 392-404, https://doi.org/10.1016/j.jhydrol.2015.09.050, 2015.

Xiong, Y., Zhao, W., Wang, P., Paw U, K. T., and Qiu, G. Y.: Simple and applicable method for estimating evapotranspiration and its components in arid regions, J. Geophys. Res.: Atmos., 124, 9963-9982, https://doi.org/10.1029/2019JD030774, 2019.

Xu, W., Sun, C., Zuo, J., Ma, Z., Li, W., and Yang, S.: Homogenization of monthly ground surface temperature in China

635 during 1961–2016 and performances of GLDAS reanalysis products, J. Clim., 32, 1121-1135, https://doi.org/10.1175/JCLI-D-18-0275.1, 2019.

Yang, Y., Liu, Y., Li, M., Hu, Z., and Ding, Z.: Assessment of reanalysis flux products based on eddy covariance observations over the Tibetan Plateau, Theor. Appl. Climatol., 138, 275-292, https://doi.org/10.1007/s00704-019-02811-1, 2019.

Yao, Y., Liang, S., Cheng, J., Liu, S., Fisher, J. B., Zhang, X., Jia, K., Zhao, X., Qin, Q., Zhao, B., Han, S., Zhou, G., Zhou,

640 G., Li, Y., and Zhao, S.: MODIS-driven estimation of terrestrial latent heat flux in China based on a modified Priestley–Taylor algorithm, Agric. For. Meteorol., 171-172, 187-202, https://doi.org/https://doi.org/10.1016/j.agrformet.2012.11.016, 2013.

Yao, Y., Liang, S., Li, X., Hong, Y., Fisher, J. B., Zhang, N., Chen, J., Cheng, J., Zhao, S., and Zhang, X.: Bayesian multimodel estimation of global terrestrial latent heat flux from eddy covariance, meteorological, and satellite observations, J. Geophys. Res.: Atmos., 119, 4521-4545, https://doi.org/10.1002/2013JD020864, 2014.

645 Yao, Y., Liang, S., Li, X., Chen, J., Wang, K., Jia, K., Cheng, J., Jiang, B., Fisher, J. B., Mu, Q., Grünwald, T., Bernhofer, C., and Roupsard, O.: A satellite-based hybrid algorithm to determine the Priestley–Taylor parameter for global terrestrial latent heat flux estimation across multiple biomes, Remote Sens. Environ., 165, 216-233, https://doi.org/https://doi.org/10.1016/j.rse.2015.05.013, 2015.

Yao, Y., Liang, S., Li, X., Zhang, Y., Chen, J., Jia, K., Zhang, X., Fisher, J. B., Wang, X., and Zhang, L.: Estimation of high-

650 resolution terrestrial evapotranspiration from Landsat data using a simple Taylor skill fusion method, J. Hydrol., 553, 508-526, https://doi.org/10.1016/j.jhydrol.2017.08.013, 2017.

Zeng, Z., Wang, T., Zhou, F., Ciais, P., Mao, J., Shi, X., and Piao, S.: A worldwide analysis of spatiotemporal changes in water balance-based evapotranspiration from 1982 to 2009, J. Geophys. Res.: Atmos., 119, 1186-1202, https://doi.org/10.1002/2013JD020941, 2014.

655 Zhang, K., Kimball, J. S., Nemani, R. R., Running, S. W., Hong, Y., Gourley, J. J., and Yu, Z.: Vegetation greening and climate change promote multidecadal rises of global land evapotranspiration, Sci. Rep., 5, 1-9, https://doi.org/10.1038/srep15956, 2015.

Zhang, K., Kimball, J. S., and 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.

660 Zhang, X., Liang, S., Wang, G., Yao, Y., Jiang, B., and Cheng, J.: Evaluation of the reanalysis surface incident shortwave radiation products from NCEP, ECMWF, GSFC, and JMA using satellite and surface observations, Remote Sens., 8, 225, https://doi.org/10.3390/rs8030225, 2016.

Zhang, Y., Kong, D., Gan, R., Chiew, F. H., McVicar, T. R., Zhang, Q., and Yang, Y.: Coupled estimation of 500 m and 8-day resolution global evapotranspiration and gross primary production in 2002–2017, Remote Sens. Environ., 222, 165-182,

665 https://doi.org/10.1016/j.rse.2018.12.031, 2019.

Zhao, W., Qiu, G. Y., Xiong, Y., Paw U, K. T., Gentine, P., and Chen, B.: Uncertainties caused by resistances in evapotranspiration estimation using high-density eddy covariance measurements, J. Hydrometeorol., 21, 1349-1365, https://doi.org/10.1175/JHM-D-19-0191.1, 2020.

Zhou, C., Wang, K., and Ma, Q.: Evaluation of eight current reanalyses in simulating land surface temperature from 1979 to

670 2003 in China, J. Clim., 30, 7379-7398, https://doi.org/10.1175/JCLI-D-16-0903.1, 2017.
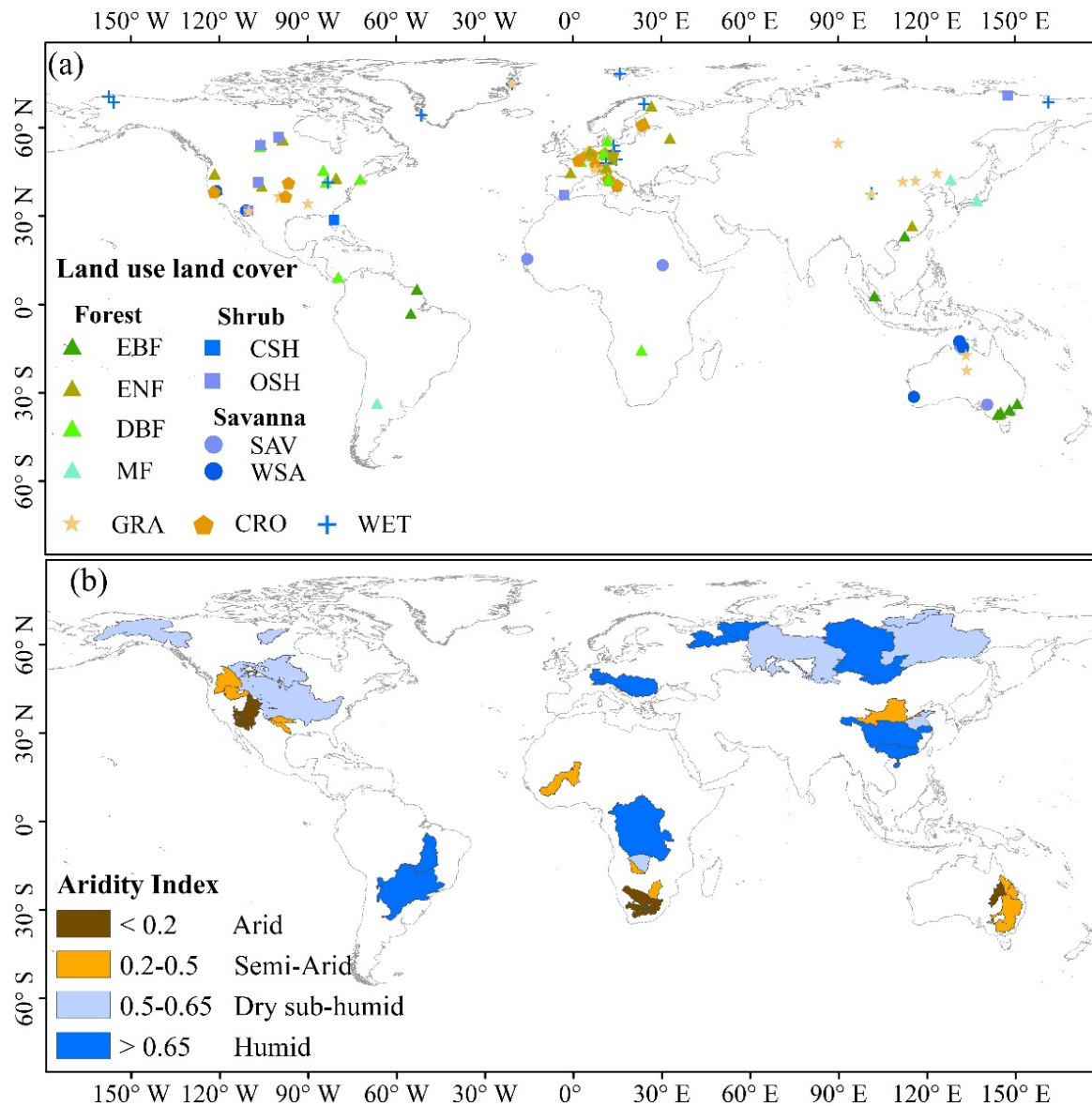
**Figure 1: Locations of the eddy covariance flux towers (a) and catchments (b) used for ET validation in this study. In this study, 126 flux towers and 34 catchments are considered. CRO denotes croplands, CSH denotes closed shrublands, DBF denotes deciduous broadleaf forests, EBF denotes evergreen broadleaf forests, ENF denotes evergreen needleleaf forests, GRA denotes grasslands, MF denotes mixed forests, OSH denotes open shrublands, SAV denotes savannas, WET denotes wetlands, and WSA denotes woody savannas; the multi-year mean aridity index in each catchment is calculated as the mean annual precipitation divided by the mean annual reference ET (Trabucco and Zomer, 2018), and the catchment classification refers to the United Nations Environment Programme (UNEP, 1997).**

680
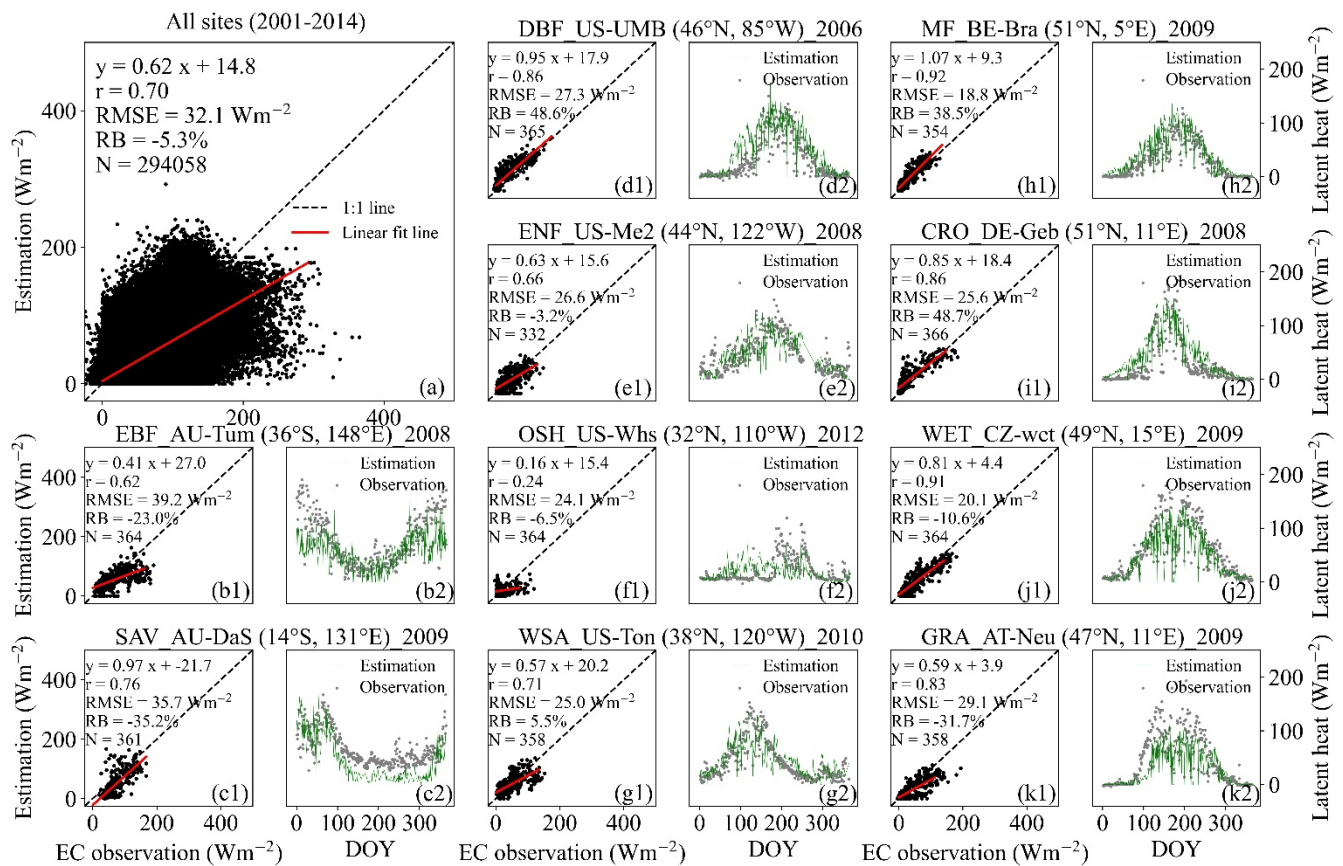
**Figure 2**: **The temporal variations in daily ET estimated from the 3T model (green line) using EC observations (gray dot). The scatter plot between EC observations and ET estimates for one selected year (with RMSE values at average level) at ten EC sites covering various biomes: (a) all 126 sites, (b1-2) EBF_AU-Tum (36°S, 148°E) in 2008, (c1-2) SAV_AU-DaS (14°S, 131°E) in 2009, (d1-2) DBF_US-UMB (46°N, 85°W) in 2006, (e1-2) ENF_US-Me2 (44°N, 122°W) in 2008, (f1-2) OSH_US-Whs (32°N, 110°W) in 2012, (g1-2) WSA_US-Ton (38°N, 120°W) in 2010, (h1-2) MF_BE-Bra (51°N, 5°E) in 2009, (i1-2) CRO_DE-Geb (51°N, 11°E) in 2008, (j1-2) WET_CZ-wet (49°N, 15°E) in 2009, (k1-2) GRA_AT-Neu (47°N, 11°E) in 2009.**

690
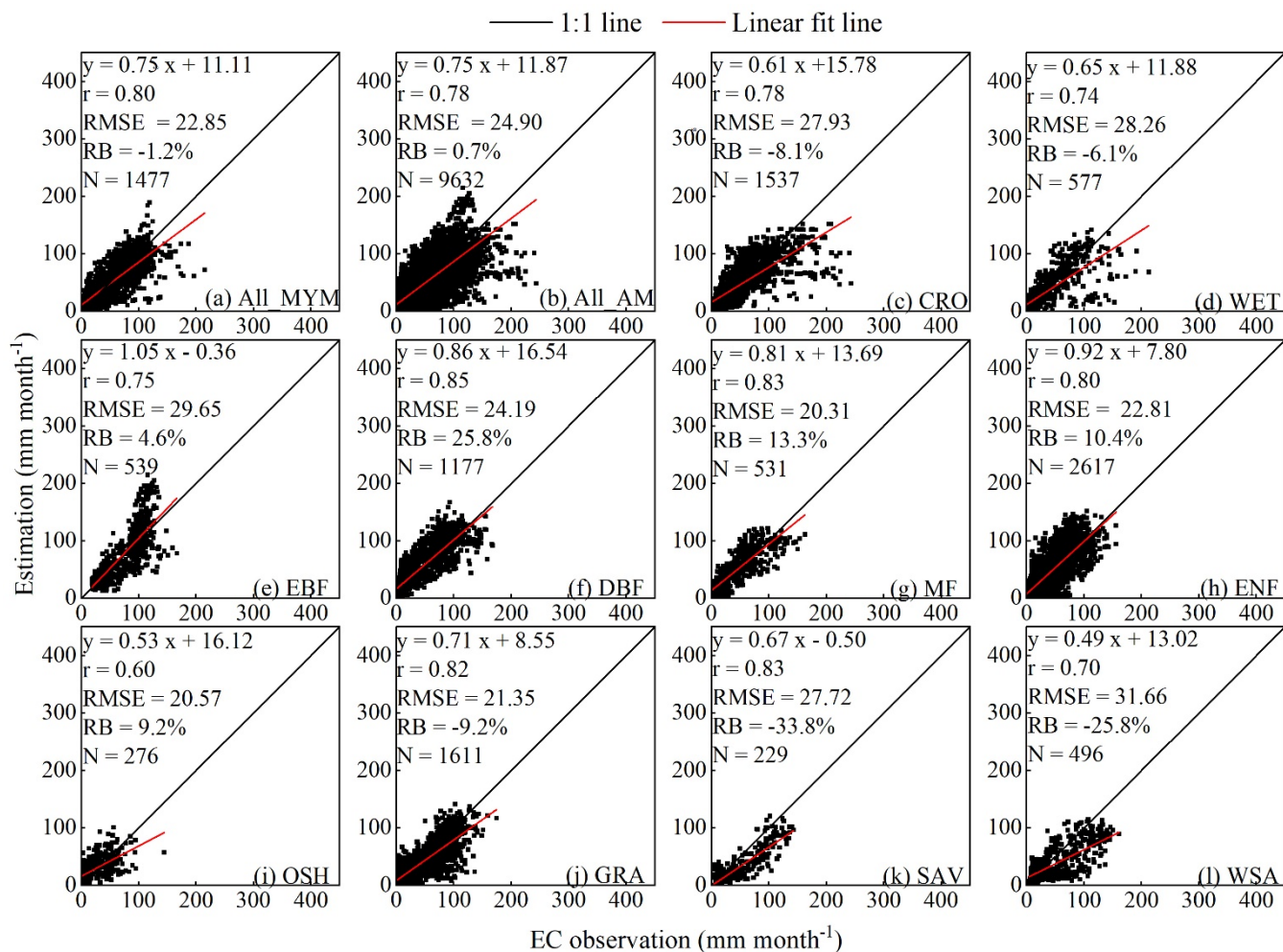
**Figure 3: Comparison of the estimated (3T model) and measured (EC tower) monthly ET values from 2003–2013, where (a) shows the data for all 126 sites on a multi-year monthly mean (MYM) scale and (b) shows the data for all sites on an annual mean (AM) monthly scale. (c)-(l) show all land use/land cover types on an annual monthly scale. The abbreviations in (c)-(l) are the same as those in Fig. 1.**

695
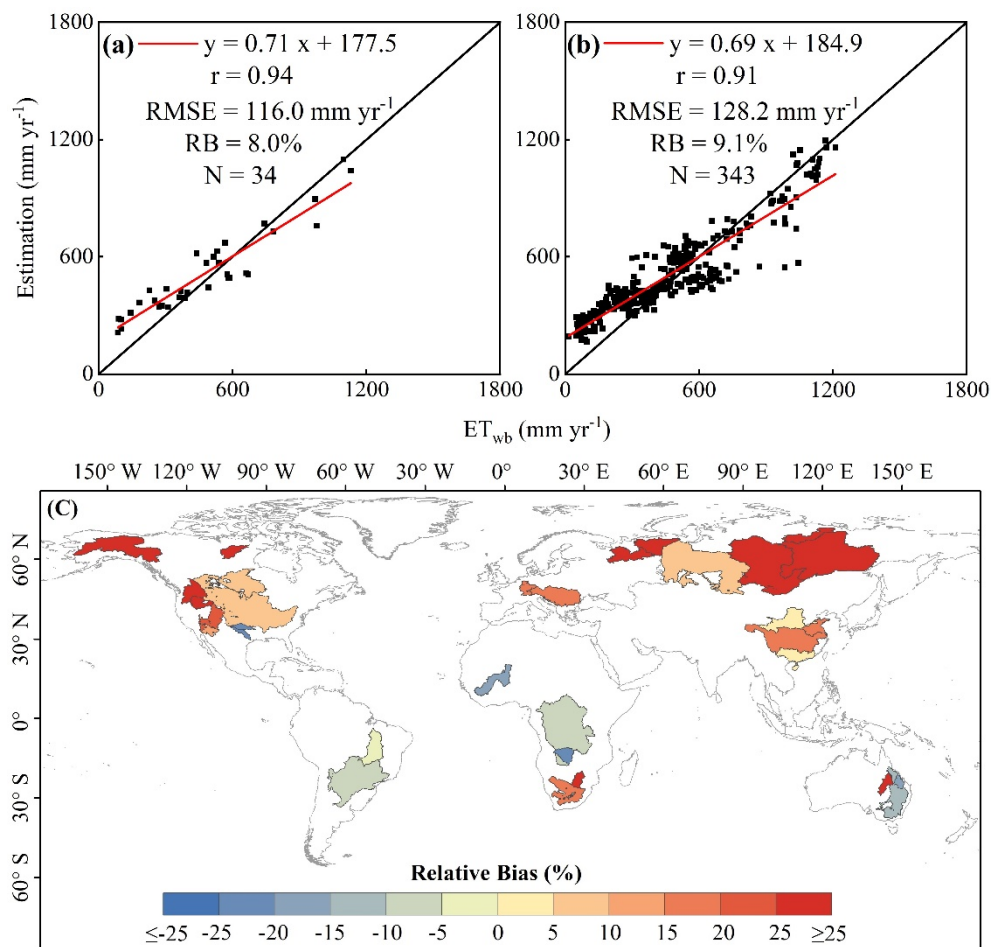
**Figure 4: Comparison of the annual model-estimated (3T model) and water-balance-based ET values during the 2003–2013 period: (a) multi-year mean annual scale, (b) annual scale and (c) relative bias (RB) in each basin.**

700
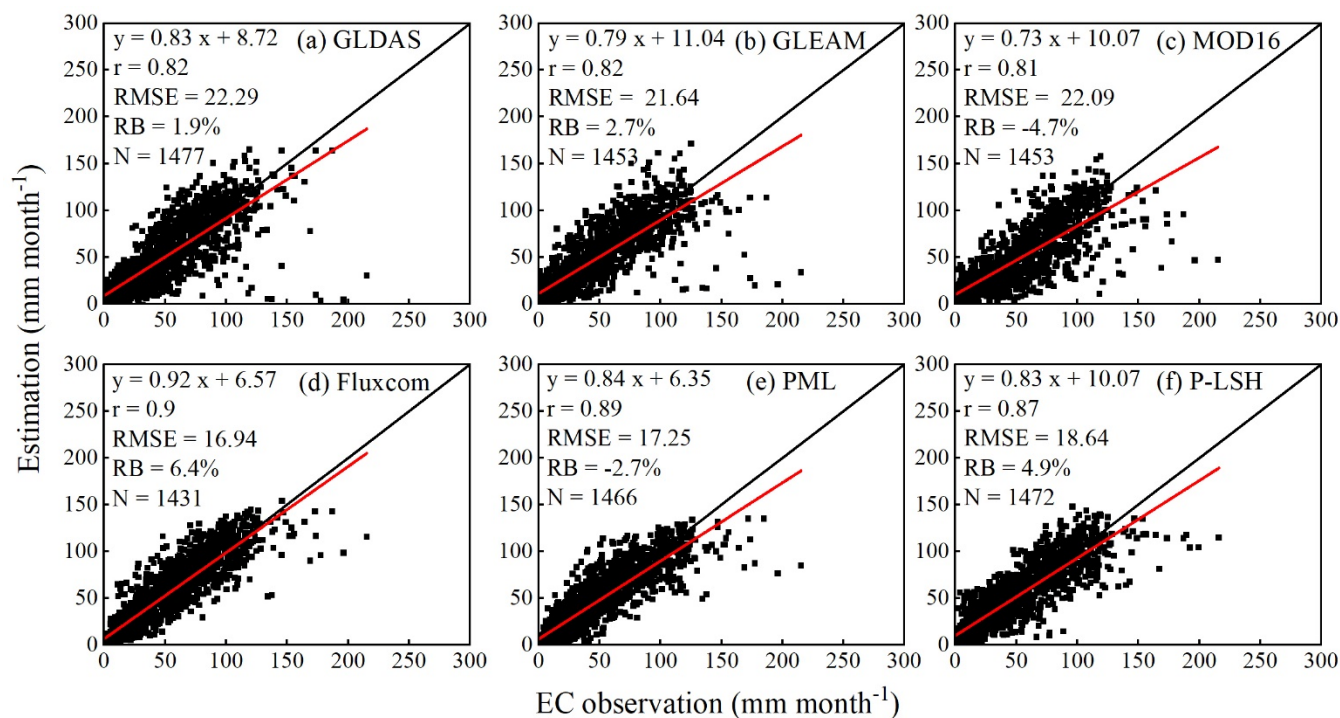
**Figure 5: Validation of 6 commonly used ET products (GLDAS, PML, P-LSH, GLEAM, Fluxcom, and MOD16) against EC tower observations. The data are monthly average ET values over the 2003–2013 period and are the same as those used in Fig. 3a.**

705
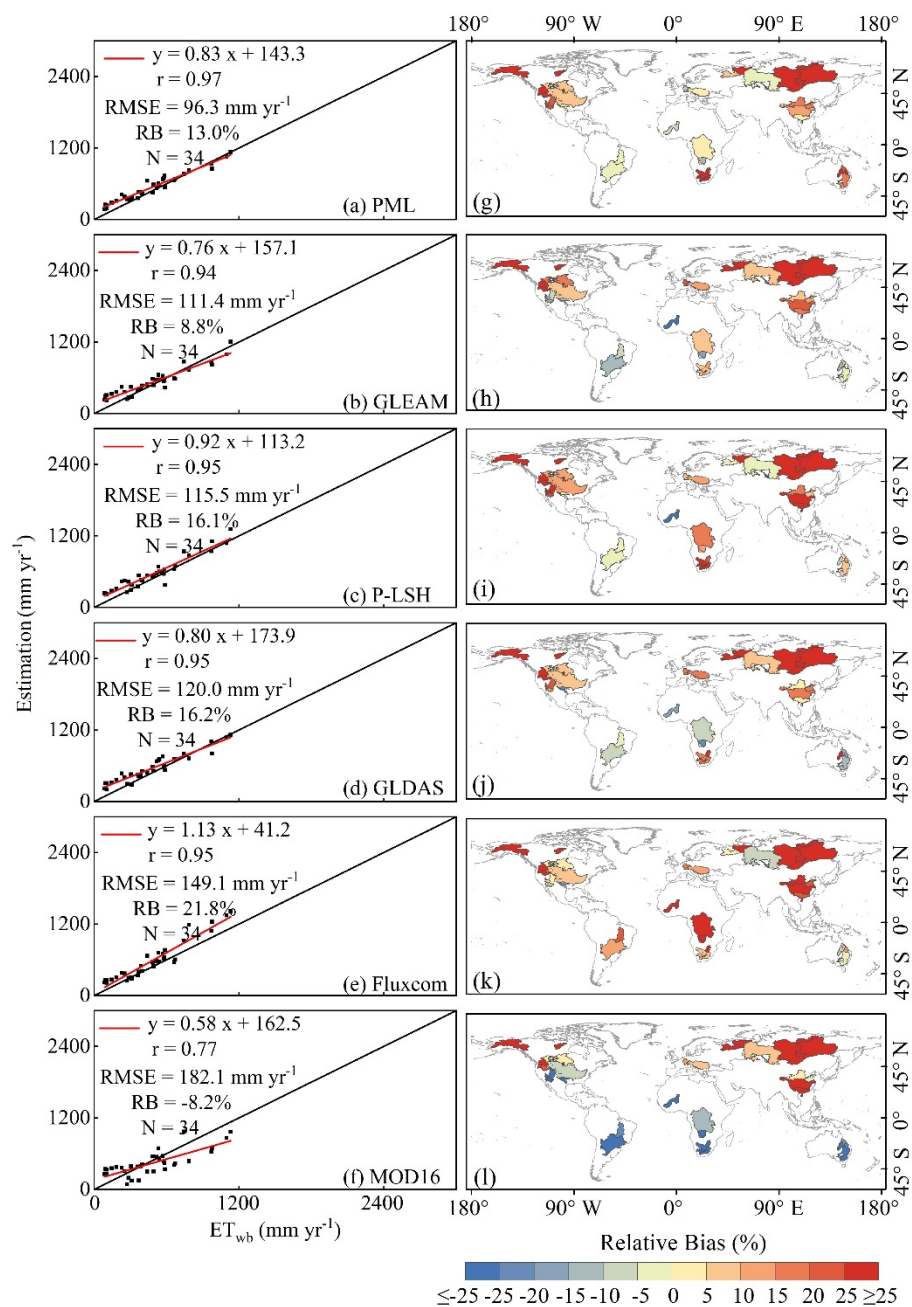
**Figure 6: Validation of 6 commonly used ET products (GLDAS, PML, P-LSH, GLEAM, Fluxcom, and MOD16) against values obtained with the catchment water balance approach. The data are yearly average values over the 2003–2013 period. The left panel shows mean annual values, and the right panel shows the relative bias (RB) in each catchment.**

710
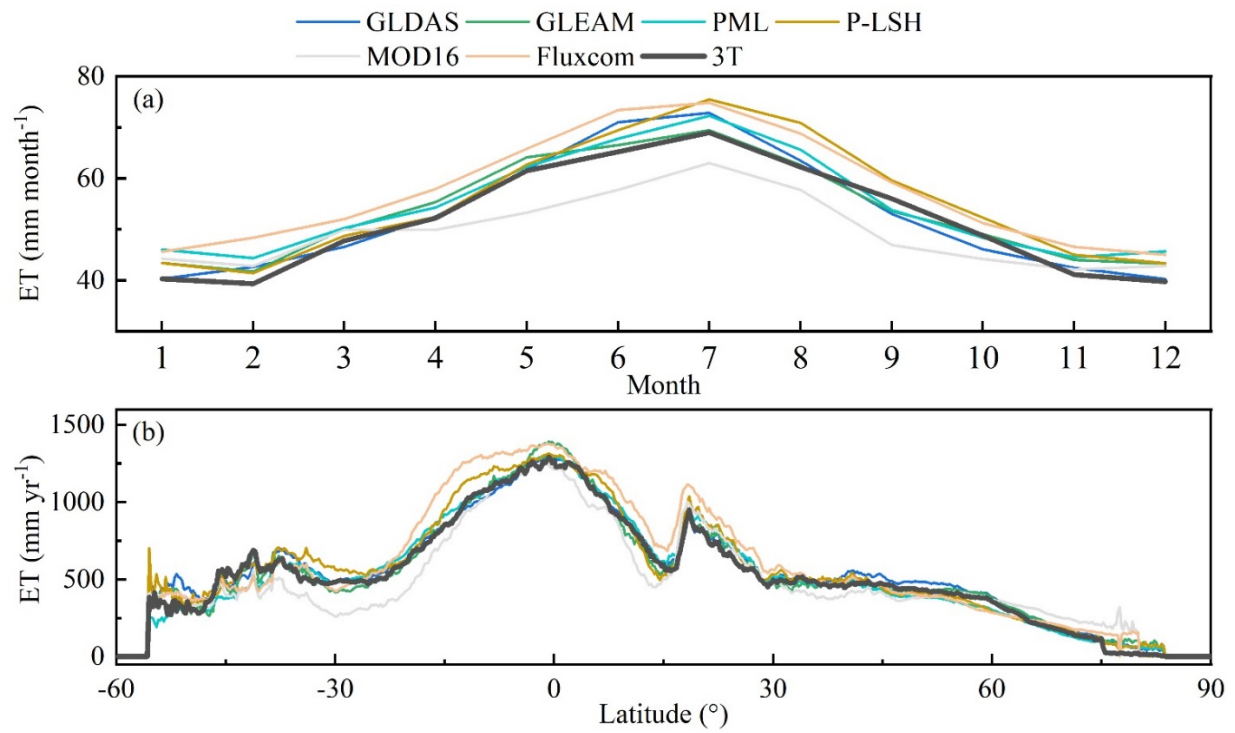
**Figure 7: (a) Monthly variation and (b) annual latitudinal distributions of the multi-year (2003–2013) mean ET value estimated with the 3T model (black line) and 6 ET products in vegetated areas (mainly excluding Greenland, Antarctica and desert areas, according to Jung et al. (2019)).**

715
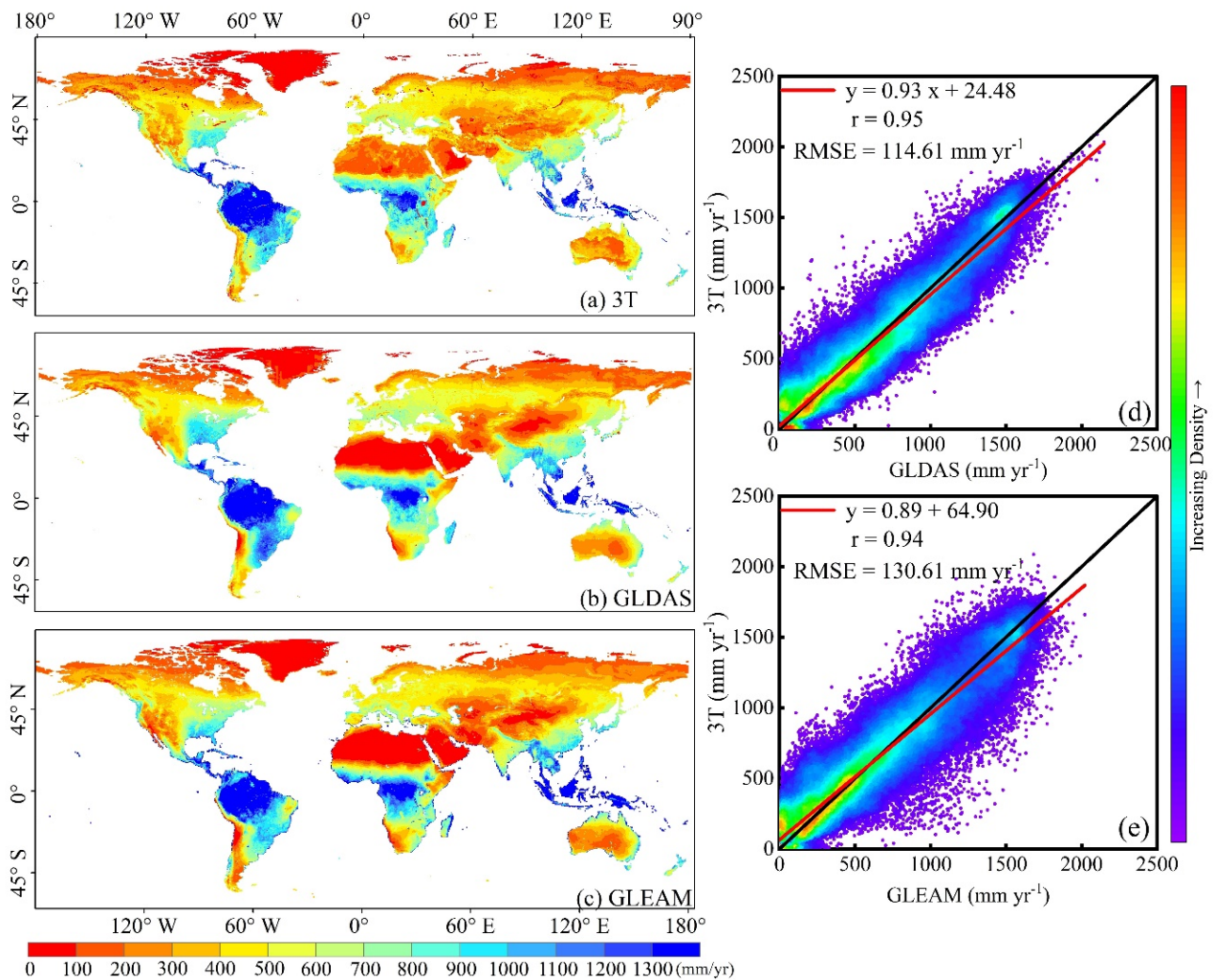
**Figure 8: Spatial pattern and pixel-to-pixel comparison of multi-year (2003–2013) global mean annual ET rates among the 3T model, GLDAS and GLEAM. Left panel: spatial ET distribution of (a) 3T model-, (b) GLDAS- and (c) GLEAM-based ET values. Right panel: pixel-to-pixel comparison of ET values between (d) the 3T model and GLDAS and (e) 3T model and GLEAM.**

720
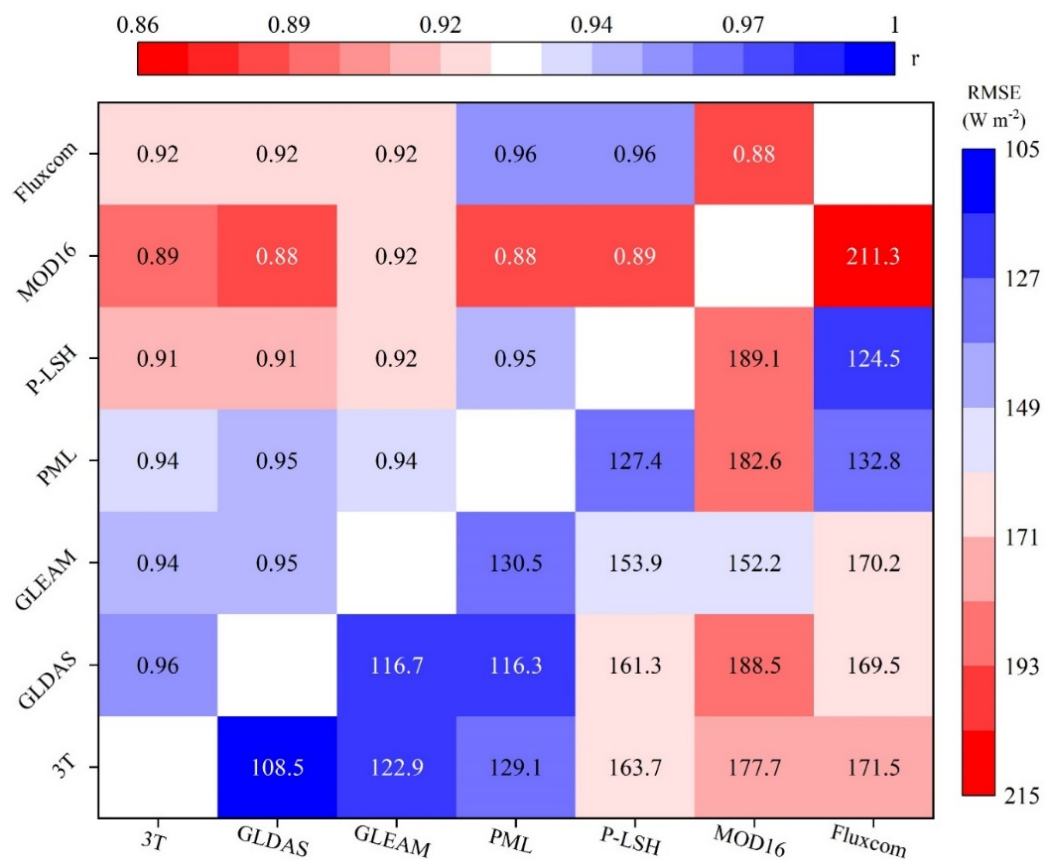
**Figure 9: RMSE and r of pixel-to-pixel (0.25° resolution) comparison of multi-year (2003–2013) mean annual ET values among the 3T model and 6 products in vegetated areas (mainly excluding Greenland, Antarctica and desert areas, according to Jung et al. (2019)).**

725
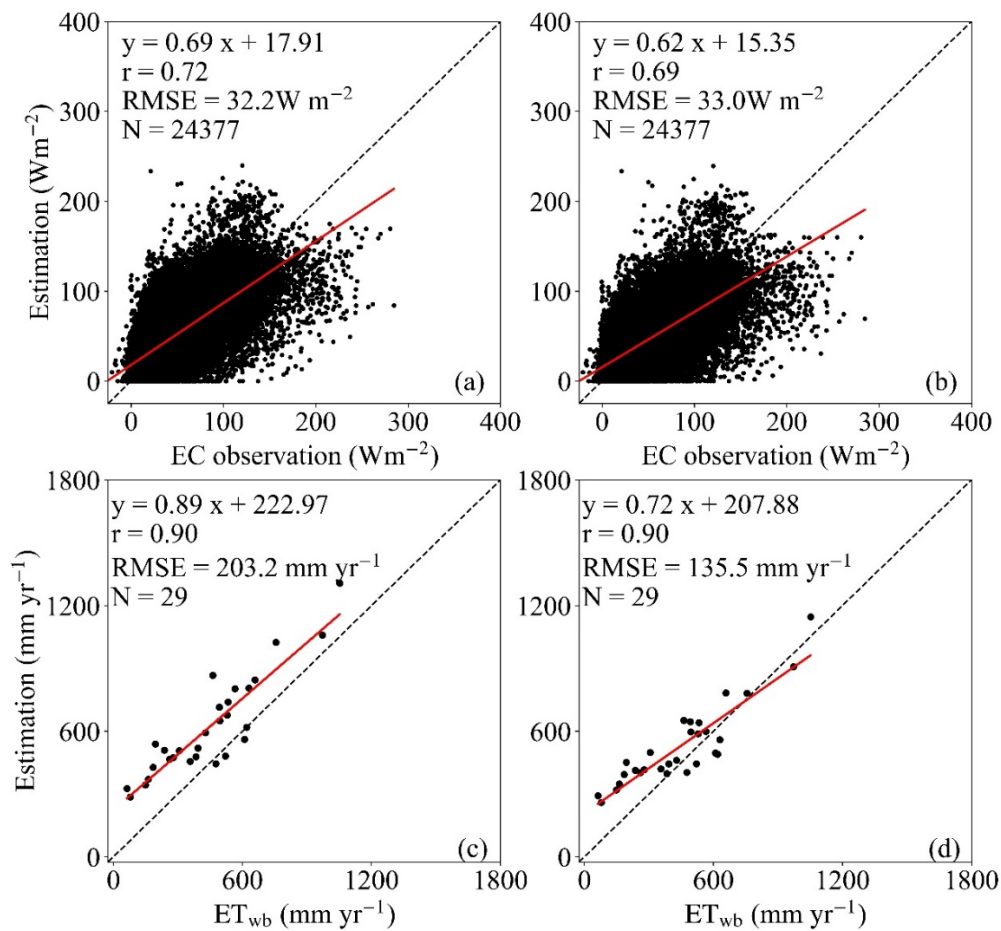
**Figure 10: Comparison of the estimated (3T model) and measured ET values in 2011 (daily ET from EC tower and annual ET from water balance equation). The left panel shows ET estimates using Köppen-Geiger climate regimes with 31 subregions at daily (a) and annual (c) scales, respectively, whereas the right panel is the same but with ET estimates using 90-110 subregions.**
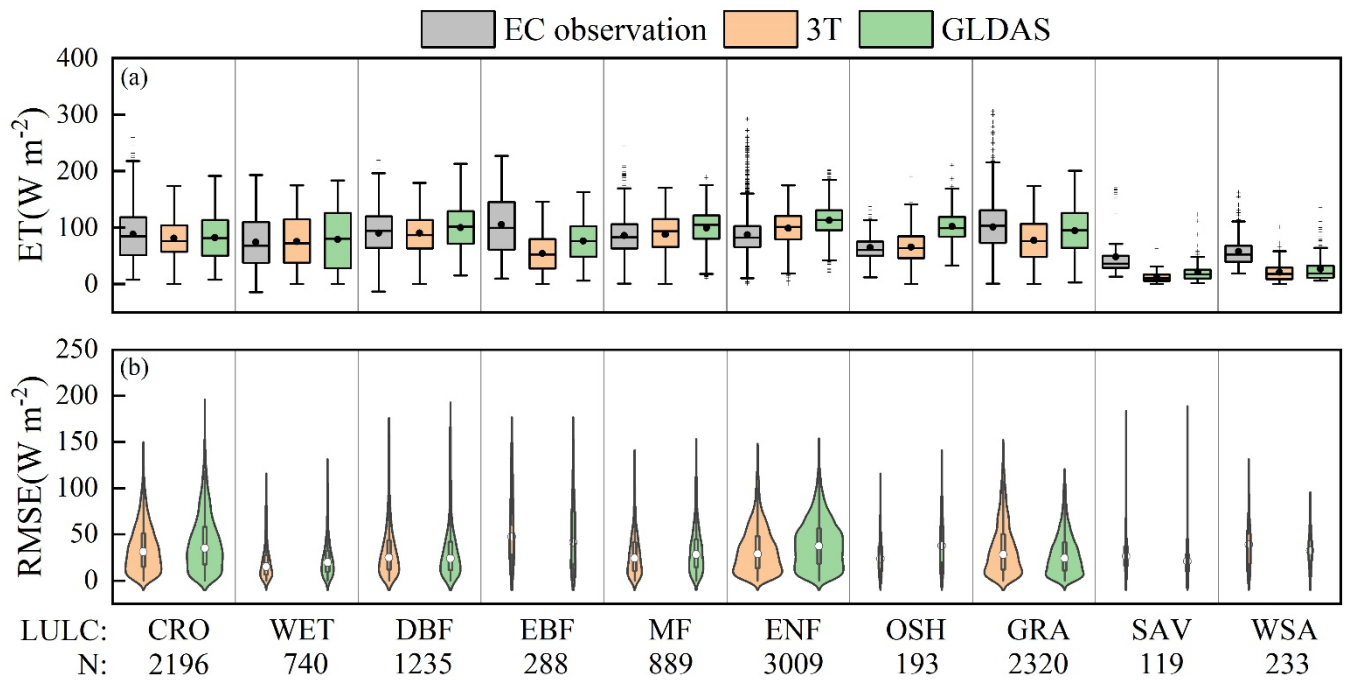
**Figure 11: Monitoring performance of the 3T model-based terrestrial ET product under extreme heat conditions in the**
735 **different biomes. The daily ET is shown in energy units. In the box plot (a), the black point indicates the mean, while central line in the box indicates the median value. The edges of the box indicate the 25th and 75th percentiles, and the whiskers indicate the outlier values. In the violin plot (b), the white point indicates the median value, and a wider violin plot indicates denser data for the same RMSE value. N denotes the number of data points.**
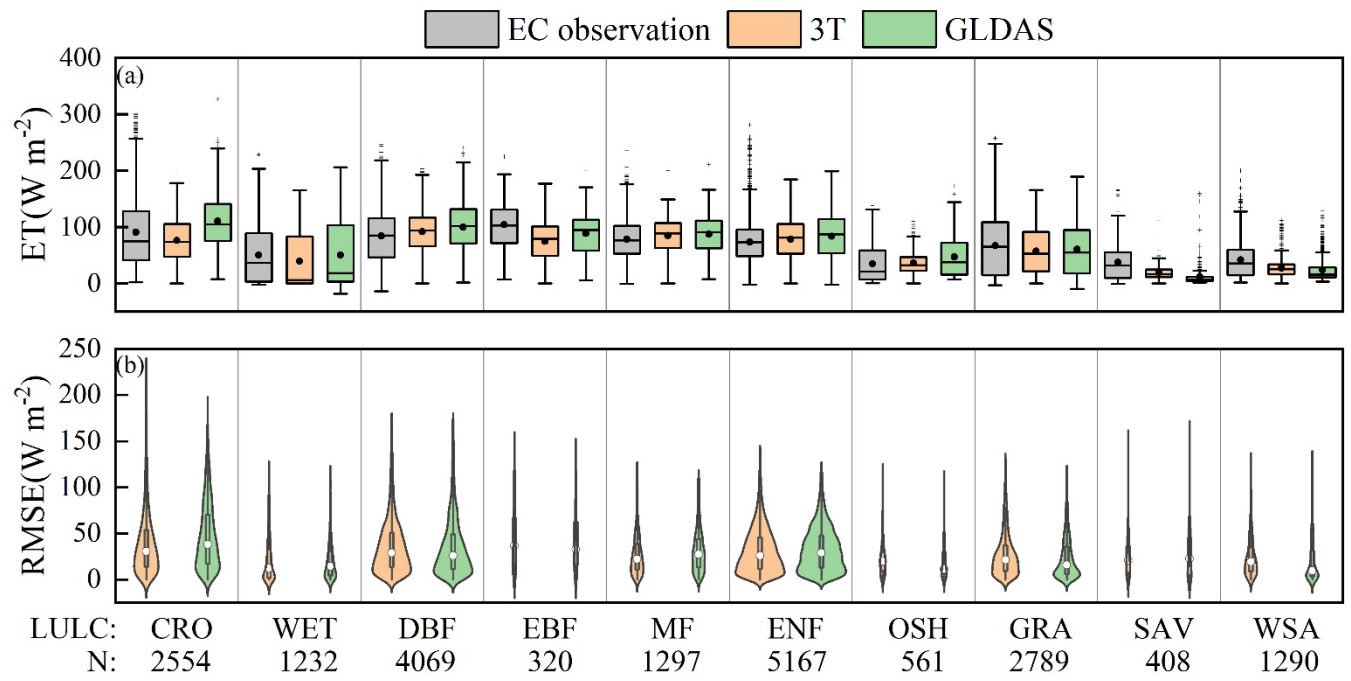
**Figure 12: Monitoring performance of the 3T model-based terrestrial ET product under extreme atmospheric drought conditions in the different biomes. The symbols are the same as those in Figure 11.**

| LULC: | CRO | WET | DBF | EBF | MF | ENF | OSH | GRA | SAV | WSA |
|---|---|---|---|---|---|---|---|---|---|---|
| N: | 1498 | 106 | 2046 | 650 | 567 | 3539 | 364 | 2284 | 453 | 823 |
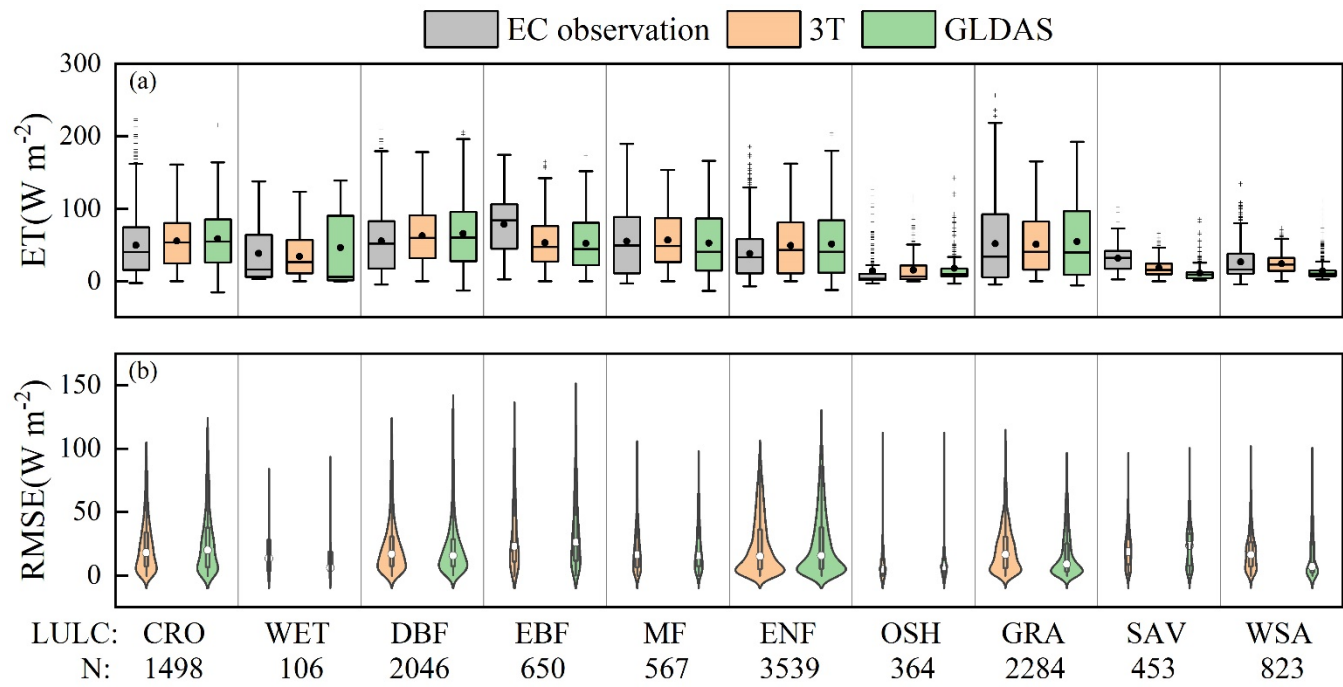
**Figure 13: Monitoring performance of the 3T model-based terrestrial ET product under extreme soil drought conditions in the different biomes. The symbols are the same as those in Figure 11.**

**Table 1: Input datasets for the three-temperature (3T) model-based global ET product.**

| Model input | Datasets | Spatial -temporal resolution | Available data coverage | Reference |
|---|---|---|---|---|
| $R_{ns}$, $R_{nl}$ LST $T_a$ | GLDAS_NOAH 025_3H_2.1 | 0.25° 3-hourly | 2000-2020 | Beaudoing and Rodell, 2020; Rodell et al., 2004 |
| NDVI | MOD13C2 | 0.05° monthly | 2001-2020 | Didan et al., 2015 |

Note: $R_{ns}$: net shortwave radiation; $R_{nl}$: net longwave radiation; LST: land surface temperature; $T_a$: air temperature; NDVI: normalized difference vegetation index.

**Table 2: Information on the typical ET products used to cross validate the ET estimates of the three-temperature (3T) model in this study.**

| ET products | Method | Spatial-temporal resolution | Reference |
|---|---|---|---|
| Fluxcom | Machine learning | 0.083° monthly | Jung et al., 2019 |
| GLDAS | Land surface models | 0.25° 3-hourly & monthly | Beaudoing and Rodell, 2020; Rodell et al., 2004 |
| GLEAM | Priestley-Taylor equation | 0.25° monthly | Martens et al., 2017 Miralles et al., 2011 |
| MOD16 | Penman-Monteith equation | 0.05° monthly | Mu et al., 2011 |
| P-LSH | with different resistance | 0.05° monthly | Zhang et al., 2015 |
| PML | parameterization methods | 0.083° 8-day | Zhang et al., 2019 |

**Table 3: Multi-year (2003–2013) average ET values considering the water depth (mm yr$^{-1}$) and volume (km$^3$ yr$^{-1}$) of the different products used in this study for the global land surface.**

| ET products | ET rate (mm yr$^{-1}$) | ET volume ($\times 10^3$ km$^3$ yr$^{-1}$) |
|---|---|---|
| 3T | $546 \pm 22$ | $73.8 \pm 3.0$ |
| Fluxcom | $549 \pm 3$ | $74.2 \pm 0.4$ |
| GLDAS | $551 \pm 10$ | $74.5 \pm 1.3$ |
| GLEAM | $544 \pm 6$ | $73.6 \pm 0.7$ |
| MOD16 | $468 \pm 6$ | $63.3 \pm 0.8$ |
| P-LSH | $551 \pm 8$ | $74.5 \pm 1.0$ |
| PML | $542 \pm 12$ | $73.2 \pm 1.7$ |

Note: global land surface has an area of $1.35 \times 10^8$ km$^2$, excluding Antarctica. Fluxcom and MOD16 do not provide ET values in Greenland and desert areas.