# **Evaluation of evapotranspiration models using different LAI and meteorological forcing data from 1982 to 2017**

Huiling Chen[a], Gaofeng Zhu[a,*], Kun Zhang[b], Jian Bi[a], Xiaopeng Jia[c], Bingyue Ding[d], Yang Zhang[a], Shasha Shang[a], Nan Zhao[a], Wenhua Qin[a]

[a] Key Laboratory of Western China's Environmental Systems (Ministry of Education), Lanzhou University, Lanzhou, China

[b] Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Beijing, China

[c] Northwest Institute of Eco-Environment and Resources, CAS, Lanzhou, China.

[d] College of Arts and Science Department of Mathematics, Miami University, Oxford, Ohio, USA.

Correspondence to: Gaofeng Zhu (zhugf@lzu.edu.cn).

**Abstract:**

We evaluated the performance of three global evapotranspiration (ET) models using the multiple sets of LAI and meteorological data from 1982 to 2017, and investigated the uncertainty in ET simulations from the model structure and forcing data. The three ET models were the Simple Terrestrial Hydrosphere model (SiTH), Priestly-Taylor Jet Propulsion Laboratory model (PT-JPL ) and MODIS ET algorithm (MOD16). Comparing the observed with simulated monthly ET by the three models over 43 Fluxnet sites, we found that SiTH overestimates ET for forests, but it performed better than the other two models over short vegetation. MOD16 and PT-JPL models performed well for forests, but poorly in dryland biomes. At the catchment scale, all models perform well expect over some tropical and high latitudinal catchments. At the global scale, SiTH highly overestimated ET in tropics, while PT-JPL underestimated ET between 30°N and 60°N and MOD16 underestimated ET between 15°S and 30°S. This study also revealed that the estimated ET by PT-JPL were largely influenced by the uncertainty in meteorological data, while the estimated ET by SiTH and MOD16 were relatively non-sensitive to the forcing data sets. In addition, the results suggested that the long-term variations in estimated ET trend were greatly influenced by the uncertainty in LAI data.

**Key words:** Evapotranspiration; LAI; Uncertainty; SiTH; MOD16; PT-JPL

## 1. Introducation

Global terrestrial evapotranspiration (ET) is an important nexus between land surface, vegetation and atmosphere, and accurately estimating global land ET is of great significance to study global hydrological cycle, energy exchange, carbon cycle and climate change (Trenberth et al., 2009; Wang and Dickinson, 2012; Fisher et al., 2008; Jung et al., 2010; Miralles et al., 2014). With the rapid developments in remote sensing, numerous global ET models have been developed in recent decades (Norman et al.,1995; Bastiaanssen et al., 1998; Su, 2002; Cleugh et al., 2007; Mu et al., 2007, 2011; Fisher et al., 2008; Leuning et al., 2008; Jung et al., 2009; Zhang et al., 2010; Miralles et al., 2011; Zhu et al., 2019). In the framework of energy balance theory, the majority of these models use the Penman-Monteith equation (P-M equation) (Monteith, 1965) or the Priestley-Taylor approach (P-T approach) (Priestley and Taylor, 1972) to estimate ET. For instance, MOD16 which is the core algorithm of NASA's MODIS evapotranspiration product uses the P-M equation to simulate global ET (Mu et al., 2011). Fisher et al. (2008) proposed a simple, less data-driven and accurate ET model (PT-JPL) to estimate ET on basis of P-T approach. Recently, Zhu et al. (2019) developed a Simple Terrestrial Hydrosphere model (SiTH) to estimate ET based on the P-T approach and the groundwater-soil-plant-atmosphere continuum (GSPAC) theory. These models have been widely used to study regional or global hydrological cycles (Vinukollu et al., 2011a,b; Long et al., 2014; Ramoelo et al., 2014; Ershadi et al., 2014, 2015; Zhang et al., 2019; Hu et al., 2015; Michel et al., 2016; Miralles et al., 2016).

Despite these progresses in model developments, there are still some insufficiencies in systematic inter-comparisons and evaluations of the model performances. First, there is a lack of systematic assessments of the impact of forcing data uncertainties on model performances. As we known, both the vegetation characteristics (LAI) and meteorological variables (i.e., radiation, temperature, precipitation, humidity and air pressure) may have significant influences on model behaviors. For example, LAI can influence the amount of absorbed solar radiation and its distributions between plant canopy and soil surface, which ultimately have significant influences on plant transpiration and soil evaporation (Good et al., 2015; Wang et al., 2014; Wei et al., 2017; Kala et al., 2014). The meteorological conditions (i.e., temperature, humidity and wind speed) regulate the atmospheric evaporation demand, which may have a significant influence on ET as the result of global change (Jung et al., 2010; Zhang et al., 2016; Zeng et al., 2018). Over the past 30 years, it has been well documented that there is a constant increase in LAI (Earth greening) (Chen et al., 2019; Jiang et al., 2017; Zhu et al., 2016), and continuous increasing in land temperature (IPCC, 2018). However, the increasing magnitude and their distributions in different LAI and meteorological datasets were also large (Jiang et al., 2017; Jia et al., 2018; Vinukollu et al., 2011b). Thus, it's urgently needed to evaluate the impacts of the forcing data uncertainties on the estimates of ET. Second, the majority of previous studies have focused on evaluating the performances of one specific model over different sites or inter-comparing the performances of different models at single (or a few) locations. For example, Zhang et al. (2019) evaluated the performance of the PT-JPL over 43 Fluxnet sites. Ershadi et al. (2014) systemtically compared the performances of four ET models over different sites and boimes. To select the best candidate ET model for global applications, it is needed to comprehensively evaluate and intercompare the performances of different models at different (local, regional and global) spatial scales across different biomes and climate conditions. Third, there are still great uncertainties in the long-term changes in ET over the past few decades. For instance, some studies reported that global land ET increased from 1982 to 1998 and then there is a sharp decline until to 2008 (Jung et al., 2010; Zhang et al., 2015; Zhang et al., 2016; Yan et al., 2013). Others studies suggested that an upward trend is observed from 1982 to 2000 and an obvious recovery in ET may have started from 2007 (Mueller et al., 2013; Miralles et al., 2014). To clearly describe the multi-decadal trend in global terrestrial ET, we need to simulate ET using different models with multiple forcing datasets. So that we can properly assess the influences of model structures and forcing data uncertainties on long-term trend of land ET.

In this study, we intercompare the performances of three process-based ET models (MOD16, PT-JPL and SiTH) at different spatial scales, and obtain a long-term trend of ET based on different combinations of LAI and meteorological datasets. Specifically, the goal of this study is to (i) evaluate the performance of three process-based models from local to regional and global scales using different forcing datasets, (ii) analyse the uncertainties due to different model structure and parameterizations, and (iii) explore the uncertainty of long-term temporal ET trend in response to climate change and LAI increasing.

## 2. Methods and data

### 2.1 Models

The SiTH model proposed by Zhu et al. (2019) is a relatively new satellite-based ET model at daily temporal resolution. Based on the framework of the groundwater-soil-plant-atmosphere continuum (GSPAC), SiTH uses well-established hydrological models to simulate important hydrological variables (i.e., groundwater, soil moisture, and runoff). Then, the potential evapotranspiration calculated by using P-T equation was constrained down to actual ET through the plant physiological factor and soil moisture conditions. In SiTH, the total ET consists of canopy interception evaporation, soil evaporation and vegetation transpiration. Soil evaporation is constrained to occur in the first soil layer, while plant transpiration can use both soil water and groundwater.

The MOD16 model, which was proposed by Mu et al. (2007; 2011), estimated ET based on the P-M equation to calculate potential evaporation (Penman and Menteith, 1948). It distributes the available energy into the components of surface soil and vegetation through fractional total vegetation cover. Then, the soil evaporation includes the evaporation from the saturated soil surface and the moist soil surface. Furthermore, canopy water loss includes evaporation from the wet canopy surface and transpiration from the dry surface. Finally, it limits potential ET to actual ET through vegetation physiological factors and meteorological factors. In MOD16, evapotranspiration is equal to the sum of wet canopy evaporation, vegetation transpiration, and bare soil evaporation at day-time and nighttime periods.

PT-JPL is a relatively simple (input data and parameters are reduced), accurate model for estimating actual ET (Fisher et al., 2008; Ershadi et al., 2014; Miralles et al., 2016; Zhang et al., 2017). First, PT-JPL estimates potential evapotranspiration based on the P-T equation (Pristley and Taylor, 1972). Then, plant physiological and ecological constraints (i.e., LAI, green canopy ratio, vegetation temperature and vegetation moisture) are used to limit potential plant transpiration and atmospheric constraints (vapour pressure deficit and relative humidity) to limit potential soil evaporation to actual ET. PT-JPL divides the actual evapotranspiration into three components: canopy transpiration, soil evaporation and interception evaporation.

### 2.2 Input data

The input data of the above three models includes leaf area index (LAI), net radiation (Rn), air temperature (Ta), precipitation (P), air pressure (Pa), relative humidity (RH) and land cover (LC) (Supplementary Table 1). To investigate the influences of vegetation and meteorological variables on ET estimations, three sets of LAI data and two sets of meteorological data were used in this study.

The three long-term LAI products are GLOBMAP (Liu et al., 2012), GLASS (Xiao et al., 2016), and GIMMS LAI3g (Zhu et al., 2013). The GLOBMAP LAI is generated in 8 km and 16-day/8-day resolution from 1981 to 2017, produced by using Advanced Very High Resolution Radiometer (AVHRR) and Moderate Resolution Imaging Spectroradiometer (MODIS) satellite data, and it can be accessed from http://www.globalmapping.org/. The GLASS product is provided in 0.05° and 8 daily and 1 km resolution spanning from 1982 to 2000 (Xiao et al., 2016), produced by

NASA's Long Term Data Record (LTDR) project using NOAA/AVHRR surface reflectance datasets (http://www.glcf.umd.edu/). The GIMMS LAI3g product (version 01) is generated at 1/12° spatial resolution from 1982 to 2011 (http://sites.bu.edu/cliveg/datacodes/; Zhu et al., 2013), based on feed-forward neural networks.

Two sets of meteorological reanalysis data based on observational data retrieval and assimilation are used in this study. The first one is the Modern-Era Retrospective analysis for Research and Applications Version 2 (MERRA-2) from NASA's Global Modeling and Assimilation Office (https://disc.sci.gsfc.nasa.gov/) (Bosilovich et al., 2016). It provides near-surface air pressure and temperature, specific humidity, precipitation, and net radiation at a spatial resolution of 0.5°×0.625° on hourly temporal resolution from 1982 to 2017. The second is the latest ERA-5 produced by European Centre for Medium-Range Weather Forecast (ECMWF) (https://cds.climate.copernicus.eu/). This dataset includes near-surface air pressure and temperature, dew point temperature, precipitation, and net radiation spanning 1982 to 2017 at a spatial resolution of 31km and an hour temporal resolution (Hersbach et al., 2016). In addition, we use static land cover data from MCD12C1 in 2001 (Friedl et al., 2010), because its changes are relatively small on a global scale with time (Zhang et al., 2016). All driving data was interpolated to a 0.25°×0.25° spatial resolution based on the non-linear spatial interpolation method (Zhao et al., 2005). A total of 18 ensembles ET products are obtained from the three ET models with different combinations of inputs data (Supplementary Table 2).

### 2.3 Data used for evaluation

At the local scale, the FLUXNET 2015 (https:// fluxnet.fluxdata.org) from global network of eddy-covariance towers was used for model evaluations (Fisher et al., 2008; Mu et al., 2007, 2011; Zhang et al., 2017; Ershadi et al., 2014). Here, a total of 43 flux sites were selected to cover a wide range of biomes with the energy closure ranging from 70% to 90%. These sites can be divided into 7 different vegetation types on the basis of the IGBP classification, and included the cropland (CRO, 7 sites), deciduous broad-leaved forest (DBF, 4 sites), evergreen broad-leaved forest (EBF, 5 sites), evergreen coniferous forest (ENF, 8 sites), the grass (GRA ,13 sites), the mixed forest (MF, 2sites) and open shrubland (OSH ,4 sites) (Figure 1).

At the catchment scale, the water balance dataset for 32 major (i.e., >200,000 km$^2$) river catchments developed by Pan et al. (2012) was used to evaluate the model performance at regional scale (Supplementary Table 3). This dataset is considered to be the best available water budget dataset (Li et al., 2013; Zeng et al., 2015; Zhang et al., 2017), and includes monthly precipitation, ET, streamflow, and the change in water storage from 1984 to 2006.

At the global scale, the Model tree ensembles (MTE) product (Jung et al., 2009) spans from 1982-2011 at monthly temporal resolution and 0.5° spatial resolution. The MTE model integrates observed ET at the FLUXNET sites with satellite remote sensing and surface meteorological data in a machine-learning algorithm. It is widely used for the comparison and verification of model performances (Miralles et al., 2014; Zhang et al., 2016; Zhu et al., 2019). The Global Land Evaporation Amsterdam Model

(GLEAM) calculates ET from satellite observations based on the P-T equation (Miralles et al., 2011), and performed well in ET estimated (Miralles et al., 2011; Michel et al., 2016).

### 2.4 Analysis method

The statistical measures used to evaluate model performance include the coefficient of determination ($R^2$), slope and the Nash-Sutcliffe efficiency coefficient (NSE) (Nash and Sutcliffe, 1970; Legates and McCabe, 1999). The NSE is a normalized statistic that determines the relative magnitude of the residual variance compared to the measured data variance. It wascomputed as:

$$NSE = 1 - \frac{\sum_{t=1}^{T}[O(t) - S(t)]^2}{\sum_{t=1}^{T}[O(t) - \overline{O}]^2} \quad (1)$$

where O(t) is the observed ET, S(t) is the simulated ET, and Ō is the mean of observed values. NSE values range between -∞ to 1. When the NSE values is closer to 1, the simulation is better.

## 3. Results

### 3.1 Model evaluation at local scale

The model performances on estimating monthly ET driving by different inputs data were compared with the observations from 43 Fluxnet sites (Figure 2). The SiTH model overestimates ET when using all the forcing datasets. The linear regression slope between observed and simulated ET ranges from 1.18 to 1.25, especially for the MERRA-2 (slope=1.23~1.25). The correlation coefficients ($R^2$) of SiTH are highest (greater than 0.73) among the three models, suggesting that the estimated ET by SiTH agreed better with the observed ET. When the same LAI data is used, the estimated ET by SiTH using ERA-5 data is better than that using MERRA-2 data. On the contrary, when the same meteorological data was used, there were small differences in simulated ET between different LAI datasets. Thus, it seemed that the influences of the meteorological data on the performances of SiTH is larger than that of the LAI data. For MOD16 model, it performed relatively well in simulating ET by using different forcing data with the regression slopes close to 1 (slope=0.92~1.0). Importantly, it seemed that the differences in forcing data have small influences on the performances of MOD16 model. However, the correlation coefficients ($R^2$) between observed and simulated ET by MOD16 is generally lower than the other models, indicating the consistencies between estimated and observed ET were relatively low for MOD16, especially using GLOBMAP LAI data ($R^2$=0.55) (Figures 2a, b). For PT-JPL model, its performances using MERRA-2 (slope=0.97~0.99, $R^2$=0.70~0.72) were much better than that using ERA-5 (slope=0.71~0.72, $R^2$=0.61~0.66), indicating that meteorological data have significant influences on model performances.

Due to the differences in model structure and parameterization, the model performances in ET simulations varied over different land surfaces (Massman and Lee, 2002; McCabe et al., 2005; Richardson et al., 2006; Williams et al., 2009; Vinukollu et al., 2011a; Ershadi et al., 2014). Figure 3 shows the performances of the three models across the different biomes. For the forest biomes (ENF, EBF, DBF, MF), the SiTH model generally overestimated ET, while the MOD16 and PT-JPL models performed relatively well at these biomes. The MOD16 model overestimates ET at the

MF biome. However, over the short vegetation types (i.e., GRA, CRO, and OSH), SiTH performs well, while the other two models underestimated ET (slope＜0.8). It was also observed that the MOD16 significantly underestimate ET over the OSH ecosystems (slope=0.25~0.2) (Figure 3). The simulated ET by the three models had good consistency with site-observed ET over forest biomes ($R^2$>0.4). However, the MOD16 and PT-JPL models do not capture the ET dynamics in dryland biomes (OSH) (MOD16:$R^2$=0.02~0.12; PT-JPL:$R^2$=0.12~0.46). On the contrary, the SiTH model was satisfactory across dryland biomes, with $R^2$ values ranging from 0.52 to 0.79. The SiTH and MOD16 models generate negative NSE in forests (except DBF), because they overestimated ET significantly. PT-JPL model has a greater NSE values (closer to 1) in forests. The SiTH produces high NSE for short vegetation, and the PT-JPL, especially MOD16 has the NSE values lower than 0 in OSH vegetation. In addition, compared with SiTH and MOD16 models, PT-JPL model showed great variations in ET simulations within a specific biome.

### 3.2 Model evaluation at catchment scale

At regional scale, SiTH tended to overestimate ET driving by different LAI and meteorological data (Figure 4). The regression slope between observed and estimated ET by SiTH ranged from 1.22 to 1.36. The SiTH model agreed better with WBE data than MOD16 and PT-JPL (SiTH:$R^2$=0.89 ~0.90). The MOD16 model also slightly overestimated ET with values of slope being 1.03 to 1.16, and the $R^2$ values were lower than the other two models ($R^2$=0.71~0.80). The PT-JPL model performed well. The regression slope between observed and PT-JPL estimated ET ranged from 0.87 to

1.03 with $R^2$ varying from 0.79 to 0.86. The estimated ET by MOD16 and SiTH models using different forcing data showed relatively little differences, suggesting that the MOD16 and SiTH models are no-sensitivity to uncertainties of the forcing data. For the PT-JPL model, the differences of estimated ET are greater when using different meteorological datasets.

The performances of the three models over each basin were compared in Figure 5. Both MOD16 and SiTH models overestimated ET over the majority of the catchments, while PT-JPL performs relatively well in most catchments. The estimated ET by three models (especially SiTH) shows high consistency with water balance ET, with the $R^2$ greater than 0.8. But the values of $R^2$ in three models are very low in several basins (Amazon, Congo, Mekong, MOD16 in Aral, Indus, Murray and Olenek ). The PT-JPL model has better NSE (closer to 1) in most catchments than the other two models. Generally, the three models performed poorly over catchments in the tropical rainforest areas (Amazon, Congo, and Mekong). In addition, MOD16 also performed poorly in high latitudes regions (Indigirk, Olenek and Yukon), and Zhujiang region. This may be due to large discrepancies in LAI and meteorological data sets (Jiang et al., 2017; Jia et al., 2018), or lack of a robust description of snow and ice evaporation at high latitudes.

**3.3 Evaluation of ET at global scales**

At global scale, we calculated the annual average ET during 1982-2011 using these models forced by six different combinations of inputs (Figure 6). The spatial pattern of simulated ET by the three models were very similar. The highest annual ET was found over the amazon basin, the Congo rainforest and the southeast Asia near the equator, while the lowest annual ET were in the north Africa, most areas of central Asia, the southwestern United States, the Central and western Australia and some parts of high latitude. The average total annual ET from 1982 to 2011 were (76.87±2.98)× $10^3$ $km^3$, (71.68 ±2.82) × $10^3$ $km^3$, and (61.25±1.92) × $10^3$ $km^3$ for SiTH, MOD16, and PT-JPL model, respectively. These values fell in the ranges from 54.9 × $10^3$ $km^3$ to 85 × $10^3$ $km^3$ reported by previous studies (Oki and Kanae, 2006; Jung et al., 2010; Wang-Erlandsson et al., 2014; Miralles et al., 2016). In addition, the mean annual global land ET calculated from MET and GLEAM during the same period were 63.34 × $10^3$ $km^3$ and 65.7×$10^3$ $km^3$, respectively. These values were slightly lower than that estimated by SiTH and MOD16 model, but very close to that estimated by PT-JPL model. The latitudinal average of ET simulated by the three model during 1982-2011 were also presented Figure 6, and showed similar latitudinal pattern. Relative high ET values were observed near the equator area and near 20°N. However, the estimated ET in tropical regions by SiTH model was higher than that estimated by the other two models. The PT-JPL estimated ET was relatively low for the latitudes 30°N - 60°N , and estimated ET by MOD16 was low for the latitudes between 15°S and 30°S (Vinukollu et al., 2011b).

Figure 7 shows the ensembles of multi-decadal ET anomalies (1982-2017) using three process-based ET models forced by the six different combinations of inputs. Generally, the range of the ensembles of global ET anomalies shows large fluctuations, and the median of long-term global ET ensembles agreed better with the global ET trend by MET (Figure 7a). There were two peaks in the median of global ET ensembles in 1998 (median ET anomalies =17.17 mm/year) and 2010 (median ET anomalies =14.79 mm/year), respectively. Both 1998 and 2010 were El Niño years. There was an increasing trend from 1982 to 1998, and a decreasing trend from 1998 to 2008. After 2008, the ET reached its second summit in 2010. Since the summit in 2010, the global ET fluctuantly declined but reached high positive anomaly in 2016 (mean ET anomalies = 12.56 mm/year).

To evaluate the impact of LAI and meteorological data uncertainties on long-term ET trend, we classified 18 sets of ET products in Supplementary Table 2 into two categories: (1) the estimates of ET from three LAI datasets with different combinations of models and meteorological datasets, which were mainly used to investigate the influences of different LAI datasets on the ET estimations (Figure 7b); and (2) the estimates of ET from two meteorological datasets with different combinations of models and LAI datasets, which were mainly used to investigate the influences of different meteorological datasets on ET estimations (Figure 7c). In Figure 7b, the estimates of ET using the GLASS dataset deviated from those using the GLOBMAP and GIMMS datasets. The estimated ET using the GLASS dataset were lower than those by the other two LAI datasets during 1988-1993 period, while the estimated ET using the GLASS dataset were higher than those by the other two LAI datasets during 1999-2004 period. In Figure 7c, the estimates of ET using the MERRA-2 dataset were very similar to that using the ERA5 data, and their inter-quartile range of ET overlapped greatly and varied synchronously. Also, the median values of ensembles ET using the ERA5 and MERRA-2 datasets were consistent with the ET value estimated by MET. This indicated that the influences of LAI datasets on the estimated long-term variations in global ET were higher than those of meteorological datasets.

## 4. Discussion

### 4.1 Analyzing the performance of models

The three models with different structural complexity and process parameterizations have been developed to predict global ET. Hence, these models are expected to present various performance in estimating ET (Vinukollu et al., 2011a; Mueller et al., 2013). In this study, we found that the SiTH model overestimated ET over some specific biomes (i.e., forest) and in the tropical regions. However, this model exhibited relatively high consistency with the observations ( $R^2$=0.6~0.88). In this model, the P-T coefficient α is a dimensionless factor associated with the Bowen ratio to limit evaporation, and its value of 1.26 is derived from the data of daily fluxes at saturated land sites and open water (Priestley and Taylor, 1972). Many studies revealed that the value of α for forests may be below 1.26. For examples, Komatsu et al. (2005) reported that the value of α is 0.83 ± 0.15 at deciduous forests and α =0.63 ± 0.2 at coniferous forests. Cho et al. (2012) discovered that the mean value of α for deciduous forests is 1.01 and for the coniferous forests is only 0.75. Sanches et al. (2010) found that α values for forests are around 0.65. Therefore, the overestimates of ET by SiTH is due to the high α value used in this model for forest ecosystems. On the contrary, SiTH has a good performance over short vegetation ecosystems (i.e., grassland, cropland and shrubland). For the well-watered short vegetations, the value of α=1.26 is confirmed in the literature (Priestley and Taylor, 1972). Pereira et al. (2007) reported that the value of α at the irrigated croplands ranged from 1.17 to1.35. Some researchers found that the value of α close to the 1.26 over shrublands (Owe and Van de Griend, 1990; Caylor et al., 2005). So, the SiTH performed relatively well over short vegetation ecosystems (i.e., grassland, cropland and shrubland) (Figure 4). Notice that the α value may systematically vary on the daily and seasonal cycles (Tongwane et al., 2017; Assouline et al., 2016; Komatsu et al., 2005). So, it should take this variation into account in the future studies and optimize the value of α in SiTH for forests to improve its accuracy in ET simulations.

We also found that both PT-JPL and MOD16 performed poor in dryland biomes (OSH), which is consistent with the previous studies (Zhu et al., 2016a; Garcia et al., 2013; Zhang et al, 2017; Vinukollu et al., 2011b; Sun et al., 2012; Velpuri et al., 2013; Michel et al., 2016; Zhang et al, 2019; Ershadi et al., 2014). In arid areas, soil moisture is the main factor to that influences ET processes. However, the PT-JPL and MOD16 models used the atmospheric moisture conditions (i.e., air temperature, RH and VPD) to reflect soil moisture constraints on ET (Fisher et al., 2008; Mu et al., 2007, 2011), rather than directly using the soil moisture to constrain the ET. Recently, Novick et al. (2016) reported that atmospheric moisture conditions may be significantly correlated with soil moisture at month to year time scales, but they tended to be nearly decoupled at the daily and hourly time scales. Thus, the PT-JPL and MOD16 models using the atmospheric moisture conditions to limit soil moisture may not properly describe the restricts of soil moisture on ET in arid regions. Furthermore, the soil moisture constraint was calculated as $RH^{VPD/\beta}$ in MOD16 and PT-JPL models. The parameter β is the sensitivity of soil moisture constraint to VPD, and plays an important role in accurate estimation of soil evaporation (Fisher et al., 2008; García et al., 2013; McCabe et al., 2016; Zhu et al., 2016a; Zhang et al., 2017). Zhang et al. (2017; 2019) found that β was the most sensitive parameter and its values in arid area were lower than that in humid regions, resulting in low soil evaporation due to soil moisture stress. On the contrary, the SiTH model directly uses soil water content to describe soil moisture limitation, and performs relatively well in arid areas (Zhu et al., 2019). Finally, plants have deep roots in arid regions (Fan et al., 2013, 2017) and can utilize deeper soil moisture or even groundwater to maintain growth (Thompson et al., 2011). However, only SiTH among the three models took the influences of groundwater into account during ET modeling .

**4.2 Impact of the uncertainties of forcing data on ET**

Models tend to exhibit different behavior when forced with different input data (Vinukollu et al., 2011b; Ferguson et al., 2010; Ershadi et al., 2014). From the evaluation of monthly estimated ET by the three models using different combinations of forcing datasets at local and catchment scales (Figure 2 and Figure 4), the SiTH model overestimated ET with all forcing datasets. The differences of simulated ET by SiTH model are relatively small by using different forcing data, although a slight improvement in performances were observed by using ERA5 meteorological dataset. The MOD16 model performed well and robust using different combinations of the forcing datasets, especially using the MERRA-2 meteorological data. Then, it seemed that MOD16 model is relatively non-sensitive and stable to the forcing datasets. On the contrary, the differences in ET estimated by PT-JPL model are large when using different meteorological forcing datasets. The PT-JPL performs well when using the ERA5 meteorological datasets. Generally, the forcing data had relatively little influence on the estimated ET in the SiTH and MOD16 models. But the meteorological data has more influence on the estimated ET at monthly scale than LAI data in PT-JPL model.

Moreover, the differences were found in estimated global ET anomalies by three models using different combination of forcing datasets, but the ensemble median of global ET anomalies agreed well with the MTE-estimated global land ET anomalies (Figure 7a). It indicates that ensemble-model method can well capture the uncertainties in ET estimates (Ershadi et al., 2014; Zhang et al., 2016; Vinukollu et al., 2011b; Mueller et al., 2013). Generally, the global ET shows an increasing trend from 1982 to 1998. After the summit in 1998, global ET shows a decreasing trend from 1998 to 2008. (Figure 7a). This agrees well with the results of previous studies (Jung et al., 2010; Yan et al., 2013; Zhang et al., 2015; Zhang et al., 2016). Some authors thought that the decline in global ET from 1998 to 2008 was caused by ENSO-induced anomalous dry conditions and consequent limited moisture supply, especially in the Southern Hemisphere (Jung et al., 2010; Yan et al., 2013). However, the decline of ET was transient, and global land ET reached another summit in 2010 (Figure 7a). It demonstrated a transition from El Niño phase to the La Niña phase in

2010 with high precipitation (Poulter et al., 2014), leading to a high ET (Yan et al., 2013). Thus, the decline of ET after 1998 was a transient variation but not a constant decline signal.

Comparing the global ET anomalies at annual scale under different forcing datasets, we found that the influences of LAI datasets on the estimated long-term variations in global ET were higher than those of meteorological datasets (Figure 7b and c). This result was consistent with previous studies which found that vegetation greening is main driver to the multi-decadal ET trend since 1980s (Zhang et al., 2015; Zhang et al., 2016; Zeng et al., 2018; Forzieri et al., 2020; Piao et al., 2020). The different meteorological variables (i.e., net radiation, temperature,precipitation and relative humidity ) have opposite or negative effects on ET process, which may blur the capabilities of ET to identify climate trends at the annual scale (Zhang et al., 2015). In Figure 7b, the estimates of ET anomalies using the GLASS dataset showed large inconsistency with those using the GIMMS and GLOBMAP LAI datasets. Comparing the four long-term LAI products, Jiang et al. (2017) found that interannual variabilities of GLASS LAI products shows large differences with other LAI products. The differences of estimated ET anomalies using the different meteorological datasets is relatively small. The ensemble median values of global ET using ERA5 and MERRA-2 datasets are in agreement with the MTE-estimated global land ET anomalies. Thus, it seemed that the influences of meteorological datasets on global ET estimates can be ignored partially due to their relatively good qualities.

## 5. Conclusions

In this study, we evaluated the performances of three process-based ET models in ET simulations at multiple scales by using variuos LAI and meteorological forcing datasets. The results showed that SiTH simulated well in dryland short vegetation ecosystems, but overestimated ET in forest ecosystems because the P-T coefficient (α) may be set too high in this model. The PT-JPL and MOD16 models performed well in forests, but poorly in dryland biomes due to their improper description of soil moisture stress based on atmospheric moisture conditions. Similar model performances were observed at both catchment and global scales. To obtain proper long-term global ET estimates, the multi-model ensemble approach is a proper choice. We found that the ensembles median of global ET anomalies from different models and forcing datasets showed good consistency with that obtained by the MTE method. Generally, the LAI datasets have larger influences on the global ET estimates than the meteorological datasets. In further studies, we will pay more attentions in optimizing the P-T coefficient (α) over different vegetation types for SiTH to improve its accuracy in ET simulations over forest ecosystems. Finally, more studies are need to quantify the contributions of different driving factors to the variations in global ET, and to figure out the mechanisms in controlling global ET changes.

**Acknowledgements:**

We thank all the principal investigators and their teams of all the dataset used in this study. This work used eddy covariance data acquired and shared by the FLUXNET community, including these networks: AmeriFlux, AfriFlux, AsiaFlux, CarboAfrica, CarboEuropeIP, CarboItaly, CarboMont, ChinaFlux, Fluxnet-Canada, GreenGrass, ICOS, KoFlux, LBA, NECC, OzFlux-TERN, TCOS-Siberia, and USCCC, available at https://fluxnet.fluxdata.org. The water balance datasets are provided by Pan Ming et al. The GLOBMAP datasets are available at http://www.globalmapping.org/. The GLASS datasets are acquired from http://www.glcf.umd.edu/. The MERRA-2 reanalysis datasets are available at https://disc.sci.gsfc.nasa.gov. The ERA5 reanalysis datasets are acquired from https://cds.climate.copernicus.eu/. The MCD12C1 products used in this study are available at https://lpdaac.usgs.gov/. The GLEAM datasets are available at https://www.gleam.eu/. The MTE datasets are acquired from http://www.bgc-jena.mpg.de/geodb/. This research was supported by the National Key Research and Development Program of China (2018YFC0406602), and National Natural Science Foundation of China (41871078).

**References:**

Allen, M. R. & Ingram, W. J., 2002. Constraints on future changes in climate and the hydrologic cycle. Nature 419, 224–232.

Assouline, S., Li, D., Tyler, S., Tanny, J., Cohen, S., Bou-Zeid, E., Parlange, M., and Katul, G. G., 2016. On the variability of the Priestley-Taylor coefficient over water bodies, Water Resour. Res. 52, 150–163, doi:10.1002/2015WR017504.

Bastiaanssen, W. G. M., Menenti, M., Feddes, R. A., and Holtslag, A. A. M., 1998. A remote sensing surface energy balance algorithm for land (SEBAL)–1.

Formulation, J. Hydrol. 212(1–4), 198–212.

Bosilovich, M. G., Lucchesi, R., and Suarez, M., 2016. MERRA-2:File Specification. GMAO Office Note No. 9 (Version 1.1), 73 pp, available from http://gmao.gsfc.nasa.gov/pubs/office _notes.

Caylor, K. K., Shugart, H. H., and Rodriguez-Iturbe, I., 2005. Tree canopy effects on simulated water stress in southern African savannas, Ecosystems. 8, 17–32.

Chen, J. M., Ju, W., Ciais, P. et al., 2019. Vegetation structural change since 1981 significantly enhanced the terrestrial carbon sink. Nat Commun. 10, 4259. https://doi.org/10.1038/s41467 -019-12257-8.

Cho, J., Oki, T., Yeh, P.-F., Kim, W., Kanae, S., Otsuki, K., 2012. On the relationship between the Bowen ratio and the near-surface air temperature. Theor. Appl. Climatol. 108 (1-2), 135–145.

Cleugh, H.A., Leuning, R., Mu, Q., and Running, S.W., 2007. Regional evaporation estimates from flux tower and MODIS satellite data. Remote Sensing of Environment. 106, 285–304.

Hersbach, H, Dee, D., 2016. ERA5 reanalysis is in production, ECMWF Newsletter, p 147.

Ershadi, A., McCabe, M. F., Evans, J. P., Chaney, N. W., and Wood, E. F., 2014. Multi-site evaluation of terrestrial evaporation models using FLUXNET data. Agr. Forest Meteorol. 187, 46–61. http://doi:10.1016/j.agrformet.2013.11.008.

Ershadi, A., McCabe, M. F., Evans, J. P., & Wood, E. F., 2015. Impact of model structure and parameterization on Penman–Monteith type evaporation models.

Journal of Hydrology, 525, 521-535. https://doi.org/10.1016/j.jhydrol.2015.04.008.

Fan, Y. , Li, H. , & Miguez-Macho, G., 2013. Global patterns of groundwater table depth. Science, 339(6122), 940-943.

Fan, Y. , Miguez-Macho, G. , Jobbágy, Esteban G, Jackson, R. B. , & Otero-Casal, C., 2017. Hydrologic regulation of plant rooting depth. Proceedings of the National Academy of Sciences, 201712381.

Ferguson, C. R., Sheffield, J., Wood, E. F., & Gao, H., 2010. Quantifying uncertainty in a remote sensing-based estimate of evapotranspiration over continental USA, International Journal of Remote Sensing, 31:14, 3821-3865. http://doi:10.1080/01431161.2010.483490.

Fisher, J. B., Tu, K. P., & Baldocchi, D. D., 2008. Global estimates of the land–atmosphere water flux based on monthly AVHRR and ISLSCP-II data, validated at 16 FLUXNET sites. Remote Sensing of Environment. 112, 901–919.

Forzieri, G., Miralles, D. G., Ciais, P., et al. 2020. Increased control of vegetation on global terrestrial energy fluxes. Nat. Clim. Chang. https://doi.org/10.1038/s41558-020-0717-0.

Friedl, M. A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., and Huang, X., 2010. MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets.Remote Sensing of Environment, 114(1):168–182.

García, M., Sandholt, I., Ceccato, P., Ridler, M., Mougin, E., Kergoat, L., Morillas, L., Timouk, F., Fensholt, R., and Domingo, F., 2013. Actual evapotranspiration in drylands derived from in situ and satellite data: Assessing biophysical constraints. Remote Sensing of Environment 131, 103–118.

Good, S. P., Noone, D., & Bowen, G., 2015. Hydrologic connectivity constrains partitioning of global terrestrial water fluxes. Science. 349 (6244), 175–177.

IPCC, 2018. Summary for Policymakers. In: Global warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways,in the context of strengthening the global response to the threat of climate change, sustainable development,and efforts to eradicate poverty. World Meteorological Organization, Geneva, Switzerland,32 pp.

Jia, A., Liang, S., Jiang, B., Zhang, X., & Wang, G., 2018. Comprehensive assessment of global surface net radiation products and uncertainty analysis. Journal of Geophysical Research: Atmospheres, 123, 1970–1989. https://doi.org/10.1002/2017JD027903.

Jiang, C., Ryu, Y., Fang, H., Myneni, R., Claverie, M., Zhu, Z., 2017. Inconsistencies of interannual variability and trends in long-term satellite leaf area index products. Glob Change Biol. 23, 4133–4146. https://doi.org/10.1111/gcb.13787.

Jung, M., Reichstein, M., and Bondeau, A., 2009. Towards global empirical upscaling of FLUXNET eddy covariance observations: validation of a model tree ensemble approach using a biosphere model. Biogeosciences 6, 2001–2013.

Jung, M., Reichstein, M., Ciais, P., Seneviratne, S. I., Sheffield, J., Goulden, M. .L, Bonan, G., Cescatti, A., Chen, J.Q., de Jeu R., Dolman, A.J, Eugster, W., Gerten, D., Gianelle, D., Gobron, N., Heinke, J., Kimball, J., Law, B. E., Montagnani, L., Mu, Q. Z., Mueller, B., Oleson, K., Papale, D., Richardson, A. D., Roupsard, O., Running, S., Tomelleri, E., Viovy, N., Weber, U.,Williams, C., Wood, E., Zaehle, S., Zhang, K., 2010. Recent decline in the global land evapotranspiration trend due to limited moisture supply. Nature. 467, 951–954.

Kala, J., Decker, M., Exbrayat, J., Pitman, A. J., Carouge, C., Evans, J. P., Abramowitz, G., and Mocko, D., 2014. Influence of Leaf Area Index Prescriptions on Simulations of Heat, Moisture, and Carbon Fluxes. J. Hydrometeor. 15, 489–503. https://doi.org/10.1175/JHM-D-13-063.1.

Komatsu, H., 2005. 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(19), 3873–3896.

Legates, D. R., & McCabe, G. J., 1999. Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. Water Resources Research, 35, 233–241. https://doi.org/10.1029/1998WR900018.

Leuning, R., Zhang, Y. Q., Rajaud, A., Cleugh, H., and Tu, K., (2008). A simple surface conductance model to estimate regional evaporation using MODIS leaf area index and the Penman-Monteith equation: MODIS-LAI-BASED EVAPORATION MODEL. Water Resources Research 44.

Liu, Y., Liu, R., & Chen, J. M., 2012. Retrospective retrieval of long-term consistent global leaf area index (1981-2011) from combined AVHRR and MODIS data. Journal of Geophysical Research: Biogeosciences. 117,G04003.

Long, D., Longuevergne, L., & Scanlon, B. R., 2014. Uncertainty in evapotranspiration from land surface modeling, remote sensing, and GRACE satellites. Water Resources Research. 50(2), 1131–1151. https://doi.org/10.1002/2013WR014581.

Massman, W. J., and Lee, X., 2002. Eddy covariance flux corrections and uncertainties in long-term studies of carbon and energy exchanges. Agric. Forest Meteorol. 113(1–4), 121–144.

McCabe, M. F., Ershadi, A., Jimenez, C., Miralles, D. G., Michel, D., & Wood, E. F., 2016. The GEWEX LandFlux project: evaluation of model evaporation using tower-based and globally gridded forcing data. Geoscientific Model Development. 9(1), 283–305. https://doi.org/10.5194/gmd-9-283-2016.

Michel, D., Jiménez, C., Miralles, D. G., Jung, M., Hirschi, M., Ershadi, A., Martens, B., McCabe, M.F., Fisher, J. B., Mu, Q. Z., et al., 2016. The WACMOS-ET project – Part 1: Towerscale evaluation of four remote-sensing-based evapotranspiration algorithms. Hydrology and Earth System Sciences. 20, 803–822. https://doi.org/10.5194/hess‐20‐803‐2016.

Miralles, D. G., De Jeu, R. A. M., Gash, J. H., Holmes, T. R. H., and Dolman, A. J., 2011. Magnitude and variability of land evaporation and its components at the global scale. Hydrology and Earth System Sciences. 15, 967–981.

Miralles, D. G., Jiménez, C., Jung, M., Michel, D., Ershadi, A., McCabe, M. F., Hirschi, M., Martens, B., Dolman, A. J., Fisher, J. B., et al. 2016. The WACMOS-ET project – Part 2: Evaluation of global terrestrial evaporation data sets. Hydrology and Earth System Sciences. 20, 823–842.

Miralles, D. G., van den Berg, M., Gash, J., et al. 2014. El Niño–La Niña cycle and recent trends in continental evaporation. Nature Clim Change. 4, 122–126. https://doi.org/10.1038/ nclimate 2068.

Monteith, J. L., 1965. Evaporation and environment, in The State and Movement of Water in Living Organisms, Symposium of the Society of Experimental Biology. pp. 205–234 , Cambridge Univ. Press, Cambridge.

Mu, Q. Z., Heinsch, F. A., Zhao, M., & Running, S. W., 2007. Development of a global evapotranspiration algorithm based on MODIS and global meteorology data. Remote Sensing of Environment. 111, 519−536.

Mu, Q. Z., Zhao, M., & Running, S. W., 2011. Improvements to a MODIS global terrestrial evapotranspiration algorithm. Remote Sens. Environ. 115 (8), 1781–1800.

Mueller, B. et al., 2013. Benchmark products for land evapotranspiration: LandFlux-EVAL multi-dataset synthesis. Hydrol. Earth Syst. Sci. 17, 3707–3720 .

Nash, J. E., Sutcliffe, J. V., 1970. River flow forecasting through conceptual models part I-a discussion of principles. J. Hydrol. 10 (3), 282–290.

Norman, J. M., Kustas, W. P., and Humes, K. S., 1995. Source approach for estimating soil and vegetation energy fluxes in observations of directional radiometric surface-temperature. Agric. For. Meteorol. 77(3–4), 263–293.

Novick, K., Ficklin, D., Stoy, P., et al. 2016. The increasing importance of atmospheric demand for ecosystem water and carbon fluxes. Nature Clim Change. 6, 1023–1027. https://doi.org/10.1038/nclimate3114.

Oki T, Kanae S., 2006. Global hydrological cycles and world water resources. Science. 313,1068-1072.

Owe, M. and Van de Griend, A. A., 1990. A daily surface moisture model for large area semi-arid land application with limited climate data. J. Hydrol. 121, 119–132.

Pan, M., Sahoo, A. K., Troy, T. J., Vinukollu, R. K., Sheffield, J., & Wood, E. F., 2012. Multisource estimation of long-term terrestrial water budget for major global river basins. Journal of Climate. 25(9), 3191–3206. https://doi.org/10.1175/JCLI-D-11-00300.1.

Penman, H. L., 1948. Natural evaporation from open water, bare soil and grass, P. Roy. Soc. Lond. A. 193, 120–145.

Pereira, A. R., Green, S. R., Villa Nova, N. A., 2007. Sap flow, leaf area, net radiation and the Priestley-Taylor formula for irrigated orchards and isolated trees. Agricultural Water Management. v. 92, p. 48–52. https://doi.org/10.1016/j.agwat.2007.01.012.

Piao, S., Wang, X., Park, T., et al. 2020. Characteristics, drivers and feedbacks of global greening. Nat Rev Earth Environ. 1, 14–27. https://doi.org/10.1038/s43017-019-0001-x.

Pinzon, J., & Tucker, C., 2014. A non-stationary 1981-2012 AVHRR NDVI3g time series. Remote Sensing, 6, 6929–6960.

Priestley, C. H. B., and Taylor, R. J., 1972. On the assessment of surface heat flux and evaporation using large-scale parameters. Mon. Weather Rev. 100, 81–92.

Poulter, B., Frank, D., Ciais, P. et al., 2014. Contribution of semi-arid ecosystems to interannual variability of the global carbon cycle. Nature 509, 600-603. https://doi.org/10.1038/nature13376

Ramoelo, A., Majozi, N., Mathieu, R., Jovanovic, N., Nickless, A., & Dzikiti, S., 2014. Validation of Global Evapotranspiration Product (MOD16) using Flux Tower Data in the African Savanna, South Africa. Remote Sensing. 6(8), 7406–7423. https://doi.org/10.3390/rs6087406.

Richardson, A. D., Hollinger, D. Y., Burba, G. G., Davis, K. J., Flanagan, L. B., Katul, G. G.,William, M. J., Ricciuto, D. M., Stoy, P. C., Suyker, A. E., Verma, S. B., Wofsy, S. C., 2006. A multi-site analysis of random error in tower-based measurements of carbon and energy fluxes. Agric. Forest Meteorol. 136 (1–2), 1–18.

Sanches, L., & Alves, M. C., Campelo Júnior, J. H., Nogueira, J. S., Dalmagro, H. J., 2010. Estimativa do coeficiente Priestley–Taylor em floresta monodominante cambarazal no Pantanal. Rev. Bras. Meteorol. 25, 448–454.

Sun, J., Salvucci, G. D., and Entekhabi, D., 2012. Estimates of evapotranspiration from MODIS and AMSR-E land surface temperature and moisture over the Southern Great Plains. Remote Sensing of Environment. 127, 44–59.

Su, Z., 2002. The Surface Energy Balance System (SEBS) for estimation of turbulent heat fluxes. Hydrology and Earth System Sciences. 6, 85−99.

Thompson SE, Harman CJ, Konings AG, Sivapalan M, Neal A, Troch PA., 2011. Comparative hydrology across AmeriFlux sites: The variable roles of climate, vegetation, and groundwater. Water Resources Research 47: W00J07. DOI: 10.1029/2010WR009797

Tongwane, M. I., Savage, M. J., Tsubo, M., Moeletsi, M. E., 2017. Seasonal variation of reference evapotranspiration and Priestley-Taylor coefficient in the eastern Free State, South Africa. Agric. Water Manage. 187, 122–130.

Trenberth KE, Fasullo JT, Kiehl J., 2009. Earth's global energy budget. Bull Am Meteorol Soc. 90, 311-323.

Velpuri, N. M., Senay, G. B., Singh, R. K., Bohms, S., and Verdin, J. P., 2013. A comprehensive evaluation of two MODIS evapotranspiration products over the conterminous United States: Using point and gridded FLUXNET and water balance ET. Remote Sensing of Environment 139, 35–49.

Vinukollu, R.K., Meynadier, R., Sheffield, J., and Wood, E.F., 2011a. Multi-model, multi-sensor estimates of global evapotranspiration: climatology, uncertainties and trends: MULTIMODEL, MULTI-SENSOR ESTIMATES OF GLOBAL EVAPOTRANSPIRATION. Hydrological Processes 25, 3993–4010.

Vinukollu, R. K., Wood, E. F., Ferguson, C. R., Fisher, B., 2011b. Global Estimates of Evapotranspiration for Climate Studies using Multi-Sensor Remote Sensing Data:

Evaluation of Three Process-Based Approaches. Remote Sensing of Environment. 115(3). http://doi:10.1016/j.rse.2010.11. 006.

Wang-Erlandsson, L., van der Ent, R. J., Gordon, L. J., and Savenije, H. H. G., 2014. Contrasting roles of interception and transpiration in the hydrological cycle – Part 1: Temporal characteristics over land, Earth Syst. Dynam., 5, 441–469, doi:10.5194/esd-5-441-2014, 2014.

Wang, K. C., Dickinson, R. E., 2012. A review of global terrestrial evapotranspiration:observation, modeling, climatology, and climatic variability. Rev. Geophys. 50, 54.

Wang, L., Good, S. P., Caylor, K. K., 2014. Global synthesis of vegetation control on evapotranspiration partitioning. Geophys. Res. Lett. 41 (19), 6753–6757.

Wei, Z. W., et al., 2017. Revisiting the contribution of transpiration to global terrestrial evapotranspiration. Geophys. Res. Lett. 44 (6), 2792–2801.

Williams, M., Richardson, A. D., Reichstein, M., Stoy, P. C., Peylin, P., Verbeeck, H., Carvalhais, N., Jung, M., Hollinger, D. Y., Kattge, J., Leuning, R., Luo, Y., Tomelleri, E., Trudinger, C. M., Wang, Y. P., 2009. Improving land surface models with FLUXNET data. Biogeosciences 6 (7), 1341–1359.

Xiao, Z., Liang, S., Wang, J., Xiang, Y., Zhao, X., & Song, J., 2016. Long-time-series global land surface satellite leaf area index product derived from MODIS and AVHRR surface reflectance. IEEE Transactions on Geoscience and Remote Sensing. 54, 5301–5318.

Yan, H., Yu, Q., Zhu, Z. C., Myneni, R. B., Yan, H. M., Wang, S. Q., and

Shugart, H. H., 2013. Diagnostic analysis of interannual variation of global land evapotranspiration over 1982–2011: Assessing the impact of ENSO. J. Geophys. Res. Atmos. 118, http://doi:10.1002/jgrd.50693.

Zeng, Z., Peng, L., & Piao, S., 2018. Response of terrestrial evapotranspiration to Earth's greening. Curr. Opin. Environ. Sustain. 33, 9–25.

Zhang, K., Kimball, J. S., Nemani, R. R., Running, S. W., 2010. A continuous satellite-derived global record of land surface evapotranspiration from 1983 to 2006. Water Resources Research. 46: W09522. http://doi: 10.1029/2009WR008800.

Zhang, K., Kimball, J., Nemani, R., et al., 2015. Vegetation Greening and Climate Change Promote Multidecadal Rises of Global Land Evapotranspiration. Sci Rep. 5, 15956. https://doi.org/10.1038/srep15956.

Zhang, K., Kimball, J.S., and Running, S.W., 2016. A review of remote sensing based actual evapotranspiration estimation: A review of remote sensing evapotranspiration. Wiley Interdisciplinary Reviews: Water 3, 834–853.

Zhang, K., Ma, J. Z., Zhu, G. F., Ma, T., Han, T., and Feng, L. L., 2017. Parameter sensitivity analysis and optimization for a satellite-based evapotranspiration model across multiple sites using Moderate Resolution Imaging Spectroradiometer and flux data, J. Geophys. Res. Atmos. 122, 230–245. http://doi:10.1002/2016JD025768.

Zhang, K., Zhu, G. F., Ma, J. Z., Yang, Y. Q., Shang, S. S., & Gu, C. J., 2019. Parameter analysis and estimates for the MODIS evapotranspiration algorithm andmultiscale verification. Water Resources Research. 55, 2211–2231. https://doi.org/10.1029/2018WR023485.

Zhang, Y. Q., et al. 2016. Multi-decadal trends in global terrestrial evapotranspiration and its components. Sci. Rep. 6, 19124. http://doi: 10.1038/srep19124.

Zhao, M., Heinsch, F. A., Nemani, R. R., & Running, S. W., 2005. Improvements of the MODIS terrestrial gross and net primary production global data set. Remote Sensing of Environment. 95(2), 164–176. https://doi.org/10.1016/j.rse.2004.12.011.

Zhu, G. F., Zhang, K, Chen, H. L., Wang, Y. Q., Sun, Y. H., Zhang, Y., Ma, J. Z., 2019. Development and evaluation of a simple hydrologically based model for terrestrial evapotranspiration simulations. Journal of Hydrology. 577. https://doi.org/10.1016/ j.jhydrol.2019.123928.

Zhu, G. F., Zhang, K., Li, X., Liu, S. M., Ding, Z. Y., Ma, J. Z., Huang, C. L., Han, T., and He, J. H., 2016a. Evaluating the complementary relationship for estimating evapotranspiration using the multi-site data across north China. Agricultural and Forest Meteorology. 230–231, 33–44.

Zhu, Z., Bi, J., Pan, Y., Ganguly, S., Anav, A., Xu, L., ... Myneni, R. B., 2013. Global data sets of vegetation leaf area index (LAI)3g and fraction of photosynthetically active radiation (FPAR)3g derived from global inventory modeling and mapping studies (GIMMS) normalized difference vegetation index (NDVI3G) for the period 1981 to 2011. Remote Sensing. 5, 927–948.

Zhu, Z., Piao, S., Myneni, R., et al. 2016b. Greening of the Earth and its drivers.

Nature Clim Change. 6, 791–795. https://doi.org/10.1038/nclimate3004.

## Figure lists

**Figure 1.** Location of the 43 FLUXNET sites. The biomes types are identified based on the International Geosphere-Biosphere Programme (IGBP) biome classification.

**Figure 2**. Scatter plots of the simulated ET of three models using different forcing datasets versus measured ET at 43 flux sites. On the left panels, the same meteorological dataset of ERA5 and LAI datasets of (**a**) GLOBMAP, (**c**) GLASS, and (**e**) GIMMS was used to simulate ET, respectively. On the right panels, the same meteorological dataset of MERRA-2 and LAI datasets of (**b**) GLOBMAP, (**d**) GLASS, and (**f**) GIMMS was used to estimate ET, respectively. The black dotted line represents the 1:1 line.

**Figure 3.** Comparison of the simulated ET by three models using different forcing datasets versus observed ET at different biomes. Boxplots show the slope, $R^2$ and NSE statistical significance of simulated ET. The central solid line of each box shows the the median. The bottom and top of boxes represent the 25th and 75th percentiles, respectively. The lower and upper whiskers indicate the minimum and maximum values, respectively. The circles represent outliers.

**Figure 4.** Scatter plots of the simulated ET by three models using different forcing datasets versus measured ET at 32 catchments. On the left panels, the same meteorological datasets of ERA5 and different LAI datasets of (**a**) GLOBMAP, (**c**)

GLASS, and (**e**) GIMMS was used to estimate ET, respectively. On the right panels, the same meteorological datasets of MERRA-2 and different LAI datasets of (**b**) GLOBMAP, (**d**) GLASS, and (**f**) GIMMS was used to estimate ET, respectively. The black dotted line represents the 1:1 line.

**Figure 5.** Comparison of the simulated ET of three models versus water balance ET at 32 catchments. The bottom, middle, and top panels represent the slope, $R^2$, and NSE, respectively. The dots are the median of simulated ET under different forcing datasets. Error bar illustrates the mean ±1 s.d.

**Figure 6.** Spatial patterns of annual mean land evapotranspiration for the SiTH, MOD16, PT-JPL, MTE, and GLEAM from 1982 to 2011. The right panel shows the latitudinal profiles of ET from each models between 55°S and 80°N.

**Figure 7.** Global land ET anomalies. (**a**). Ensembles of global ET anomalies from 1982 to 2017 under all models and the forcing datasets. (**b**). The estimates of ET anomalies from three LAI datasets with different combinations of models and meteorological datasets. The green line (GIMMS), blue line (GLASS) and red line (GLOBMAP) represent the median of ET ensembles using the GIMMS, GLASS and GLOBMAP LAI datasets, respectively. (**c**). The estimates of ET anomalies from two meteorological datasets with different combinations of models and LAI datasets. The blue line (MERRA-2) and pink line (ERA5) is the median of ET ensembles using the MERRA-2 and ERA5, respectively. The shading area indicates the inter-quartile of ET ensembles using different LAI and meteorological datasets.