Coupling physical constraints with machine learning for satellite-derived evapotranspiration of the Tibetan Plateau

Ke Shang [a], Yunjun Yao [a,*], Zhenhua Di [b], Kun Jia [a], Xiaotong Zhang [a], Joshua B. Fisher [c], Jiquan Chen [d], Xiaozheng Guo [a], Junming Yang [a], Ruiyang Yu [a], Zijing Xie [a], Lu Liu [a], Jing Ning [a], Lilin Zhang [e]

[a] *State Key Laboratory of Remote Sensing Science, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China*
[b] *State Key Laboratory of Earth Surface Processes and Resource Ecology, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China*
[c] *Schmid College of Science and Technology, Chapman University, 1 University Drive, Orange, CA 92866, USA*
[d] *Department of Geography, Environment, and Spatial Sciences, Michigan State University, East Lansing, MI 48823, USA*
[e] *Faculty of Geo-Information and Earth Observation (ITC), University of Twente, 7500 AE Enschede, the Netherlands*

ARTICLE INFO

ABSTRACT

More accurate and process-based satellite evapotranspiration (ET) estimation for the Tibetan Plateau (TP)—the Third Pole of the world—have long been of major interest in hydrometeorology. Combining recent advances in satellite-based ET mechanistic algorithms and data-oriented methods allows ET hybrid modeling by coupling physical constraints with machine learning (ML). Specifically, we developed two hybrid models, a surface conductance-based ML model (ML-Gs) and a soil evaporation-based ML model (ML-Es), to estimate regional ET on the TP. These hybrid models have biophysical framework, under which one of the parameters or components is modeled using ML. Hybrid models make ML complementary to the process-based ET framework, which to find an optimal junction between well physical mechanism and high model performance. The daily ET estimates were evaluated at 28 eddy covariance flux tower sites, as well as by comparison with two process-based ET algorithms (a Penman–Monteith-based ET-PM algorithm and a Priestley-Taylor-based ET-PT algorithm) and a data-oriented pure ML method. The hybrid models decreased the root-mean-square-error (RMSE) of two physical algorithms (1.11 mm/day for ET-PM, 1.09 mm/day for ET-PT) to 0.50 mm/day, and increased the Kling-Gupta efficiency (KGE) (0.35 for ET-PM, 0.36 for ET-PT) to 0.92. Our hybrid models also showed improved performance (KGE of 0.65) than pure ML (KGE of 0.62) at data-sparse regions as well as for the responses to extreme weather events. It indicates that our approach does not only boost the ET simulation accuracy, but also improve the physical understanding of ML-based ET estimation. More importantly, ML-Es focuses on the ET components on the TP and is more well-defined than ML-Gs. An innovation of our approach is that for data-sparse regions and extreme cases, the more robust physical mechanism was coupled, the better generalization performance of hybrid model could achieve. The spatiotemporal ET patterns based on our hybrid models were consistent with the variations in local climatic regions and could provide critical information on the understanding of hydrological processes under the global and regional climate changes.

1. Introduction

Terrestrial evapotranspiration (ET) is a crucial component of the energy budget and hydrological cycle (Kool et al., 2014; Trenberth et al., 2009). The Tibetan Plateau (TP), with an average elevation of >4000 m, is considered the "Third Pole" of the world. The water and heat transfer of the TP may be a key regional driver and amplifier of global climate change (Dong et al., 2012; Zhang et al., 2018). Additionally known as the "Water Tower of Asia", the TP is also the water source for surrounding lowlands (Immerzeel et al., 2010). Therefore, accurate regional ET quantification in TP is of great significance for understanding the ecosystem response to climate change as well as for enumerating the feedback mechanisms between water resources and ecosystem functioning (Fisher et al., 2017; Jasechko et al., 2013; Jung

et al., 2010).

In situ observations can provide accurate point ET measurements, but their sparse distribution in TP limits the accurate acquisition of regional ET. Remote sensing observations combined with process-based flux equations are an effective strategy to quantify regional ET of this region (Chen, 2021; Chen and Liu, 2020). Conventional process-based algorithms, including the Penman-Monteith (PM)-based equation (Monteith, 1965), Priestley-Taylor (PT)-based equation (Priestley and Taylor, 1972), Surface Energy Balance System (SEBS) (Su, 2002), Single-Source model (Bastiaanssen et al., 1998), Two-Source model (Kustas and Norman, 1997; Norman et al., 1995) and empirical/semiempirical algorithms (Wang et al., 2010a), have been widely used for the generation of global and regional satellite ET products (Chen et al., 2019; Fisher et al., 2020; Liang et al., 2020; Martens et al., 2017; Mu et al., 2011). However, regional ET quantification in TP remains a challenging task due to complex physical and biological controls on ET and high variation in geophysical conditions of TP (Allen et al., 2011; Smith et al., 2019). The SEBS-estimated multiyear (2001–2018) average ET of TP was 496 ± 23 mm/year (Han et al., 2021), while the multiyear (1982–2016) average value based on Penman-Monteith-Leuning Version 2 (PML V2) was 353 ± 24 mm/year (Ma and Zhang, 2022). The process-based ET models differ substantially when applied in TP (Han et al., 2021; Ma and Zhang, 2022; Ma et al., 2015; Wang et al., 2018; Yuan et al., 2021). One possible reason is that, as ET substantially changes over time, the static parameterizations of some process-based models limit the dynamic capture of ET throughout different plant functional types, especially for heterogeneous surfaces. With the deluge of satellite- and ground-based observations, machine learning (ML) approaches are increasingly used in hydrology, especially for ET prediction at regional and global scales (Bodesheim et al., 2018; Jung et al., 2019; Li et al., 2018a; Lu and Zhuang, 2010; Trajkovic et al., 2003). Data-oriented ML ET models can collect patterns and insights from the stream of satellite and meteorological data and automatically extract spatiotemporal features to provide more powerful and flexible predictions (Reichstein et al., 2019). However, these data-oriented ET estimates may not be suitable for extrapolation to heterogeneous and data-sparse regions because they are empirically based (Shang et al., 2021; Zhao et al., 2019). There is scope for the further development of ML-based ET models to better adapt to hydroscientific applications.

The coupling of process-based models and ML provides alternative ways to model complex phenomena from natural physical processes (Karpatne et al., 2017). Physical models and ML approaches can be treated as complements to each other instead of independent scientific paradigms (Reichstein et al., 2019). At present, similar studies have been carried out. Zhao et al. (2019) proposed a physics-constrained hybrid model to conserve the surface energy budget of the ET process, which used a modified PM equation as a loss function of the artificial neural network (ANN) to estimate ET at the site scale. However, an important question is whether hybrid model is capable at regional scales with ML replaced the specific process formulations. Koppa et al. (2022) trained a deep learning model to embed a transpiration stress (St) formulation within process-based ET algorithms to yield a global hybrid ET model. But this hybrid model focuses on the ML modeling of transpiration stress of vegetation, which is not suitable for the high-altitude and vegetation-sparse TP. There are still great limitation and uncertainty of these hybrid models used in the TP.

On the sparsely vegetated TP, soil evaporation (ETs) tends to dominate the total ET. Globally, the ETs accounts for 20% to 40% of total ET (Lawrence et al., 2007; Miralles et al., 2016; Schlesinger and Jasechko, 2014; Talsma et al., 2018), and the ratio of TP is >60% (Wang et al., 2018). Hence, the quantification of ETs is vital for the accurate ET estimation on this region. However, there is no consensus on a clear paradigm to describe the complex feedback mechanisms, which has led to substantial uncertainties in the estimation of ETs (Yuan et al., 2021). This provides an opportunity for the coupling of ETs with ML. Besides, the PM-based ET algorithms have been proved to have substantial errors

by many studies, the mainly challenge is to determine the surface conductance (gs) for different climatic, vegetation, and soil water conditions (Mu et al., 2011; Wang and Dickinson, 2012). This is because that the gs varies widely with changes of actual environmental and biological conditions, which makes it hardly to be calibrated (Ma et al., 2015). Despite studies were carried out to estimate gs (Kochendorfer et al., 2011; Nishida et al., 2003; Zhao et al., 2016), due to insufficient theoretical bases and substantial uncertainties, the ET estimates exhibit wide discrepancies across models with different using of gs. The ML methods thus provide an opportunity to replace the traditional expression of gs. However, hybrid models with different coupling frameworks have not been substantially compared, and the linkages between constrained physical mechanism and coupling performance are not clear.

Aiming to mitigate these issues, we explored two promising strategies of ET hybrid modeling for the TP, which apply ML a complement to physical models. The objectives of this study were to (1) develop two hybrid models, a surface conductance-based ML model and a soil evaporation-based ML model; (2) evaluate the two hybrid models based on EC observations at 28 flux tower sites and by comparison with process-based physical models and a data-oriented pure ML model; and (3) implement the mapping of multiyear (2003–2018) mean ET using the two hybrid models to investigate their spatiotemporal performance over the TP.

2. Model description

We proposed two hybrid models, a surface conductance-based machine learning model (ML-Gs) and a soil evaporation-based machine learning model (ML-Es) (Fig. 1). These hybrid models have biophysical framework, under which one of the parameters or components is modeled using ML. The ET estimation of the hybrid models for the TP were compared with a pure ML method and two process-based ET algorithms (Fig. 1).

The same set of input variables for the hybrid models (ML-Gs and ML-Es) and pure ML model were selected: incident solar radiation (Rs), air temperature (Ta), relative humidity (RH), normalized difference vegetation index (NDVI), soil moisture (SM) and land surface temperature (LST). These variables were chosen because they are key parameters in ET mechanisms and involve a wide range of data sources, including meteorological, optical, thermal, and microwave satellite data. The logics of the ET algorithms are described below.

2.1. Physics-guided hybrid ET models

2.1.1. Surface conductance-based machine learning model

To monitor global ET on decadal scales using meteorological data complemented with remote sensing observations, Wang et al. (2010a) provided a semiempirical expression based on PM framework:

$$\lambda E = \lambda E_E + \lambda E_A \tag{1}$$

$$\lambda E_E = \frac{\Delta}{\Delta + \gamma} \bullet (R_n - G) \bullet g_s \tag{2}$$

$$\lambda E_A = \frac{\gamma}{\Delta + \gamma} \bullet VPD \bullet g_a \bullet g_s \tag{3}$$

where λE is the latent heat flux (W m^{-2}), λE_E is the energy control on ET (W m^{-2}), λE_A is the atmospheric control on ET (W m^{-2}), λ is the latent heat of evaporation (J kg^{-1}), Δ is the slope of the saturated vapor pressure vs temperature curve (k Pa °C^{-1}), γ is the psychrometric constant (k Pa °C^{-1}), R_n represents the net radiation and G is the soil heat flux in W m^{-2}. g_s and g_a are the surface and aerodynamic conductance (m s^{-1}), respectively, and VPD represents the vapor pressure deficit of the air (Pa). This semiempirical ET algorithm can be expressed as:
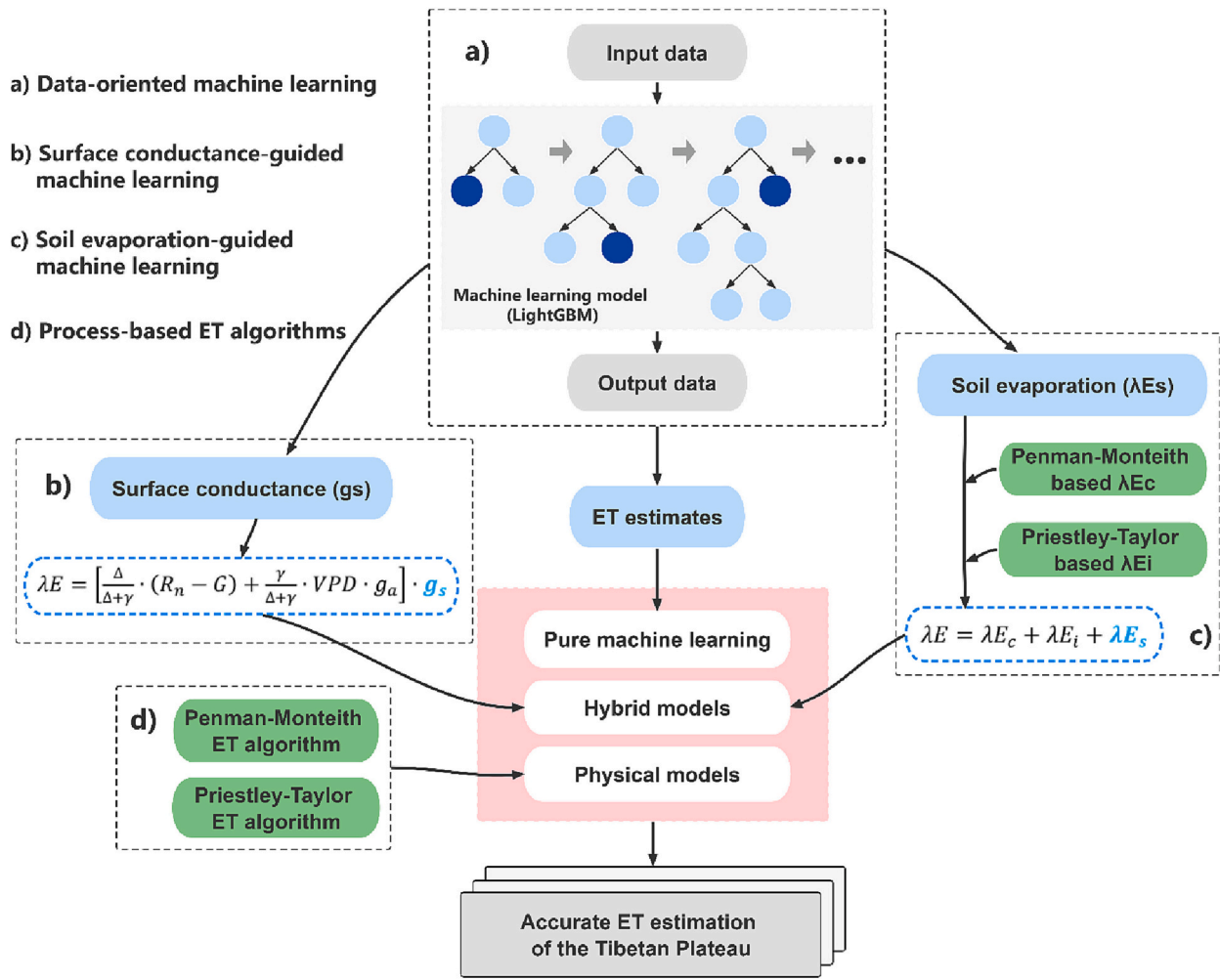
Fig. 1. Development framework of our study. a) Data-oriented pure machine learning, b) and c) physics-guided hybrid ET models, and d) process-based ET algorithms. λE is the latent heat flux and λ is the latent heat of vaporization. λE_c stands for canopy transpiration, λE_i stands for interception evaporation and λE_s represents for soil evaporation.

$$\lambda E = \left[\frac{\Delta}{\Delta + \gamma} \bullet (R_n - G) + \frac{\gamma}{\Delta + \gamma} \bullet VPD \bullet g_a \right] \bullet g_s \qquad (4)$$

The g_a can be parameterized by the wind speed (WS) using $0.26 \bullet (1 + 0.54 \bullet WS)$ (Parlange et al., 1995; Shuttleworth, 2007; Yang et al., 2013). g_s was parameterized by Wang et al. (2010a) using the relative humidity deficit (RHD) and vegetation indices (VI). This algorithm functions well where net radiation and stomatal conductance data are available and the vegetation is not water stressed (Wang et al., 2010b; Yao et al., 2014).

To enhance the simulation performance of Eq. (4) at the TP, we replaced the empirical expression of g_s with an ML model as follows:

$$g_s \sim ML(Rs, Ta, RH, NDVI, SM, LST) \qquad (5)$$

To train this ML model, the input variables include Rs, Ta, RH, NDVI, SM and LST. We have no available ground observations for g_s, and we used the latent heat flux (LE) obtained from EC observations to infer g_s based on Eqs. 1–4. The inferred g_s was used as the target variables to train this ML model. Daily ET estimates based on this ML-Gs algorithm were generated at 1-km spatial resolution.

2.1.2. Soil evaporation-based machine learning model

To further explore the potential possibilities of the hybrid modeling framework, we developed a more complete coupling scheme by calculating ET as the sum of the canopy transpiration (λE_c), interception

evaporation (λE_i) and soil evaporation (λE_s) (Eq. 6). For the expression of canopy transpiration, we used a MOD16-based ET algorithm (Mu et al., 2011), which was calibrated by Yuan et al. (2021):

$$\lambda E = \lambda E_c + \lambda E_i + \lambda E_s \qquad (6)$$

$$\lambda E_c = \frac{(\Delta \bullet R_{nc} + \rho \bullet C_p \bullet VPD \bullet f_c / r_a^c) \bullet (1 - f_w)}{\Delta + \gamma \bullet (1 + r_s^c / r_a^c)} \qquad (7)$$

$$r_a^c = \frac{rh \bullet rr}{rh + rr} \qquad (8)$$

$$r_s^c = \frac{1}{C_L \bullet m(T_{min}) \bullet m(VPD) \bullet LAI} \qquad (9)$$

$$f_w = RH^4 \qquad (10)$$

where R_{nc} is the net radiation absorbed by the vegetation canopy (W m^{-2}); ρ is the air density (kg m^{-3}); C_p is the specific heat capacity of air at constant pressure (J kg^{-1} K^{-1}); f_c and f_w are the vegetation cover fraction and relative surface wetness, respectively; r_a^c is the aerodynamic resistance (s m^{-1}) which is calculated as the parallel resistance to convective (rr) and radiative (rh) heat transfer (Yuan et al., 2021), r_s^c is the surface resistance of the canopy (s m^{-1}), respectively. $m(T_{min})$ and $m(VPD)$ are multipliers that limit the potential stomatal conductance (Mu et al., 2007; Mu et al., 2011). C_L represents the mean potential stomatal

conductance per unit leaf area and was recalibrated to 0.0038 (m s^{-1}) by Yuan et al. (2021) using measurements from five eddy covariance (EC) sites over the TP.

λE_i can be calculated by the formula developed by Fisher et al. (2008) based on the Priestley-Taylor equation (Priestley and Taylor, 1972):

$$\lambda E_i = \alpha \frac{\Delta}{\Delta + \gamma} R_{nc} f_w \tag{11}$$

where α is the PT coefficient of 1.26 for water bodies.

To more accurately simulate soil evaporation, we adopted an ML model to calculate the ETs over the TP:

$$\lambda E_s \sim ML(Rs, Ta, RH, NDVI, SM, LST) \tag{12}$$

To train this ML model, the input variables include Rs, Ta, RH, NDVI, SM and LST. We have no available ground observations for ETs, and we LE obtained from EC observations to infer ETs based on Eqs. 6–11. The inferred ETs was used as the target variables to train this model. The daily ET estimates based on this ML-Es model were estimated with 1-km spatial resolution.

2.2. Data-oriented machine learning model

We selected Light Gradient Boosting Machine (LightGBM) as the ML method used in the hybrid modeling in Eqs. (5) and (12). LightGBM is an enhanced implementation of a widely used ML algorithm—gradient boosting decision tree (GBDT)—based on decision tree algorithms. The GBDT is a popular ML model with superior performance (Friedman, 2001). Despite its state-of-the-art simulation ability in many geoscience-dominant ET prediction studies (Fan et al., 2018; Shang et al., 2020; Wang et al., 2021), the efficiency and scalability of GBDT remain unsatisfactory when the data volume is large (Ke et al., 2017).

To tackle the time consumption problem of GBDT, LightGBM estimates the information gain by using data instances with larger gradients, and proposes a greedy algorithm to scan the approximation ratio, which effectively reduces the dimension of features and ensures the accuracy of split point determination (Ke et al., 2017). A contrastive study on various datasets showed that LightGBM speeds up the training processes of multiple GBDT implementations by up to >20 times while obtaining comparable accuracy (Ke et al., 2017). Moreover, LightGBM can be conveniently implemented in Python. The efficiency, accuracy, and interpretability of LightGBM are the main reasons why we chose it as the representation of various ML approaches to implement the coupling investigation.

We constructed our hybrid models with LightGBM in Python; detailed information on the Python implementation can be found on GitHub (https://github.com/Microsoft/LightGBM). The main parameters to fit our model include the max_depth, num_leaves, subsample, colsample_bytree, learning_rate, etc. These parameters can be tuned to optimal values to fit specific datasets by cross-validation. Through careful tuning, we adopted 10, 150, 0.8, 0.8 and 0.1 for the above parameters, respectively.

In addition to the coupled modeling in ML-Gs and ML-Es, a data-oriented pure ML ET model using LightGBM was also calculated. The comparison between physics-guided hybrid models (ML-Gs, ML-Es) and pure ML model can help us to understand the impacts and constraints of physical mechanisms. We used Rs, Ta, RH, NDVI, SM and LST as the input variables of LightGBM and used EC observations as the target output variable to directly implement the upscaling of ET from flux towers to regional scales without any physical constraints. The daily pure ML ET estimates were generated at 1-km spatial resolution.

2.3. Comparison with process-based ET algorithms

Two process-based ET algorithms driven by satellite and meteorological variables were used to investigate the performance

improvements achieved by the coupling of ML, including a traditional PM-based algorithm and a classic PT-based algorithm.

(1) Penman–Monteith ET algorithm

We adopted a traditional ET algorithm from the MODIS global ET framework (Mu et al., 2007; Mu et al., 2011), which is based on a conventional PM equation (Monteith, 1965):

$$\lambda E_{PM} = \frac{\Delta(R_n - G) + \rho \bullet C_p \bullet VPD/r_a}{\Delta + \gamma \bullet (1 + r_s/r_a)} \tag{13}$$

where r_a and r_s are the aerodynamic resistance and surface resistance (s m^{-1}), respectively. Mu et al. (2011) considered the daily ET as the sum of evaporation from the wet canopy, dry canopy and the soil surface. The surface resistances used in MODIS ET for different plant function types were determined by the Biome Properties Look-Up Table (BPLUT) (Mu et al., 2011). The alpine grasslands and alpine meadows are the main plant functional types of TP (Wang et al., 2018), so we adopted parameter values for grassland in the BPLUT for the optimum air temperature and Vapor Pressure Deficit (VPD) constraints, which reduces the uncertainty from incorrect classification of land cover types. More detailed algorithm information can be found in Mu et al. (2011).

We generated this ET-PM product mainly using the Rn, Ta, Tmin, RH, leaf area index (LAI), and fraction of absorbed photosynthetically active radiation (FPAR) as the forcing data. The daily ET estimates (ET-PM) using the PM algorithm over the TP were produced at 1-km spatial resolution.

(2) Priestley-Taylor ET algorithm

We used a novel Priestley-Taylor-based model proposed by Fisher et al. (2008) to generate daily ET estimates (ET-PT) over the TP. This algorithm was validated during the NASA ECOsystem Spaceborne Thermal Radiometer Experiment on Space Station (ECOSTRESS) mission (Fisher et al., 2020) and can work well on a per-pixel basis at the global scale without ground-based calibration or model tuning (Fisher et al., 2008). The algorithm partitions ET into three parts: soil evaporation, canopy transpiration and interception evaporation:

$$\lambda E_{PT} = \lambda E_{s\_PT} + \lambda E_{c\_PT} + \lambda E_{i\_PT} \tag{14}$$

$$\lambda E_{s\_PT} = \alpha \frac{\Delta}{\Delta + \gamma}(R_{nS} - G)[f_w + f_s(1 - f_w)] \tag{15}$$

$$\lambda E_{c\_PT} = \alpha \frac{\Delta}{\Delta + \gamma} R_{nC} f_c f_{pt} f_{pm}(1 - f_w) \tag{16}$$

$$\lambda E_{i\_PT} = \alpha \frac{\Delta}{\Delta + \gamma} R_{nC} f_w \tag{17}$$

where R_{nS} (W m^{-2}) is the net radiation at the soil surface (R_n exp $(-k_{Rn}\text{LAI})$) (Fisher et al., 2008); k_{Rn} is 0.6 (Impens and Lemeur, 1969); LAI is total leaf area index; R_{nC} (W m^{-2}) is the net radiation to the canopy ($R_n - R_{nS}$); f_s is the soil moisture (SM) constraint; and f_{pt} and f_{pm} are the plant temperature constraint and plant moisture constraint, respectively. More detailed information about this algorithm can be found in Fisher et al. (2008). The forcing data of ET-PT mainly require the Rn, Ta, Tmax, RH, NDVI, LAI, and FPAR. The daily ET-PT estimates of the TP were generated with 1-km spatial resolution.

2.4. Water evaporation and glacier sublimation estimation

For the hybrid modeling of ET over the TP, the water evaporation and glacier sublimation were calculated separately using physical models because there are not enough available EC observations for corresponding biomes to train the ML. The water evaporation was

calculated based on the Penman equation (Penman and Keen, 1948), as shown in Eq. (18), where a_1 is 6.43, a_2 and a_3 are wind function coefficients (taken as 0.5 and 0.54 for open water). The glacier sublimation was estimated using the equation provided by Kuzmin (1953) (Eq. 19), where b_1 was 0.18, and b_2 was 0.098.

$$\lambda E_{water} = \frac{\Delta(R_n - G) + \gamma \bullet a_1 \bullet (a_2 + a_3 \bullet WS) \bullet VPD}{\Delta + \gamma} \tag{18}$$

$$\lambda E_{ice} = (b_1 + b_2 \bullet WS) \bullet VPD \tag{19}$$

3. Data and model validation

3.1. Study domain and in situ observations

The Tibetan Plateau, with an average elevation of >4000 m, is the highest and most extensive plateau of the globe (Fig. 2 a). Surrounded by massive mountains, including the Himalayas, Pamir and Kunlun, the TP has long been known as "the roof of the world" (Liu and Chen, 2000). The TP and adjacent mountains are the sources of five major rivers in Asia, snow and glacial melt are important hydrologic processes (Immerzeel et al., 2010). Strongly influenced by the summer Indian monsoon and East Asian monsoon, the temperature zones of the TP are dominated by plateau subfrigid and plateau temperate, while the arid/humid regions comprise arid, semiarid, semihumid and humid regions with latitudes from north to south (Fig. 2 b) (Wang and Zuo, 2009; Zheng, 1999). The main plant functional types include alpine grasslands, alpine meadows, forests, alpine deserts, alpine shrubs, croplands, etc. (Wang et al., 2018).

In this study, we used 28 eddy covariance (EC) flux tower sites (Fig. 2) collected from the Tibetan Observation and Research Platform (TORP) (Ma et al., 2020; Ma et al., 2008) and the National Tibetan Plateau Data Center (TPDC, https://data.tpdc.ac.cn/), the Heihe Integrated Observatory Network (Liu et al., 2018; Xu et al., 2020), the Qilian Mountains integrated observatory network (Li et al., 2018b), the Chinese Ecosystem Research Network (CERN, http://www.cern.ac.cn/), the FLUXNET (https://fluxnet.org/) and the ChinaFLUX National Flux Observation Network (http://www.chinaflux.org/). The underlying surfaces cover five plant functional types, evergreen needleleaf forests (ENF), croplands (CRO), grasslands (GRA), barren lands (BAR) and wetlands (WET). Table 1 lists detailed information for these EC flux tower sites. For the half-hourly EC observations from all the flux towers, we rejected the sensor malfunction data and uncompleted 30 min data in the raw record (Liu et al., 2018; Xu et al., 2020). The look-up table (LUT) method was used to fill the gaps when data were missing or rejected (Falge et al., 2001). The quality-controlled half-hourly EC observations were aggregated to daily temporal resolution to remain consistent with the final ET estimates. We used the Bowen ratio closure method (Foken, 2008; Twine et al., 2000) to correct the energy imbalance of the original EC ground measurements. Negative values were excluded, and samples during rainy days were omitted to avoid the influences of rain interception and sensor saturation (Medlyn et al., 2011).

3.2. Satellite and meteorological datasets

The input satellite data sources include the 16-day NDVI from MOD13A2 with 1-km spatial resolution, 8-day LAI and FPAR from the Global LAnd Surface Satellite (GLASS) with 500-m spatial resolution (Xiao et al., 2015; Xiao et al., 2014), instantaneous 1-km all-weather LST dataset (both daytime and nighttime) from the National Tibetan Plateau Data Center (Zhang et al., 2021b), and the daily SM dataset from the
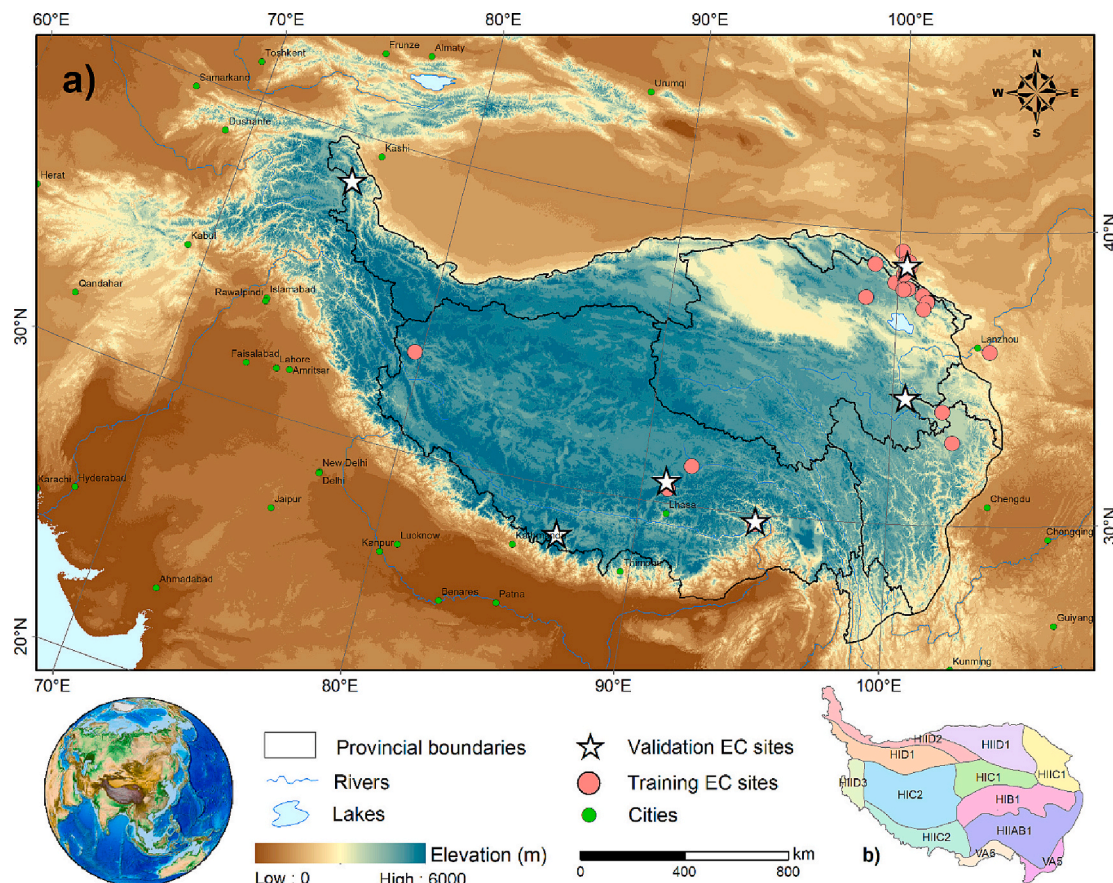

Fig. 2. Spatial distribution of a) EC flux tower sites and b) climatic subregions of the Tibetan Plateau. The IDs of climatic subregions are HI: plateau subfrigid, HII: plateau temperate, V: mid-subtropical, A: humid, B: semihumid, AB: humid/semihumid, C: semiarid, D: arid.

Table 1

Information for the 28 EC flux tower sites in the Tibetan Plateau, including the ID, site name, latitude N (Lat N), longitude E (Lon E), location, International Geosphere-Biosphere Programme land cover types (IGBP), peak leaf area index (LAI), average elevation (m), and duration of data collection.

ID	Name	Lat (N), Long (E)	IGBP	Peak LAI	Elevation (m)	Duration used in study
1	Arou	38.05, 100.46	GRA	5.17	3033	Jun 2008 – Dec 2018
2	Dashalong	38.84, 98.94	GRA	2.51	3739	Aug 2013 – Dec 2018
3	Hulugou	38.25, 99.87	GRA	3.61	3232	Sep 2011 – Dec 2016
4	Yakou	38.01, 100.24	GRA	1.65	4147	Jan 2015 – Dec 2018
5	Guantan	38.53, 100.25	ENF	3.41	2835	Jan 2008 – Dec 2011
6	Yingke	38.86, 100.41	CRO	2.85	1519	Jan 2008 – Dec 2011
7	Linze	39.33, 100.14	CRO	3.53	1399	Apr 2013 – Dec 2014
8	Daman	38.86, 100.37	CRO	3.45	1556	Jun 2012 – Dec 2018
9	Bajitan	38.92, 100.30	BAR	0.35	1562	Jun 2012 – Dec 2014
10	Huazhaizi	38.77, 100.32	BAR	0.77	1731	Jun 2012 – Dec 2018
11	Shenshawo	38.79, 100.49	BAR	0.33	1694	Jun 2012 – Dec 2014
12	Zhangye wetland	38.98, 100.45	WET	4.03	1460	Jun 2012 – Dec 2018
13	Jingyangling	37.84, 101.12	GRA	2.97	3750	Aug 2018 – Dec 2018
14	BJ	31.37, 91.90	GRA	1.88	4509	May 2008 – Dec 2016
15	QOMS	28.36, 86.95	BAR	1.09	4298	Jan 2008 – Dec 2016
16	SETORS	29.77, 94.74	GRA	4.69	3327	Jan 2008 – Dec 2016
17	NADORS	33.39, 79.70	BAR	0.46	4270	Jul 2010 – Dec 2016
18	MAWORS	38.42, 75.03	BAR	0.70	3668	Jul 2012 – Dec 2016
19	NAMORS	30.77, 90.96	GRA	1.15	4730	Jan 2008 – Dec 2016
20	Maqu	33.89, 102.14	GRA	4.89	3423	Jan 2010 – Dec 2010
21	LinZhi	29.65, 94.72	ENF	6.49	4330	Jan 2008 – Mar 2010
22	Yuzhong	35.95, 104.13	GRA	1.67	1841	Jul 2008 – Oct 2009
23	Dangxiong	30.50, 91.07	GRA	1.45	4297	Jan 2004 – Dec 2010
24	Haibei Shrubland	37.61, 101.33	GRA	5.08	3190	Jan 2003 – Dec 2010
25	Haibei Alpine Tibet	37.37, 101.18	WET	2.60	3205	Jan 2003 – Dec 2009
26	TRHR	34.35, 100.55	GRA	3.68	3960	Jan 2016 – Dec 2016
27	HongYuan	32.85, 102.59	GRA	3.78	3500	May 2015 – Dec 2017
28	GaoHanCaoDian	37.70, 98.59	GRA	3.02	3718	Sep 2018 – Dec 2018

European Space Agency Climate Change Initiative (ESA CCI) program with 0.25-degree spatial resolution. We used a wavelet transform (WT) method provided by Hu et al. (2022) combined with finer resolution satellite products (i.e., LAI, FPAR, LST) to downscale the SM dataset to 1-km spatial resolution.

The meteorological forcing data were derived from the China Meteorological Forcing Dataset (CMFD) with daily, 0.1-degree spatial temporal resolution (He et al., 2020). The CMFD continuously covers the period from 1979 to 2018. The variables used in this study include the

Ta (K), pressure (Pa), specific humidity (kg kg^{-1}), WS (m s^{-1}) and Rs (W m^{-2}). The 16-day NDVI, 8-day LAI and FPAR were linearly interpolated daily, and both the satellite and meteorological datasets were spatially resampled to a 1-km spatial resolution to maintain identical raster geometry.

In addition to ET-PM and ET-PT, the hybrid models were compared with commonly known ET products. These products included a modified satellite-based Priestley-Taylor ET product (MS-PT) which used Apparent Thermal Inertia (ATI) to quantify soil water deficit (Yao et al., 2013); a semiempirical Penman-Monteith-based ET product (ET-SEMI) which developed by 64 global flux tower sites (Wang et al., 2010a; Wang et al., 2010b); a soil moisture constrained ET product (EF-HF) which based on an empirical formulation calibrated by flux tower sites in the Northwest China (Yao et al., 2019), and a Global Land Evaporation Amsterdam Model (GLEAM, v3.3 a) ET product (Martens et al., 2017).

3.3. Model validation and evaluation metrics

(1) Cross validation and independent validation

We adopted tenfold cross-validation to randomly assess the overall model performance. In addition, we carried out independent validation by leaving out 20% (6 sites) of the available stations to evaluate the spatial simulation performance of the hybrid models. For the specific investigation of prediction performance for unknown locations and site-scarce regions, we trained the hybrid models with varying percentages of training sites (Table 2) and tested the estimation results with the same validation sites. The selection of varying training sites and independent validation stations must meet the requirements of a uniform spatial distribution and sufficient representativeness of plant functional types.

(2) Validation over extreme conditions

To verify the robustness of hybrid models, we compared the generalization performance of ET models in multiple extreme cases. For six independent validation sites, we used the 0th - 3rd percentiles and 97th - 100th percentiles extreme samples for five variables (Rs, NDVI, Ta, SM and RH) to assess the response of each ET model. For instance, the 0th - 3rd percentiles of the SM datasets were used to investigate droughts; in contrast, the 97th - 100th percentiles of SM were used to investigate irrigation or precipitation.

(3) Validation metrics

To evaluate the hybrid models against ground measurements, we selected R^2 to assess the correlations and adopted the root-mean-square-error (RMSE) and bias to quantify the closeness and differences. Moreover, we used the Kling-Gupta efficiency (KGE) (Gupta et al., 2009) to evaluate the overall performance because it incorporates the correlation

Table 2

Summary of EC flux tower sites used in the varying percentages of total available sites in the training of pure ML, ML-Gs and ML-Es.

Percentage of training sites	Training sites	Validation sites
80% (22 sites)	Arou, Dashalong, Hulugou, Yakou, Guantan, Yingke, Linze, Bajitan, Huazhaizi, Shenshawo, Zhangye, Jingyangling, BJ, NADORS, Maqu, LinZhi, Yuzhong, Dangxiong, Haibei Shrubland, Haibei Alpine Tibet, HongYuan, GaoHanCaoDian	
60% (16 sites)	Arou, Dashalong, Yakou, Guantan, Yingke, Bajitan, Huazhaizi, Shenshawo, Zhangye, BJ, NADORS, Maqu, LinZhi, Haibei Shrubland, HongYuan, GaoHanCaoDian	Daman, QOMS, SETORS, MAWORS, NAMORS and TRHR (20%, 6 sites)
40% (11 sites)	Arou, Dashalong, Guantan, Yingke, Zhangye, BJ, NADORS, Maqu, Haibei Shrubland, HongYuan, GaoHanCaoDian	
20% (6 sites)	Arou, Zhangye, BJ, NADORS, HongYuan, GaoHanCaoDian	

(r), relative variability ratio (α) and mean values ratio (β) to better understand the mismatches between estimation and observation. The KGE can be written as Eqs. (20–21), where σ_p and μ_p are the standard deviation and mean value of predictions, respectively, and σ_o and μ_o are those of observations. The model performance is better when KGE is closer to 1.

$$KGE = 1 - \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2} \qquad (20)$$

$$\alpha = \frac{\sigma_p}{\sigma_o} \qquad (21)$$

$$\beta = \frac{\mu_p}{\mu_o} \qquad (22)$$

(4) Machine learning model interpretability analysis

We used the Shapley value to analyze the interpretability of ML coupled in two hybrid models. The Shapley value was calculated by SHAP (SHapley Additive exPlanations), which was proposed to calculate explanations of model predictions by classic equations from cooperative game theory (Cubitt, 1991; Lipovetsky and Conklin, 2001). SHAP theory regards model predictions as the sum of marginal contributions of each input feature (Lundberg et al., 2018).

For a sample x_i, the feature j of this sample is x_{ij}. The predicted value of the model for this sample is y_i, and the baseline of the model (usually the mean of the target variables of all samples) is y_{base}. It can be written as:

$$y_i = y_{base} + f(x_{i1}) + f(x_{i2}) + \ldots + f(x_{ij}) \qquad (23)$$

where $f(x_{ij})$ is the SHAP value of x_{ij}. The computing of the Shapley value can be implemented by Python.

4. Results

4.1. Evaluation against ground measurements

4.1.1. Model development using EC observations

To evaluate the model simulation performance, we conducted tenfold cross-validation at all 28 EC sites (Fig. 3 a1-a4). In addition, six

independent validation sites were used to investigate the out-of-sample generalization performance for ML-Gs and ML-Es (Fig. 3 b1-b4). The inferred simulation performance of gs and ETs were also plotted. The cross-validation results show that the inferred gs performs well, with a KGE of 0.80, R^2 of 0.68 ($p < 0.01$), bias of 0.01 m s^{-1} and RMSE of 0.04 m s^{-1}. The KGE of the corresponding ET estimated by ML-Gs is as high as 0.91. Compared with gs, the inferred ETs fit the ground-based data better, with a higher KGE of 0.83 and R^2 of 0.77 (p < 0.01). The corresponding ET from ML-Es outperformed that of ML-Gs, with a KGE of 0.92 and R^2 of 0.89 (p < 0.01). The RMSEs of ML-Es and ML-Gs are 0.50 and 0.54 mm/day, respectively, and the biases of these two hybrid models are comparable.

The independent validation results shown in Fig. 3 b) indicate that the ET estimates from the two hybrid models have similar generalization skills and perform well in the out-of-sample simulation of daily ET. The ET estimation from ML-Es performs slightly better than that of ML-Gs, with a higher KGE of 0.87 and R^2 of 0.77 (p < 0.01) and a lower bias of 0.01 mm/day. The KGE and R^2 of ML-Gs are 0.86 and 0.76 (p < 0.01), respectively, and the bias and RMSE are 0.03 and 0.68 mm/day, respectively. The KGE of the independent validation for inferred gs is 0.62, which is lower than that of ETs (0.70). These two components of the ET process yield comparable R^2 values of 0.55 (gs) and 0.52 (ETs). The results from the cross-validation and independent validation indicate that these two hybrid models are stable and have the potential to simulate daily ET over the TP.

To further investigate the capacity to simulate ET, the seasonal variations of ground measurements and five ET model estimates at independent validation sites are plotted in Fig. 4. Six EC tower sites with more than half a year of available ET observations in 2011/2016 were used to compare seasonal ET estimates between the hybrid and pure physical/ML methods. These site-based evaluation results indicate that although current ET estimates from hybrid models may contain biases, they generally agree well and show great consistency with the observations. At six validation sites, by comparison with ET-PM and ET-PT, ML-Gs and ML-Es enhance the ET estimates in the dominant plant functional types of the TP (grasslands, barren lands and croplands). The pure ML shows similar fitting performance with ML-Gs and ML-Es.

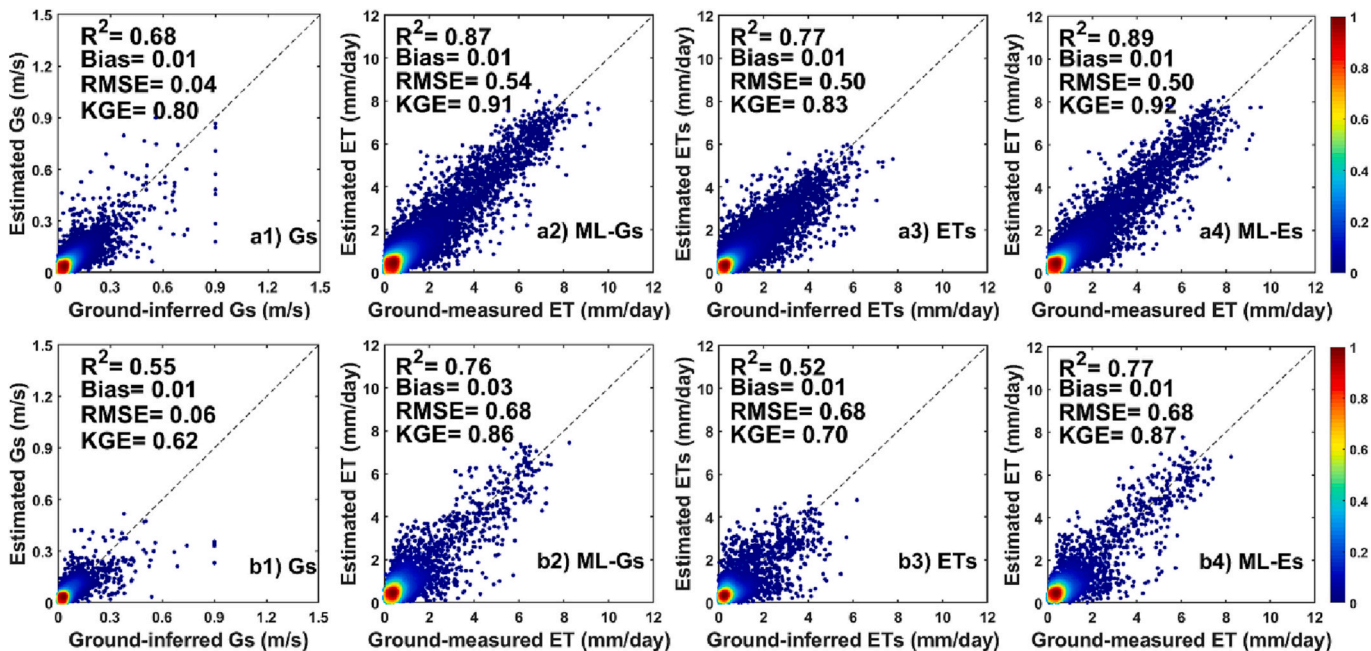We also compared the two hybrid models with commonly known ET

Fig. 3. Performance of two physics-guided hybrid models in predicting 1) surface conductance (Gs) and 2) corresponding ET from ML-Gs, 3) soil evaporation (ETs) and 4) corresponding ET from ML-Es. The letters a) and b) represent the daily hybrid model retrieval in the cross-validation and independent validation, respectively.
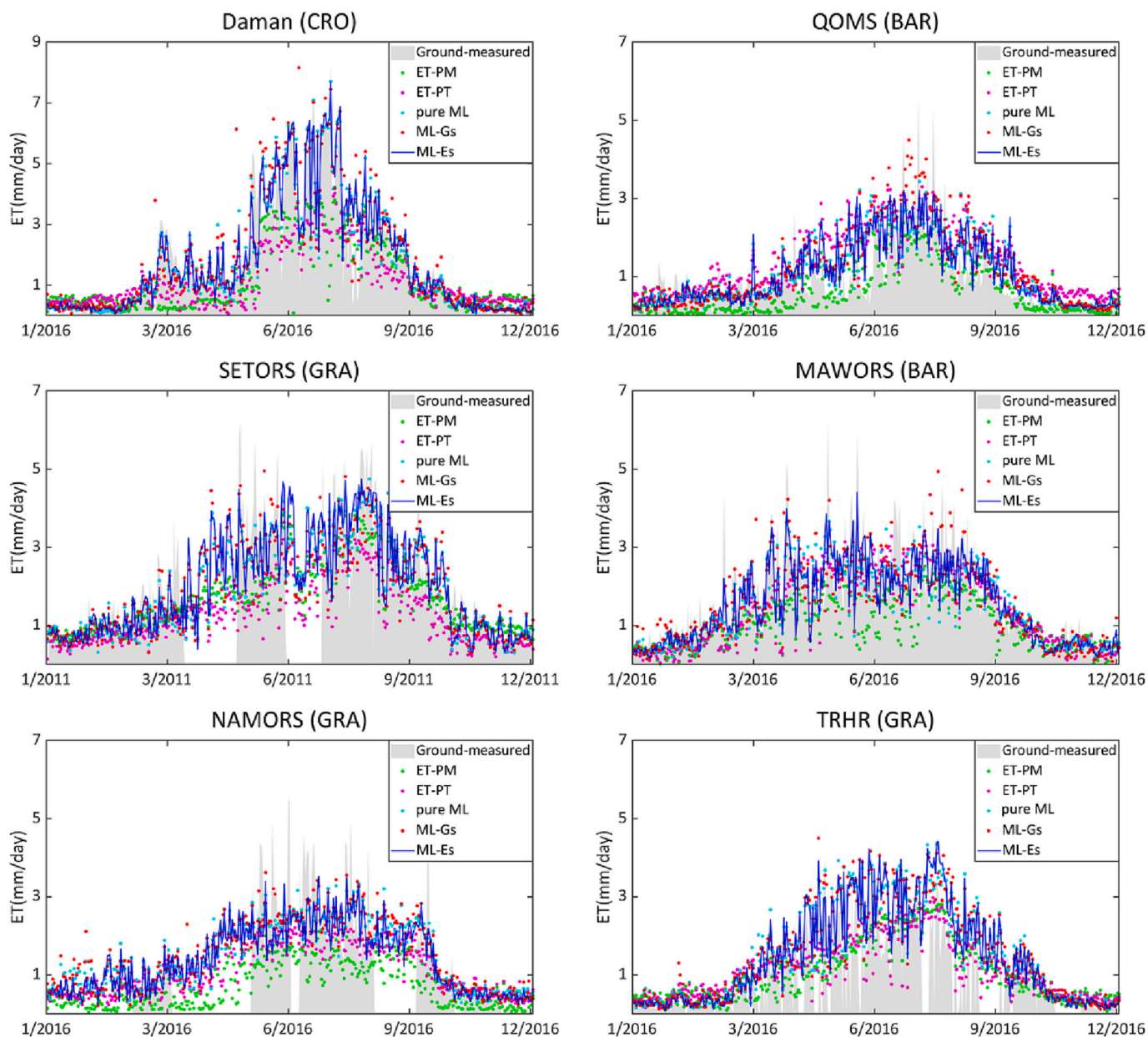
Fig. 4. Time series of the ET ground measurements (gray background), ET estimates with the hybrid models (blue lines for ML-Es, red points for ML-Gs), ET estimates from the pure machine learning (light blue points, pure ML) and ET estimates of the physical models (purple points for ET-PT, green points for ET-PM) at six validation EC sites, including Daman (CRO), QOMS (BAR), SETORS (GRA), MAWORS (BAR), NAMORS (GRA) and TRHR (GRA).

products (ET-PM, ET-PT, MS-PT, ET-SEMI, ET-HF, GLEAM). The results show that ML-Gs and ML-Es perform better at 28 flux tower sites. As Fig. 5 shows, the overall KGEs of ML-Gs and ML-Es are 0.91 and 0.92, respectively, whereas that of six compared models are ET-PM (0.35), ET-PT (0.36), MS-PT (0.40), ET-SEMI (0.25), ET-HF (0.80) and GLEAM (0.29). Among six commonly known products, MS-PT performs better at BAR sites with the KGE of 0.50, ET-SEMI performs better at ENF sites with the KGE of 0.55, ET-HF performs better at CRO, GRA and WET sites. The two hybrid models yield higher accuracies than these six compared ET products throughout different plant functional types (Fig. 5).

4.1.2. Temporal assessment of the hybrid models

Fig. 6 summarizes the performance comparison between the hybrid models, pure ML and physical models at three temporal scales. ML-Es outperforms ML-Gs at daily, monthly and annual scales. In addition, pure ML and hybrid models perform better than physical models with

the same trend. For daily ET estimates, the median KGE' values of ML-Es, ML-Gs and pure ML are between 0.76 and 0.78. The ET-PT demonstrates better performance in two physical models with a median KGE of 0.33, whereas ET-PM yields the lowest median KGE value of 0.22. The RMSE of the three ML-based models varies from 0.19 to 0.78 mm/day, and that of the two physical models ranges between 0.42 and 1.73 mm/day. For monthly simulation, pure ML yields the highest median KGE (0.81), followed by ML-Es (0.80) and ML-Gs (0.78). ET-PT and ET-PM show lower accuracies, with median KGE values of 0.31 and 0.21, respectively. The median R^2 for ML-Es, ML-Gs and pure ML vary from 0.88 to 0.90 ($p < 0.01$), whereas those of ET-PT and ET-PM are 0.45 and 0.35 (p < 0.01), respectively. For the annual scale, the highest median KGE is 0.79 from ML-Gs, followed ML-Es (0.78), pure ML (0.76), and ET-PT (0.57) and ET-PM (0.43).

All five models show increased performance from the daily scale to the annual scale. In comparison with three ML-based methods, the two
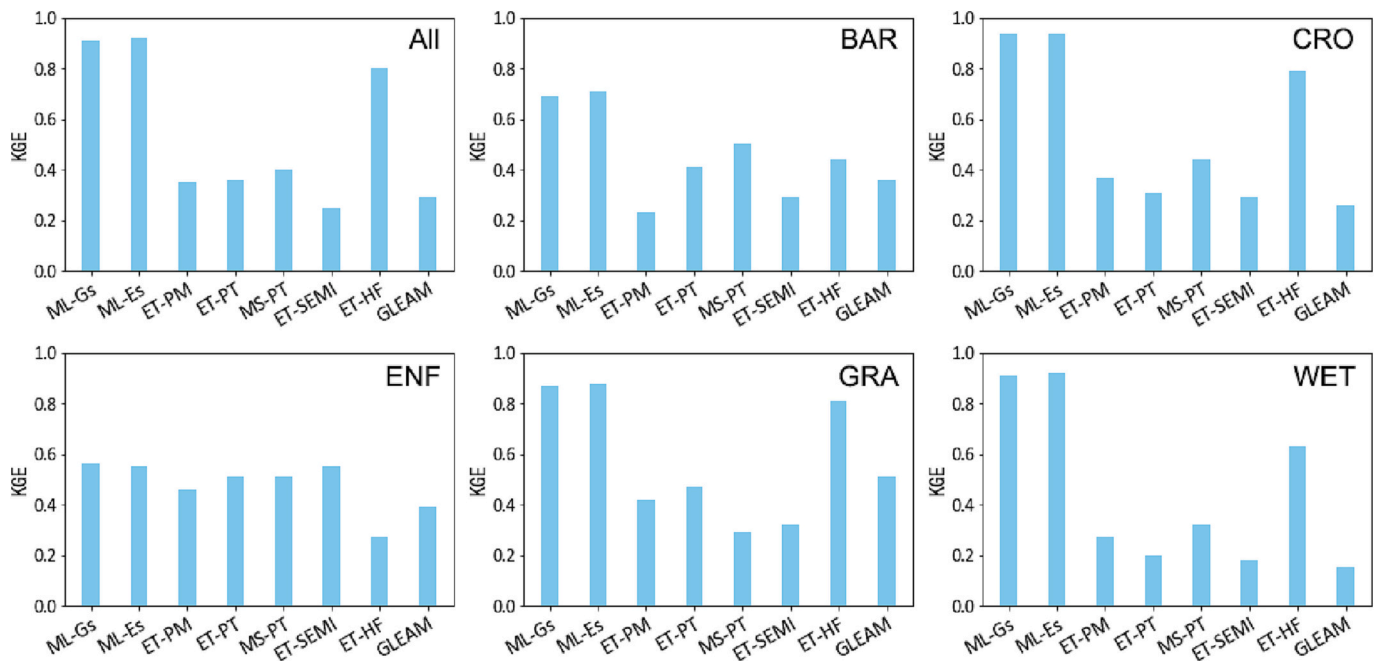
Fig. 5. Comparison of hybrid ET models and commonly known ET products for different plant functional types. All: all plant functional types, BAR: barren lands, CRO: croplands, ENF: evergreen needleleaf forests, GRA: grasslands, WET: wetlands.

physical models present lower performance and greater dispersion in terms of the four evaluation metrics. The results demonstrate that the ML approach can improve ET simulation accuracy at multiple temporal scales. In addition, both data-oriented and physics-guided ML models can generate similar and competitive fitting performance for cross-validation.

4.1.3. Spatial and anomaly assessment of the hybrid models

The ET estimates for site variability and annual anomalies were evaluated using ground measurements to assess the spatial and anomaly simulation performance of the five ET models. As Fig. 7 a) demonstrates, for the two physical ET models, ET-PM performs better than ET-PT for the site variability. The KGE of ET-PM for the ET site variability is 0.38, the R^2 is 0.76 ($p < 0.01$), and the RMSE is 0.49 mm/day. ET-PM shows underestimation at the site scale, with a bias of -0.34 mm/day. In comparison, ET-PT yields poor accuracies, with a lower KGE of 0.28 and R^2 of 0.69 ($p < 0.01$). The bias of -0.24 mm/day indicates that ET-PT presents the same underestimation trend as ET-PM. As Fig. 7 a) shows, ET-PM and ET-PT show significant underestimation at the Zhangye (WET) and Yingke (CRO) sites. For two GRA sites (SETORS and BJ), these two physical models also underestimated the site ET variability. In comparison, the three ML models better fit the ground-based data for the site variability, with a significantly higher KGE of 0.94 and R^2 values from 0.95 to 0.96 ($p < 0.01$). The bias and RMSE are also quite low. Although they show a similar KGE, ML-Es achieves higher R^2 and lower RMSE and bias, which makes it perform slightly better than ML-Gs.

For annual ET site anomalies (Fig. 7 b), ET-PT performs better than ET-PM, with a higher KGE of 0.46 and a lower RMSE of 32.17 mm/year. For the Zhangye (WET) and MAWORS (BAR) sites, the two physical models present slight underestimations at the site anomalies. In addition, at the HongYuan (GRA) and Dashalong (GRA) sites, the physical models do not capture the sharp decreases in ET interannual anomalies, and show slight overestimation. In comparison, ML-Gs, ML-Es and pure ML all perform better than ET-PM and ET-PT. The KGE of the three ML-based models varies from 0.90 to 0.91, and the R^2 values range from 0.91 to 0.93 (p < 0.01). ML-Es yields the lowest RMSE of 13.60 mm/year, followed by pure ML (14.65 mm/year) and ML-Gs (16.00 mm/year).

4.2. Impact of EC flux tower density and sample representativeness

4.2.1. Generalization performance over training sites with varying percentages

To understand the influence of the EC flux tower density on model development, the same sample data from a varying number of EC sites were used to train three ML models. The accuracies at six validation sites were compared with physical models to analyze the simulation performance for unknown locations. Fig. 8 shows that the physical models achieve similar accuracies, with KGE values of 0.40 (ET-PM) and 0.38 (ET-PT). When the available training sites decrease from 80%, 60%, 40% and 20%, the three ML models outperform the physical models with the same trend. Among them, ML-Es yields the most robust generalization performance.

Fig. 8 illustrates that the KGE values of ML-Es range from 0.87 to 0.65 with varying training sites, which outperforms those of ML-Gs and pure ML. When training sites are reduced to 60%, the KGEs of the hybrid models (ML-Es and ML-Gs) remain 0.87, while that of the pure ML model significantly drops to 0.84. When 40% of the sites are used for training, although the accuracies of these three models drop with the same trend, ML-Es still shows the highest KGE of 0.76. When only 20% of the sites are used for model training, the KGE of ML-Es is 0.65, whereas those of ML-Gs and pure ML fall to 0.64 and 0.62, respectively. However, even under this particular circumstance, the performance of these three models is still better than those of the physical models (KGE of 0.40 for ET-PM and KGE of 0.38 for ET-PT). Taken together, the hybrid models (ML-Es and ML-Gs) perform better than the pure ML model with decreasing site density. In addition, the more robust and clearer the physical mechanism coupling is, the better the performance of the hybrid model for data-sparse regions.

4.2.2. Generalization performance over extreme samples

To verify the robustness and investigate the generalization performance of extreme events for hybrid models, we compared the independent validation performance of ML-Es, ML-Gs and pure ML in multiple extreme cases (Fig. 9). A key result is that the extrapolation performance of the hybrid models (ML-Es and ML-Gs) is systematically better than that of the pure ML model in most extreme cases (Fig. 9).
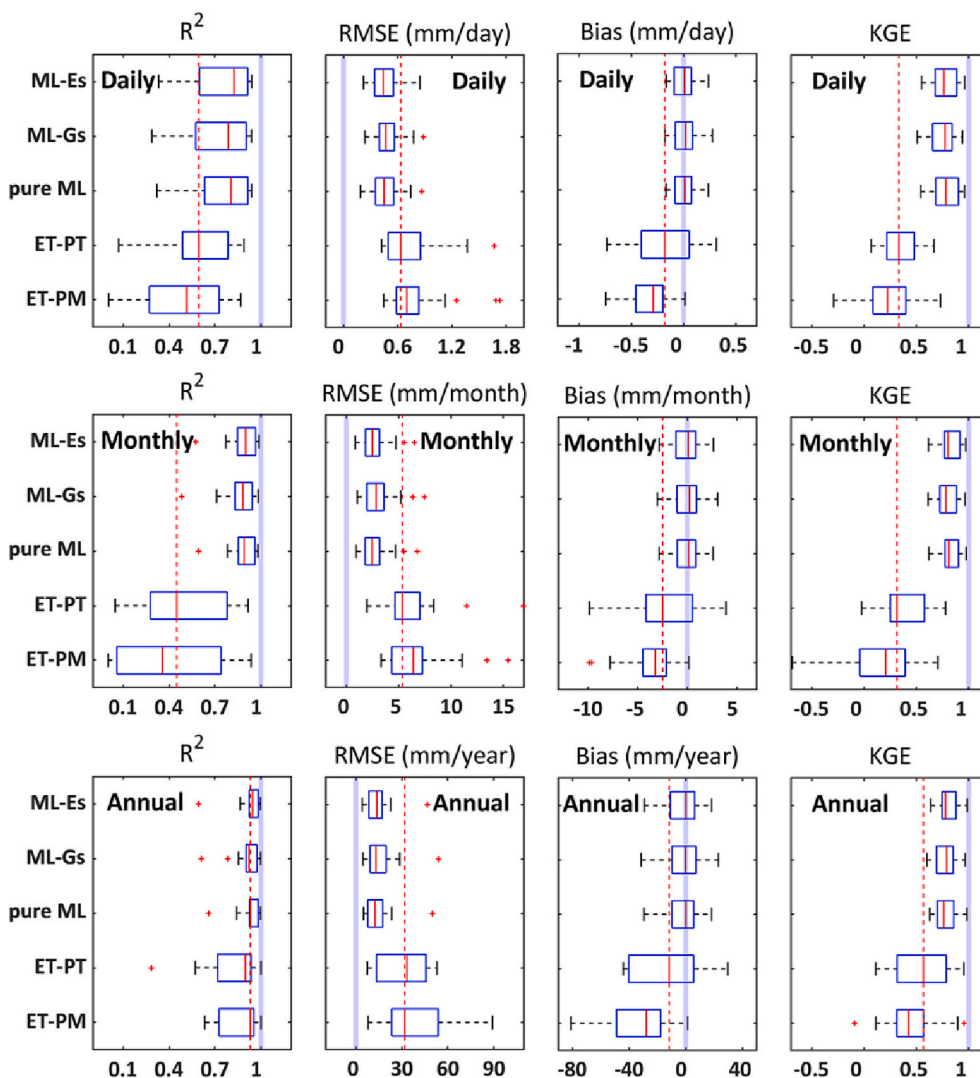
Fig. 6. Performance comparison of two hybrid models (ML-Es, ML-Gs), a pure machine learning (pure ML) model and two physical models (ET-PT, ET-PM) against ground measurements. Metrics from left to right are R^2, RMSE, bias, and KGE. Temporal scales from top to bottom are daily, monthly and annual. The red dotted lines indicate the median values of the best-performing physical model. The blue lines indicate the optimal values for all metrics.

For the 0th - 3rd percentiles of the SM and NDVI in particular, that is, for droughts and vegetation-sparse cases, the hybrid models perform much better than the pure ML model. Fig. 9 demonstrates that for drought cases, the KGEs of ML-Es and ML-Gs are 0.16 and 0.17, respectively, but that of pure ML sharply drops to 0.04. For vegetation-sparse cases, the KGE of pure ML is as low as 0.03, while those of ML-Es and ML-Gs are 0.15 and 0.14, respectively. When the RH is high, the KGE of pure ML is 0.35; ML-Es increases it by 43% and ML-Gs increases it by 34% compared with pure ML. Fig. 9 also illustrates that under most extreme circumstances, the physics-guided hybrid models yield stronger extrapolation performance than pure ML, emphasizing the capacity of hybrid models to better predict the impacts of extreme weather events on the ET process.

4.3. Mapping of the Tibetan Plateau ET from the hybrid models

Fig. 10 presents the spatial distribution of the multiyear (2003–2018) mean ET for the five ET models over the TP. With the changes in elevation and climatic region, all ET estimates show consistent spatial patterns over the TP. In general, the estimated ET decreases from the southeast to the northwest of the TP, with a maximum value above 1000 mm/year in humid mid-subtropical southeastern regions

and a minimum value of <200 mm/year on the arid subfrigid plateau over the northwestern edge of the TP.

The spatial variations of all ET model estimates are consistent with the variations of climatic subregions (Fig. 10). For the whole TP, significantly higher annual ET estimates emerge in the Yunnan Plateau (VA5) and south edge of the east Himalaya (VA6), which are humid/mid-subtropical regions. The median annual ET of the five models for VA5 and VA6 are 728.10 and 685.26 mm/year, respectively. For eastern Tibet (HIIAB1), which is located in a humid/semihumid plateau temperate region, the median annual ET from the five models is 561.75 mm/year. The regions that generate the lowest ET in the TP appear in the Kunlun Mountains (HID1) and Ali Mountains (HIIID3). They lie in arid plateau regions, and the annual ET from the five ET simulation models are <360 mm/year.

Despite great consistency in spatial patterns, the five ET estimates show regional discrepancies and uncertainties. The domain-averaged multiyear mean ET of the five models over the TP ranges from 320.03 to 426.87 mm/year. Compared with the three ML-based models, ET-PM and ET-PT provide lower annual ET estimates in the Ali Mountains (HIIID3), Qiangtang Plateau Lake basin (HIC2) and Qaidam Basin (HIID1). Affected by atmospheric circulation and vegetation coverage, the ET in the eastern TP is higher. The pure ML, ML-Gs and ML-Es better
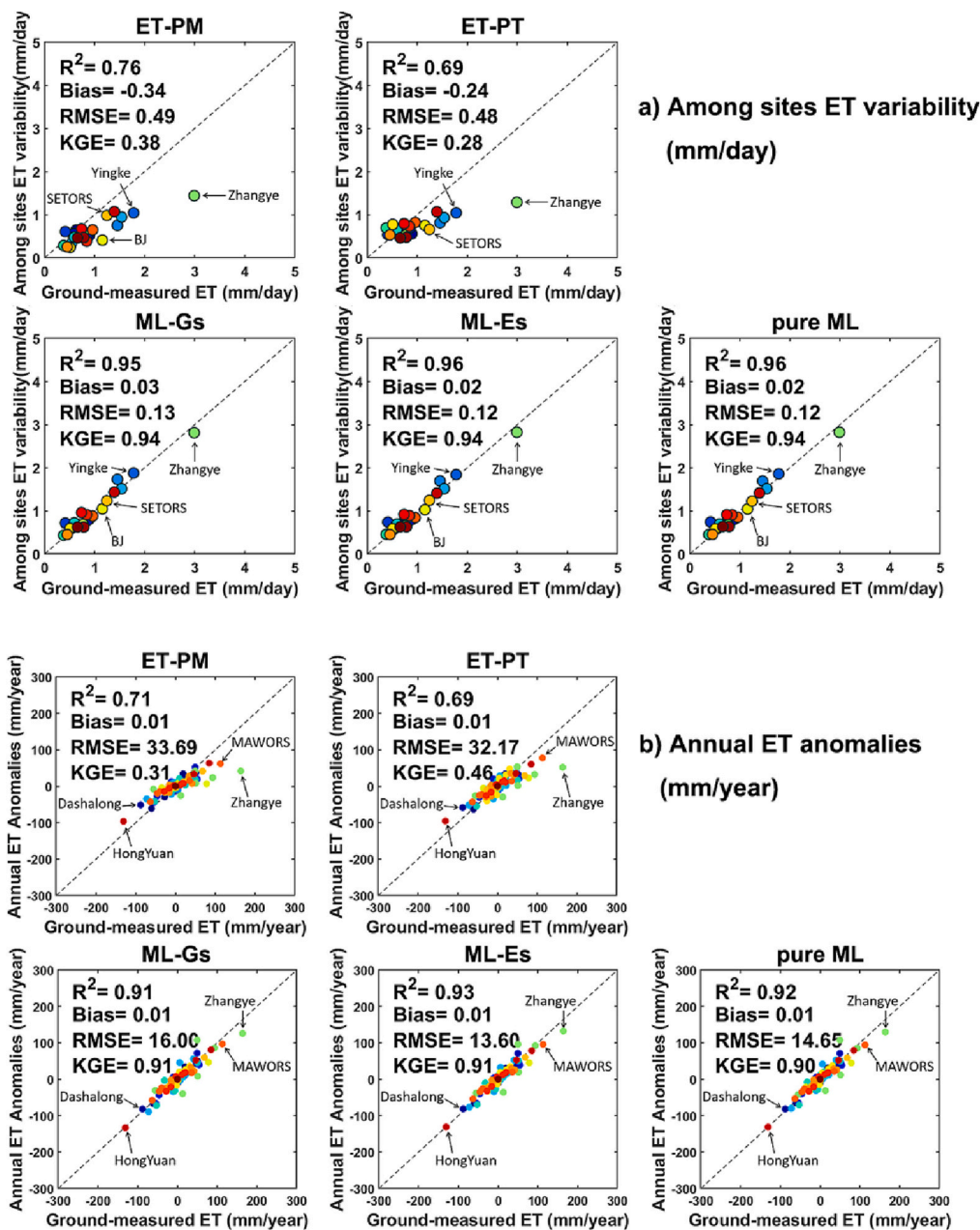
Fig. 7. The estimated ET of different models (vertical axis) versus the ground-based measurements (horizontal axis) for a) among-site ET variability and b) annual ET anomalies. For each ET models, the same colour represents the same site.

capture this characteristic than the two physical models and present more reliable estimations in eastern Tibet (HIIAB1), the Golog-Naqu hilly plateau (HIB1) and the Qilian Mountains (HIIC1). Although pure ML, ML-Gs and ML-Es generate the closest ET spatial distribution in the whole TP, they show discrepancies regionally. In the central Qaidam Basin (HIID1), which is an arid plateau temperate, the ET estimates of ML-Gs are nearly 100 mm/year higher than those of pure ML and ML-Es. This may be attributed to the discrepancies between model structures of ML-Gs and other models. ML-Gs estimates ET based on the energy control part (λE_E) and the atmospheric control part (λE_A) (Eq. 1), the aerodynamic conductance used to estimate λE_A contains significant uncertainties. Strongly influenced by topography and Asian monsoon rainfall, the ET at the southern edge of the TP is strong.

5. Discussion

5.1. Performance of physics-guided hybrid ET models

5.1.1. Capacity to simulate ET

By the coupling of physical constraints with ML method, hybrid models preserve the physical understanding and produces more accurate ET estimation on the TP. ET estimates of hybrid models generally agree well with the seasonal variations of ground measurements. As shown in Fig. 4, for Daman (CRO), ET-PM and ET-PT underestimate ET during the growing season, while ML-Gs and ML-Es significantly fit observations better. This mainly because ET at CRO site in the growing season reach the peak due to intensive irrigation. SM and LST can capture the variation of surface moisture and temperature caused by irrigation. ML-Gs and ML-Es use SM and LST as input variables to model ET, which can better fit the soil evaporation for intensively irrigated
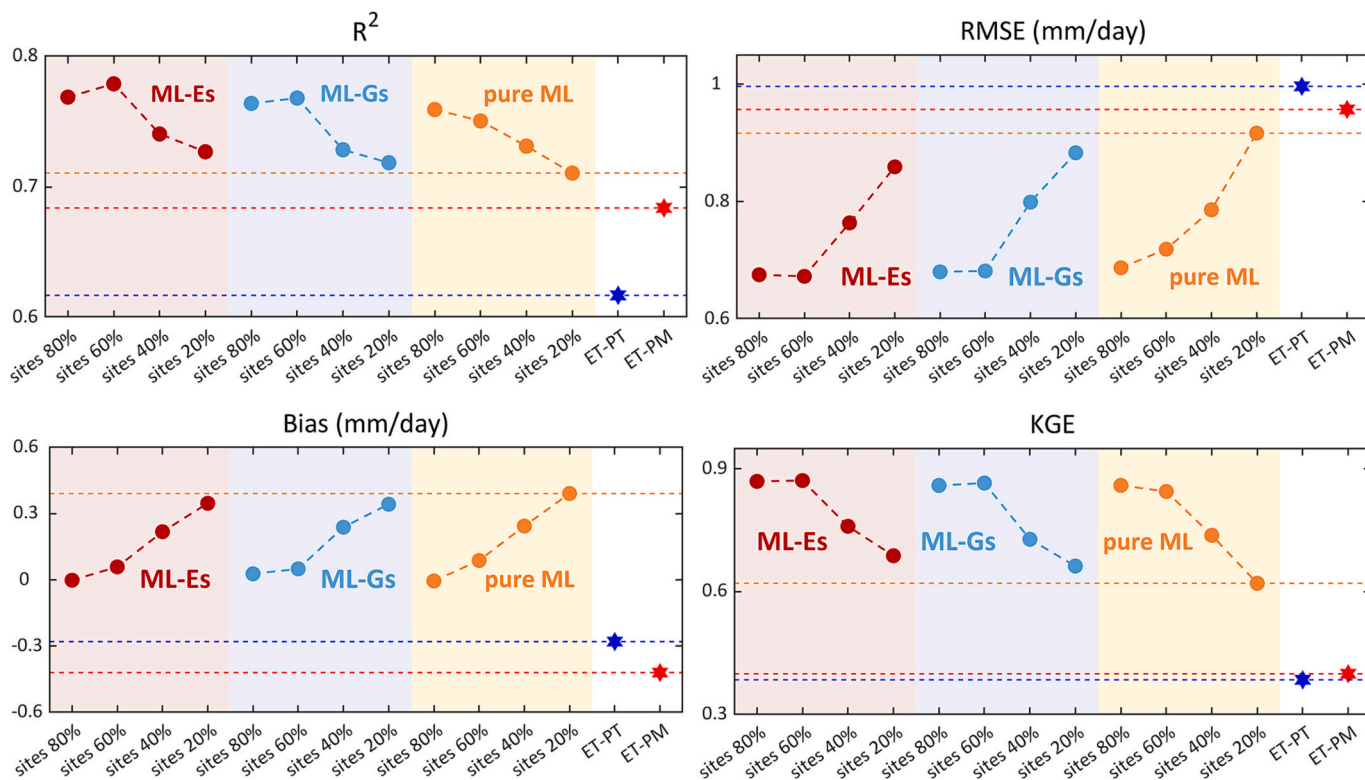
Fig. 8. Performance of two hybrid models (ML-Es, ML-Gs) and a pure machine learning model (pure ML) at 6 validation sites using varying percentages of EC flux tower sites in the model training dataset. The performance of two physical models (ET-PT and ET-PM) are plotted for comparison.
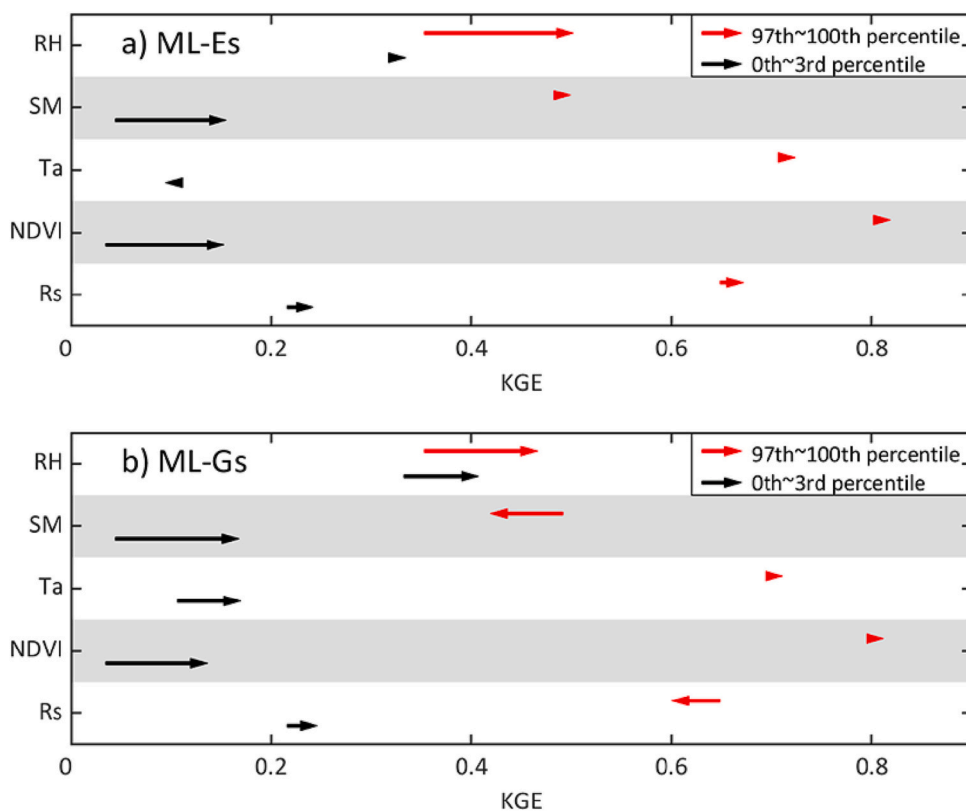

Fig. 9. Extrapolation performance of a) ML-Es and b) ML-Gs hybrid models versus a pure machine learning model (pure ML) over an extreme dataset of validation sites. The arrows show the KGE value change from the data-oriented pure ML to the physics-guided hybrid estimations. The red arrows indicate the results in the 97th - 100th percentiles of the validation dataset, and black arrows indicate the 0th - 3rd percentiles. The start of the arrows represents the KGEs of pure ML, and the end of the arrows represents the KGEs of hybrid models (a for ML-Es and b for ML-Gs). The length of the arrows represents the KGE differences between the pure machine learning and hybrid models.
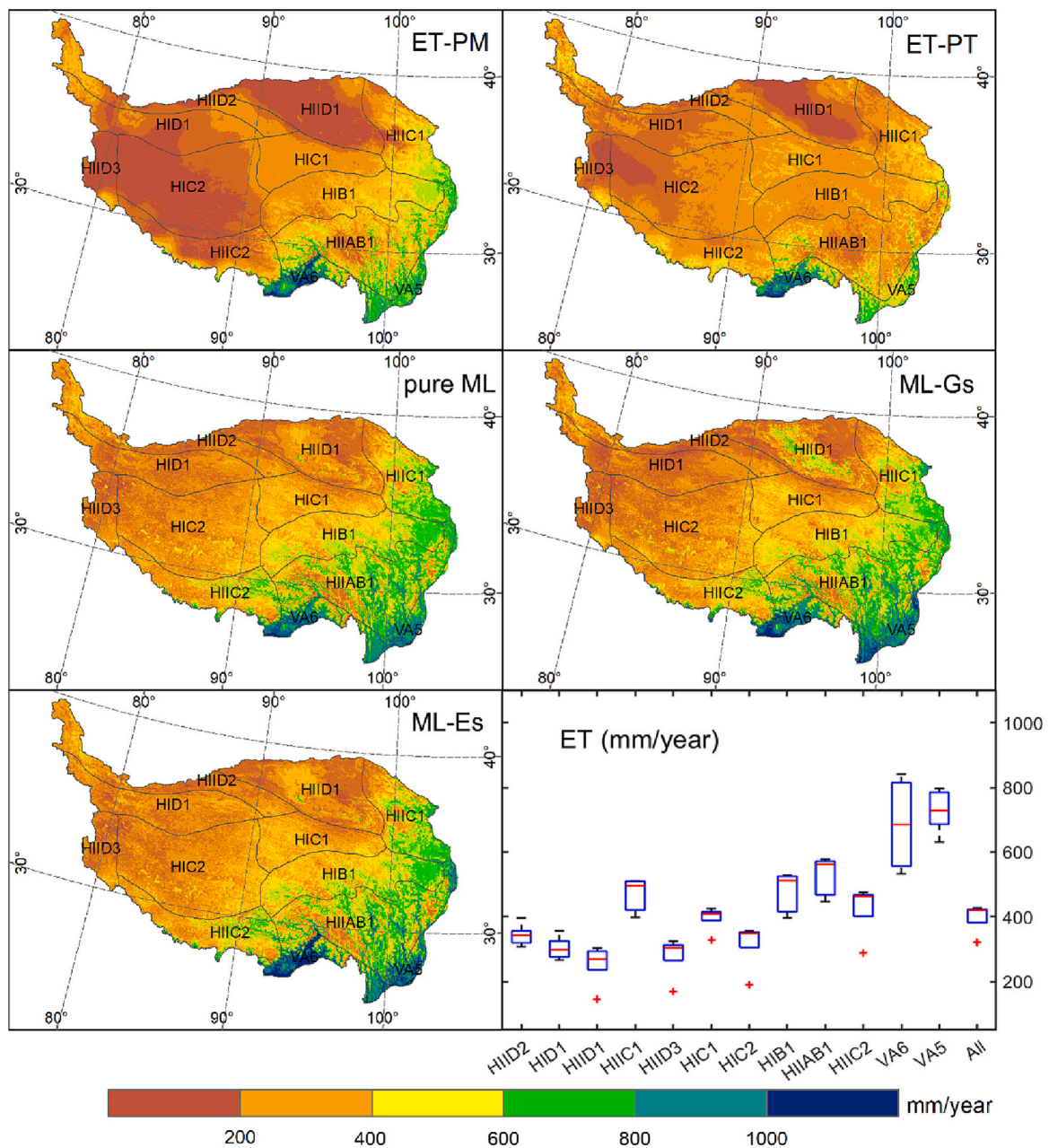
Fig. 10. Mapping of multiyear (2003–2018) mean ET and statistical boxplots of climatic regions of five ET models for the Tibetan Plateau. The five models include two physical models (ET-PM and ET-PT), a pure machine learning model (pure ML), and two hybrid models (ML-Gs and ML-Es). The IDs of climatic subregions are HI: plateau subfrigid, HII: plateau temperate, V: mid-subtropical, A: humid, B: semihumid, AB: humid/semihumid, C: semiarid, D: arid.

agricultural fields (Marshall et al., 2020). For grasslands, ET-PT slightly underestimates the high-grass sites (SETORS and TRHR) from May to September 2011, while ET-PM shows a significant underestimation trend at the short-grass site (NAMORS) in 2016. ML-Gs and ML-Es better fit the observations than two physical models, which reflects the highlights of ML-based method—the superior extraction and simulation capacity for spatiotemporal features (Reichstein et al., 2019). For the highaltitude gravel site (QOMS) and Gobi site (MAWORS), due to the complex terrain and heterogeneous landscape, physical model estimates show large discrepancies and uncertainties compared with the in situ measurements, which has been reported in similar studies (Han et al., 2021). At six validation sites, pure ML shows similar or even competitive fitting performance with ML-Gs and ML-Es. This indicates that a pure data-oriented ML model has automatic feature extraction capability and strong nonlinear simulation performance if sufficient samples are

available.

Taken together, traditional physical models tend to underperform in dry regions where moisture status is not well represented by surrounding areas (Li et al., 2009; Marshall et al., 2020). In addition, the ET underestimates by ET-PM and ET-PT are mostly in the season when the canopy is withered or in regions where the vegetation coverage is sparse, while ML-Gs and ML-Es perform better under all circumstances. The validations indicate that the hybrid models improve the simulation capacity of physical models and can be applied to a wider range of ecosystems over the TP. For regions outside the TP where there are available flux tower sites, the hyper-parameters of LightGBM used in our hybrid ET models could be re-tuned. The well trained hybrid models can be also applied to the new regions.

5.1.2. Generalization performance

Notwithstanding the improvement from the use of ML-Gs, ML-Es and pure ML in the ET simulation, pitfalls such as the risk of out-of-sample extrapolation should be avoided. Better generalization performance benefits from sufficient and representative training samples (Ball et al., 2017b). However, the coverage of in situ measurements is limited, and many regions of the global continent do not have EC sites or available observations. It is more important to enhance the prediction performance for out-of-sample locations using limited in situ measurements. Fig. 8 illustrates that ML-based models (ML-Gs, ML-Es and pure ML) outperform physical models (ET-PM and ET-PT), even when only 20% of EC sites (6 sites) were used. In addition, for most extreme cases, the hybrid models can better predict the impact of weather events on the ET

process than pure ML (Fig. 9). Taken together, we face two questions: (1) Can the pure ML be an alternative for physical models? (2) By coupling with physical constraints, can the extrapolation performance of hybrid models be practically improved?

(1) To understand the response to these two questions for TP, we conducted further investigation. We selected certain data from independent validation samples – the case of $0 < NDVI < 0.1$ and the case of $0 < VPD < 0.3$ kPa – to compare the out-of-sample model fitting performance. Fig. 11 a) shows that for the extreme case of $0 < NDVI < 0.1$, the KGE of pure ML is 0.23, which is significantly lower than those of ET-PM (0.32) and ET-PT (0.51). For the case of $0 < VPD < 0.3$ kPa (Fig. 11 b), the KGE of pure ML is 0.27; in contrast, ET-PM and ET-PT outperform pure ML with higher KGE values of 0.29 and 0.45 and lower bias and
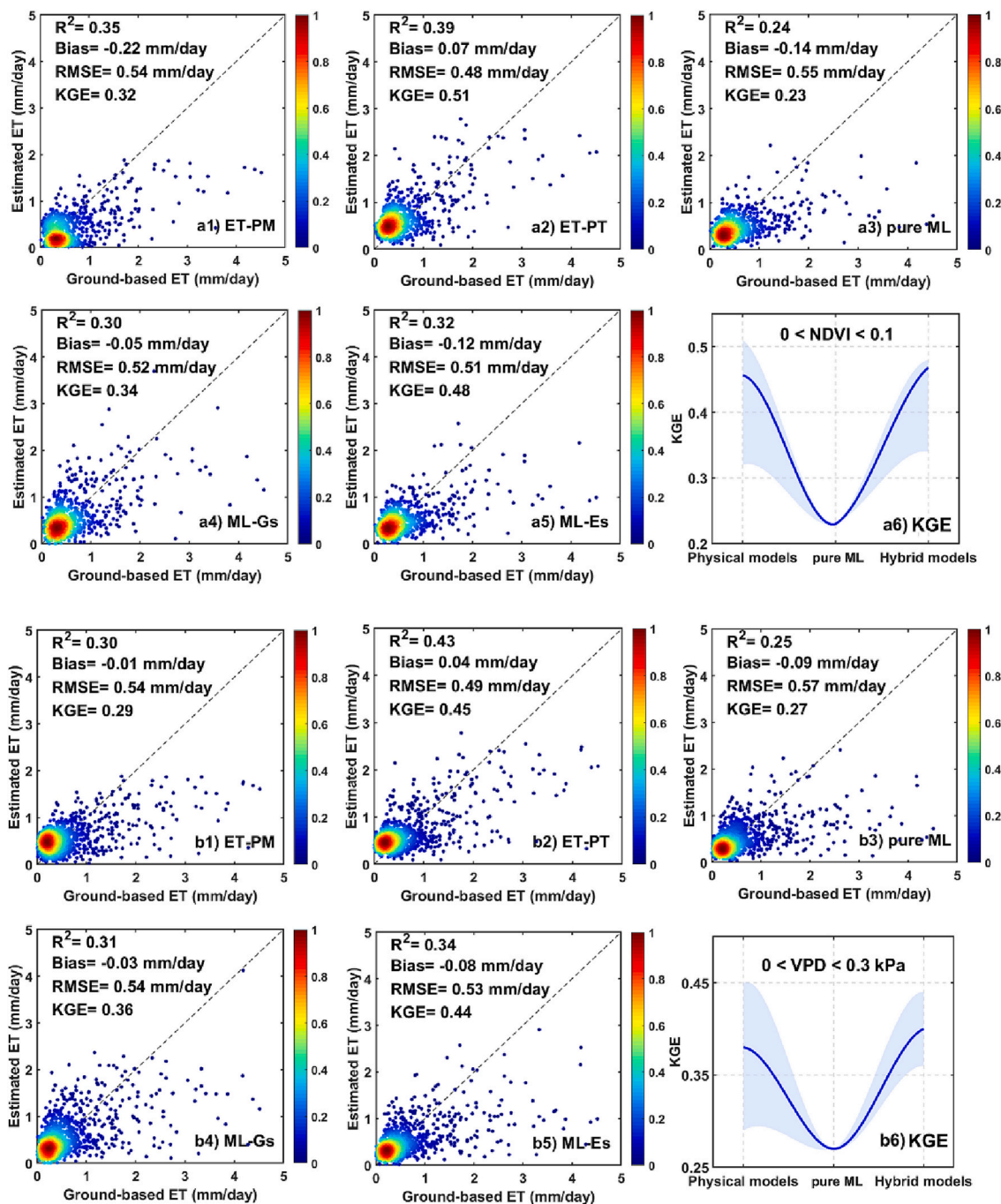

Fig. 11. Scatterplots of five models (1 ET-PM, 2 ET-PT, 3 pure ML, 4 ML-Gs, 5 ML-Es) in the extreme cases of a) $0 < NDVI < 0.1$ and b) $0 < VPD < 0.3$ kPa. The a6) and b6) are the trends of KGE for physical models, pure ML and hybrid models in these two extreme cases.

RMSE. This indicates that pure ML model is thought to be of high quality and perform well in training and even test datasets, but it deviates strongly for situations outside their valid domain (Reichstein et al., 2019). This may be attributable to the fact that geoscientific problems are often complex and physics-constrained; without an understanding of the ET process, upscale ET from flux towers at the regional scale using pure ML is completely confined to the training samples, which results in large uncertainties in the out-of-sample predictions (Yuan et al., 2020). Consequently, in the simulation of complex geoscientific problems, pure ML cannot take the place of physical models.

(2) In the case of $0 <$ NDVI<0.1 (Fig. 11 a), our hybrid models increase the R^2 of pure ML from 0.24 to 0.30 (ML-Gs) and 0.32 (ML-Es) and decrease the RMSE of pure ML from 0.55 mm/day to 0.52 mm/day (ML-Gs) and 0.51 mm/day (ML-Es). The KGE of pure ML is as low as 0.23. The KGE of ML-Gs and ML-Es are 0.34 and 0.48, respectively. When $0 <$ VPD < 0.3 kPa, the two hybrid models increase the R^2 of pure ML from 0.25 to 0.31 (ML-Gs) and 0.34 (ML-Es) (Fig. 11 b) and decrease the RMSE of pure ML from 0.57 mm/day to 0.54 mm/day (ML-Gs) and 0.53 mm/day (ML-Es). Under this circumstance, the KGE of pure ML is only 0.27; however, the KGE of ML-Gs and ML-Es are 0.36 and 0.44, respectively. This result is very encouraging, as it emphasizes that, by

integrating with physical modeling, hybrid models can practically improve generalization skills compared to pure ML. This is mainly because hybrid models involve a deep combination of physics and ML, which are optimized to simultaneously obtain high model performance and physical consistency (Yuan et al., 2020). In addition, for both extreme cases, ML-Es performs better than ML-Gs, with higher R^2 and lower RMSE. It indicates for extreme cases, hybrid model can achieve better generalization performance when more robust and well-defined physical mechanisms are coupled.

5.2. Interpretability of machine learning coupled in hybrid ET models

Improvement of the model performance is important but insufficient. The interpretability and understanding of ML models can be as crucial as the prediction accuracy in hydrometeorological applications. In this study, we used the Shapley value to analyze the interpretability of ML coupled in two hybrid models. We quantified the proportion of each variable's Shapley value as a whole as its contribution to the model.

As Fig. 12 shows, with the same model parameters, the same input variables show different contribution in the ML modeling stage of two hybrid models. For ML-Gs, the contribution of RH accounts for 25% in
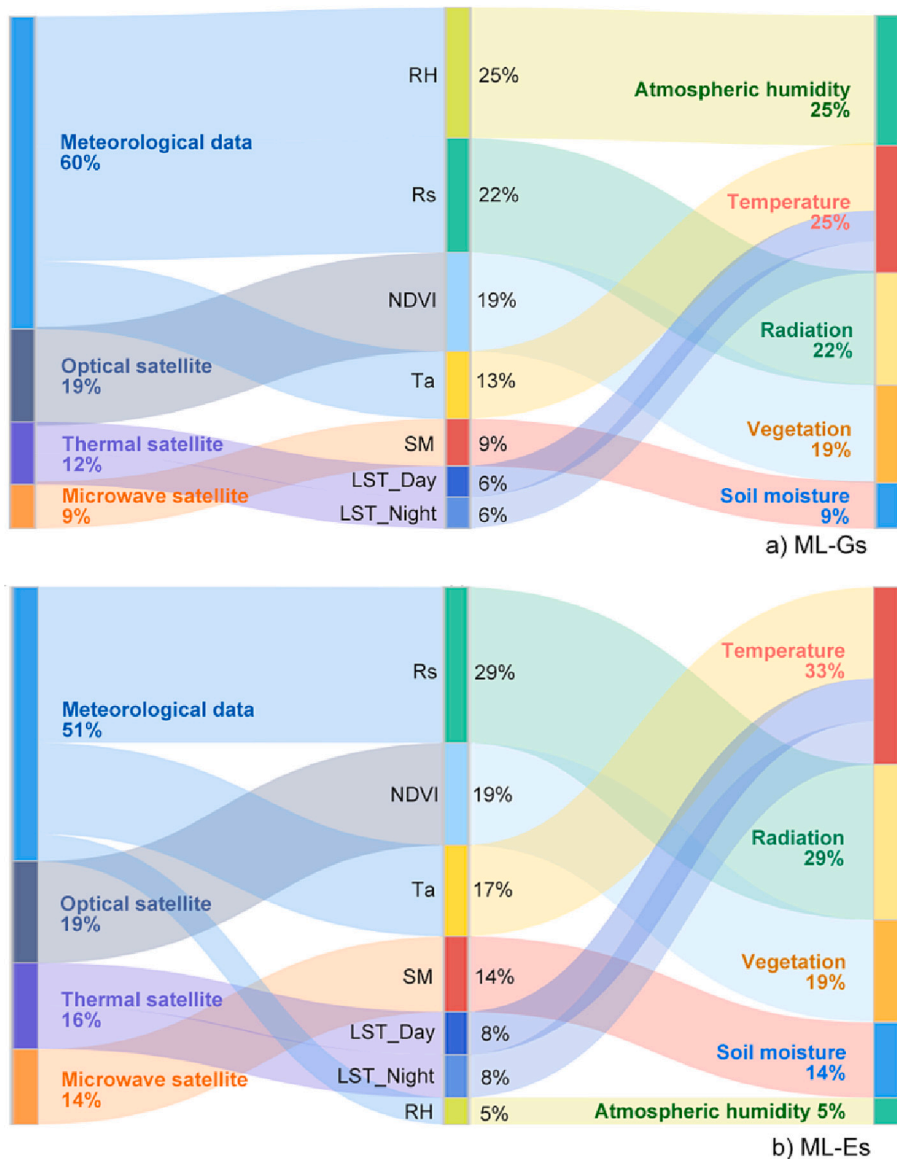

Fig. 12. The contributions of input variables in the machine learning stage of two hybrid models, a) is for ML-Gs and b) is for ML-Es. The input variables are the incident solar radiation (Rs), air temperature (Ta), relative humidity (RH), normalized difference vegetation index (NDVI), soil moisture (SM) and land surface temperature (LST_Day and LST_Night). For the data sources, the input variables include meteorological, optical satellite, thermal satellite and microwave satellite data. For the biophysical parameters, the input variables involve temperature, radiation, vegetation, soil moisture and atmospheric humidity. These contributions are expressed by Shapley values of input variables used in the machine learning modeling.

the machine learning stage of the hybrid modeling, followed by Rs (22%), NDVI (19%), Ta (13%), SM (9%), LST_Day (6%) and LST_Night (6%). From the perspective of process-based biophysical parameters, atmospheric humidity (RH) plays an important role, with a 25% contribution in the ML-Gs modeling, temperature (sum of Ta and LST) also accounts for 25%, and radiation (Rs) accounts for 22%. Studies showed the gs is mainly affected by moisture (RH or SM), Ta and VPD (Gan et al., 2018; Leuning et al., 2008; Mallick et al., 2015). Mu et al. (2011) used water cover fraction (RH^4) and LAI to calculate the conductance for canopy. Wang et al. (2010a) also parameterized gs as a linear function of RH and vegetation indices (NDVI or EVI) to estimate global ET. The interpretability of ML-Gs by the Shapley value shows common correlation with these ET algorithms.

As a soil evaporation-based ML method, the variable contribution in the machine learning stage of ML-Es shows substantial discrepancy with ML-Gs (Fig. 12). For ML-Es, Rs dominates the feature contribution in the machine learning stage of the hybrid modeling, with a proportion of 29%, and NDVI accounts for 19%, followed by Ta (17%), SM (14%), LST_Day (8%), LST_Night (8%) and RH (5%). From the perspective of biophysical parameters, the temperature (sum of Ta and LST) shows the greatest impact on the modeling of ML-Es, accounting for 33%. Radiation (Rs) also plays an important role (29%), vegetation (NDVI) contributes 19%, and soil moisture (SM) accounts for 14%. In contrast to ML-Gs, atmospheric humidity (RH) contributes the least to ML-Es. Studies have shown that the soil radiation dominates soil evaporation (Stoy et al., 2019), and apparent thermal inertia (ATI) derived from temperature can be well used as a soil moisture constraint (Yao et al., 2013). The vegetation coverage on the TP is sparse, NDVI can capture differences in ET due to variations in canopy cover, which plays an important role in the ET process of this region. The SM should be the dominating factor of ET under dry conditions (Yao et al., 2019), but the coarse spatial resolution of SM product that reduced the contribution of SM. The trends in the Shapley value agree well with previous process-based ET algorithms, which indicates that by coupling with physical constrains, ML-Es can describe the ET process as same as the physical models.

From the perspective of data sources, ML-Gs and ML-Es yield similar trends. Meteorological data plays an important role in the machine learning stage of two hybrid models, contributing >50%. One possible reason is that ET is a hydrometeorological process that is largely determined by meteorological factors (e.g., radiation, temperature, atmospheric humidity). The annual sunshine duration of the TP is >2500 h. The solar radiation plays an important role in the ET process. The monthly average Ta of the TP varies from −12.3 °C (January) to 10.1 °C (July), and the annual average Ta ranges from −1.4 °C to 0.7 °C (Zhang et al., 2021a). The temperature is also one of the dominant factors of ET process in alpine regions. The other reason is that we used three meteorological factors in the machine learning model building of hybrid models, which comprise a large proportion of meteorological data. Optical satellites provide the NDVI to characterize vegetation coverage, thermal satellites provide the LST to reflect surface water and heat flux, and microwave satellites can capture the variation in soil moisture. Our hybrid modeling approaches couple physical constraints and ML approach, as well as multiple data sources.

5.3. Uncertainties and future perspectives

5.3.1. Uncertainties in the ET estimation of hybrid models

Despite better performance than other models in reproducing spatiotemporal patterns and variability of ET, hybrid models in our approach still have some limitations and uncertainties. First, the representativeness of in situ measurements is not optimal at the pixel scale. Due to the specific geographical and climatic characteristics of the TP, the distribution of EC sites used in our study is limited and inhomogeneous. The sparse distribution of EC sites weakens the spatial representativeness of observation samples (Liu et al., 2018; Xu et al.,

2020). In addition, the mismatch of scales between satellite and in situ observations may cause 5% ~ 25% uncertainties in the spatiotemporal estimation (Li et al., 2009). The ground observations are assumed to be the "true" target in the hybrid modeling; however, there are no "true" ET observations for satellite pixel scales greater than a few kilometers—particularly on heterogeneous surfaces (Kalma et al., 2008). In addition, large eddies cannot be measured by the EC method, resulting in an energy imbalance, which can lead to approximately 5– 20% error (Foken, 2008; Mahrt, 2010). We collected as many as possible EC sites covering 2003–2018 to carry out our experiments. All the selected ground observations were quality controlled and energy imbalance corrected. The selection of training and verification sites fully considered the uniformity of spatial distribution and the involvement of typical plant function types, which to ensure the representativeness of the ground-based observations.

Second, there are uncertainties in the synergy between physical laws and machine learning. Although many well-known algorithms have been developed since the 1960s, there is still no consensus on the best way to parameterize ET on a regional scale (Yuan et al., 2021). Process-based ET algorithms exhibit large discrepancies and uncertainties across different terrestrial biomes (Jiménez et al., 2018). In our study, we have no ground-measured gs and ETs to calibrate our hybrid models, the target variables (gs and ETs) used for model training are inferred by LE observations based on ET algorithms (Eqs. 4–12). The uncertainty of inferences caused by physical models may directly affect the prediction accuracy of hybrid models. Similarly, ML-based models may fall into local optimization, which results in overfitting (Ball et al., 2017a; Ma et al., 2019). In our study, the hybrid models have been carefully tuned in the modeling of machine learning to avoid over fitting. Moreover, we used the algorithm calibrated for the TP by Yuan et al. (2021) to calculate the canopy transpiration used in ML-Es. All of the attempts were try to find an optimal junction between physical mechanism and high model performance.

Lastly, complex data sources contain great uncertainties. The hybrid models incorporate various input variables, including meteorological data and satellite-based observations. It has been reported that comparisons of meteorological variables between regional reanalysis datasets and observations at micrometeorological scales showed large biases (Rienecker et al., 2011). The analysis shows that the greatest disagreement between input variables for ET algorithms arises from the radiation dataset (Badgley et al., 2015). In addition, accurate satellite-based estimation of surface variables interacting with ET processes (e.g., SM and LST) remains a challenge, especially on heterogeneous underlying surfaces (Wang and Dickinson, 2012; Yao et al., 2017). For instance, a previous study found that the effects of topography, vegetation, precipitation, and LST greatly impact the spatial distribution of SM, resulting in large biases in the existing SM products of the TP (Liu et al., 2021). At the same time, the downscaling of coarser resolution SM products will also affect the accuracy of ET estimation (Hu et al., 2022; Sabaghy et al., 2020). Therefore, uncertainties could be inherited through errors from these input data sources.

5.3.2. Future perspectives for the hybrid modeling approach

Despite its uncertainties and limitations, our approach reveals a promising path for current and future applications. Overall, we identify three major perspectives and avenues for incorporating physical constraints into ML in hydrology. First, more comprehensive and dynamic parameterizations are needed. The process of ET is affected by the local microclimate with complex feedback mechanisms. Not all of the related parameters can be easily derived from a certain principle. There is no consensus on a framework to integrate multidimensional data sources (e.g., information from radiation, vegetation, soil, temperature, aerodynamics, water vapor pressure, or spatiotemporal autocorrelation) into a definite equation. However, ML has no restrictions on the selection of data sources, which can be used to extract knowledge from the data deluge to optimally describe the surface truth. Furthermore, instead of

assigning static parameters for different plant functional types, ML can make these parameterizations more dynamic and interdependent (Reichstein et al., 2019).

Second, complementarity between the physical model and machine learning will be explored. The interpretability and extrapolation of the physical model and the flexible data adaptability of ML can be complementary or even part of each other (Yuan et al., 2020). If formulations of a submodel or an intermediate variable are of semiempirical nature, this submodel or variable can be replaced by an ML model if sufficient observations are available (Brenowitz and Bretherton, 2018; de Bezenac et al., 2019; Gentine et al., 2018). Moreover, the physical regularization constraint can be used as the loss function of ML to replace the traditional loss function constructed by the errors between the predicted and target values (Zhao et al., 2019).

Finally, potential advantages beyond accuracy should be explored. The evaluation of ET satellite-derived models against ground-based observations is not enough, which hampers our understanding of regional natural ET. Due to the existence of the scale effect, we lack the true value at the satellite pixel scale to verify the remote sensing-based ET estimation. Facing this practical problem, the improvement of accuracy against ground measurements may not be the only aim in development of hybrid models. In addition to site-based verification, more comprehensive assessments, such as energy balance and water budget balance, the capture performance of agricultural irrigation and intensive precipitation, and feedback on extreme weather events, can provide future perspectives and avenues for exploring the structures and potential advantages of hybrid models.

6. Conclusions

We provide two sets of coupling estimates of spatiotemporally continuous ET for the TP by combining ML with physical constraints. These two hybrid models integrate the strengths of physical algorithms (theoretical basis, interpretability) and ML methods (data adaptability). The key results of our approach can be summarized as follows:

a) The evaluation results show that hybrid models enhance the performance of conventional process-based ET algorithms.
b) Although the accuracy of data-oriented pure ML is high, it cannot take the place of physical models, especially for the simulation of data-sparse regions and heterogeneous surfaces.
c) Hybrid models show better generation performance than pure ML at data-sparse regions as well as for the responses to extreme weather events.
d) For data-sparse regions and extreme cases, the more robust physical mechanism was coupled, the better generalization performance of hybrid model could achieve.

CRediT authorship contribution statement

Ke Shang: Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Visualization. **Yunjun Yao:** Conceptualization, Methodology, Formal analysis, Investigation, Resources, Writing – review & editing, Visualization, Supervision. **Zhenhua Di:** Visualization. **Kun Jia:** Visualization. **Xiaotong Zhang:** Visualization. **Joshua B. Fisher:** Writing – review & editing. **Jiquan Chen:** Writing – review & editing. **Xiaozheng Guo:** Data curation. **Junming Yang:** Data curation. **Ruiyang Yu:** Data curation. **Zijing Xie:** Data curation. **Lu Liu:** Data curation. **Jing Ning:** Data curation. **Lilin Zhang:** Data curation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Allen, R.G., Pereira, L.S., Howell, T.A., Jensen, M.E., 2011. Evapotranspiration information reporting: I. Factors governing measurement accuracy. Agric. Water Manag. 98, 899–920.

Badgley, G., Fisher, J.B., Jiménez, C., Tu, K.P., Vinukollu, R., 2015. On uncertainty in global terrestrial evapotranspiration estimates from choice of input forcing datasets. J. Hydrometeorol. 16, 1449–1455.

Ball, J., Anderson, D., Chan, C.S., 2017a. Comprehensive survey of deep learning in remote sensing: theories, tools, and challenges for the community. J. Appl. Remote. Sens. 11, 042609.

Ball, J.E., Anderson, D.T., Chan, C.S., 2017b. Comprehensive survey of deep learning in remote sensing: theories, tools, and challenges for the community. J. Appl. Remote. Sens. 11, 042609.

Bastiaanssen, W.G.M., Menenti, M., Feddes, R.A., Holtslag, A.A.M., 1998. A remote sensing surface energy balance algorithm for land (SEBAL). 1Formulation. J Hydrol. 212–213, 198–212.

Bodesheim, P., Jung, M., Gans, F., Mahecha, M.D., Reichstein, M., 2018. Upscaled diurnal cycles of land–atmosphere fluxes: a new global half-hourly data product. Earth Syst. Sci. Data 10, 1327–1365.

Brenowitz, N.D., Bretherton, C.S., 2018. Prognostic validation of a neural network unified physics parameterization. Geophys. Res. Lett. 45, 6289–6298.

Chen, J., 2021. Biophysical Models and Applications in Ecosystem Analysis. Higher Education Press, Beijing.

Chen, J.M., Liu, J., 2020. Evolution of evapotranspiration models using thermal and shortwave remote sensing data. Remote Sens. Environ. 237, 111594.

Chen, X., Massman, W.J., Su, Z., 2019. A column canopy-air turbulent diffusion method for different canopy structures. J. Geophys.Res.Atmos. 124, 488–506.

Cubitt, R., 1991. The shapley value: essays in honor of Lloyd S.Shapley. Econ. J. 101, 644–646.

de Bezenac, E., Pajot, A., Gallinari, P., 2019. Deep Learning for Physical Processes: Incorporating Prior Scientific Knowledge. In.

Dong, M., Jiang, Y., Zheng, C., Zhang, D., 2012. Trends in the thermal growing season throughout the Tibetan plateau during 1960–2009. Agric. For. Meteorol. 166–167, 201–206.

Falge, E., Baldocchi, D., Olson, R., Anthoni, P., Aubinet, M., Bernhofer, C., Burba, G., Ceulemans, R., Clement, R., Dolman, H., Granier, A., Gross, P., Grünwald, T., Hollinger, D., Jensen, N.-O., Katul, G., Keronen, P., Kowalski, A., Lai, C.T., Law, B.E., Meyers, T., Moncrieff, J., Moors, E., Munger, J.W., Pilegaard, K., Rannik, Ü., Rebmann, C., Suyker, A., Tenhunen, J., Tu, K., Verma, S., Vesala, T., Wilson, K., Wofsy, S., 2001. Gap filling strategies for defensible annual sums of net ecosystem exchange. Agric. For. Meteorol. 107, 43–69.

Fan, J., Yue, W., Wu, L., Zhang, F., Cai, H., Wang, X., Lu, X., Xiang, Y., 2018. Evaluation of SVM, ELM and four tree-based ensemble models for predicting daily reference evapotranspiration using limited meteorological data in different climates of China. Agric. For. Meteorol. 263, 225–241.

Fisher, J.B., Lee, B., Purdy, A.J., Halverson, G.H., Dohlen, M.B., Cawse-Nicholson, K., Wang, A., Anderson, R.G., Aragon, B., Arain, M.A., Baldocchi, D.D., Baker, J.M., Barral, H., Bernacchi, C.J., Bernhofer, C., Biraud, S.C., Bohrer, G., Brunsell, N., Cappelaere, B., Castro-Contreras, S., Chun, J., Conrad, B.J., Cremonese, E., Demarty, J., Desai, A.R., De Ligne, A., Foltýnová, L., Goulden, M.L., Griffis, T.J., Grünwald, T., Johnson, M.S., Kang, M., Kelbe, D., Kowalska, N., Lim, J.-H., Maïnassara, I., McCabe, M.F., Missik, J.E.C., Mohanty, B.P., Moore, C.E., Morillas, L., Morrison, R., Munger, J.W., Posse, G., Richardson, A.D., Russell, E.S., Ryu, Y., Sanchez-Azofeifa, A., Schmidt, M., Schwartz, E., Sharp, I., Sigut, L., Tang, Y., Hulley, G., Anderson, M., Hain, C., French, A., Wood, E., Hook, S., 2020. ECOSTRESS: NASA's Next Generation Mission to Measure Evapotranspiration From the International Space Station. Water Resourc. Res. 56 e2019WR026058.

Fisher, J.B., Melton, F., Middleton, E., Hain, C., Anderson, M., Allen, R., McCabe, M.F., Hook, S., Baldocchi, D., Townsend, P.A., Kilic, A., Tu, K., Miralles, D.D., Perret, J., Lagouarde, J.-P., Waliser, D., Purdy, A.J., French, A., Schimel, D., Famiglietti, J.S., Stephens, G., Wood, E.F., 2017. The future of evapotranspiration: global

requirements for ecosystem functioning, carbon and climate feedbacks, agricultural management, and water resources. Water Resour. Res. 53, 2618–2626.

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 Sens. Environ. 112, 901–919.

Foken, T., 2008. The energy balance closure problem: an overview. Ecol. Appl. 18, 1351–1367.

Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Ann. Stat. 29 (1189–1232), 1144.

Gan, R., Zhang, Y., Shi, H., Yang, Y., Eamus, D., Cheng, L., Chiew, F.H.S., Yu, Q., 2018. Use of satellite leaf area index estimating evapotranspiration and gross assimilation for Australian ecosystems. Ecohydrology 11, e1974.

Gentine, P., Pritchard, M., Rasp, S., Reinaudi, G., Yacalis, G., 2018. Could machine learning break the convection parameterization deadlock? Geophys. Res. Lett. 45, 5742–5751.

Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: implications for improving hydrological modelling. J. Hydrol. 377, 80–91.

Han, C., Ma, Y., Wang, B., Zhong, L., Ma, W., Chen, X., Su, Z., 2021. Long-term variations in actual evapotranspiration over the Tibetan plateau. Earth Syst. Sci. Data 13, 3513–3524.

He, J., Yang, K., Tang, W., Lu, H., Qin, J., Chen, Y., Li, X., 2020. The first high-resolution meteorological forcing dataset for land process studies over China. Sci. Data 7, 25.

Hu, Z., Chai, L., Crow, W.T., Liu, S., Zhu, Z., Zhou, J., Qu, Y., Liu, J., Yang, S., Lu, Z., 2022. Applying a wavelet transform technique to optimize general fitting models for SM analysis: a case study in downscaling over the Qinghai-Tibet Plateau. Remote Sens. 14, 3063.

Immerzeel, W.W., Beek, L.P.H.V., Bierkens, M.F.P., 2010. Climate change will affect the Asian water towers. Science 328, 1382–1385.

Impens, I., Lemeur, R., 1969. Extinction of net radiation in different crop canopies. Archiv für Meteorologie, Geophysik und Bioklimatologie, Serie B 17, 403–412.

Jasechko, S., Sharp, Z.D., Gibson, J.J., Birks, S.J., Yi, Y., Fawcett, P.J., 2013. Terrestrial water fluxes dominated by transpiration. Nature 496, 347–350.

Jiménez, C., Martens, B., Miralles, D.M., Fisher, J.B., Beck, H.E., Fernández-Prieto, D., 2018. Exploring the merging of the global land evaporation WACMOS-ET products based on local tower measurements. Hydrol. Earth Syst. Sci. 22, 4513–4533.

Jung, M., Koirala, S., Weber, U., Ichii, K., Gans, F., Camps-Valls, G., Papale, D., Schwalm, C., Tramontana, G., Reichstein, M., 2019. The FLUXCOM ensemble of global land-atmosphere energy fluxes. Sci.Data 6, 74.

Jung, M., Reichstein, M., Ciais, P., Seneviratne, S.I., Sheffield, J., Goulden, M.L., Bonan, G., Cescatti, A., Chen, J., 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., 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.

Kalma, J.D., Mcvicar, T.R., Mccabe, M.F., 2008. Estimating land surface evaporation: a review of methods using remotely sensed surface temperature data. Surv. Geophys. 29, 421–469.

Karpatne, A., Atluri, G., Faghmous, J., Steinbach, M., Banerjee, A., Ganguly, A., Shekhar, S., Samatova, N., Kumar, V., 2017. Theory-guided data science: a new paradigm for scientific discovery. IEEE Trans. Knowl. Data Eng. 29, 2318–2331.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T., 2017. LightGBM: a highly efficient gradient boosting decision tree. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. Curran Associates Inc, Long Beach, California, USA, pp. 3149–3157.

Kochendorfer, J., Castillo, E.G., Haas, E., Oechel, W.C., Paw, U.K.T., 2011. Net ecosystem exchange, evapotranspiration and canopy conductance in a riparian forest. Agric. For.Meteorol. 151, 544–553.

Kool, D., Agam, N., Lazarovitch, N., Heitman, J.L., Sauer, T.J., Bengal, A., 2014. A review of approaches for evapotranspiration partitioning. Agric. For. Meteorol. 184, 56–70.

Koppa, A., Rains, D., Hulsman, P., Poyatos, R., Miralles, D.G., 2022. A deep learning-based hybrid model of global terrestrial evaporation. Nat. Commun. 13, 1912.

Kustas, W.P., Norman, J.M., 1997. A two-source approach for estimating turbulent fluxes using multiple angle thermal infrared observations. Water Resour. Res. 33, 1495–1508.

Kuzmin, P.P., 1953. On method for investigations of evaporation from the snow cover. Trans. State Hydrol. Inst. 41, 34–52.

Lawrence, D.M., Thornton, P.E., Oleson, K.W., Bonan, G.B., 2007. The partitioning of evapotranspiration into transpiration, soil evaporation, and canopy evaporation in a GCM: impacts on land-atmosphere interaction. J. Hydrometeorol. 8, 862–880.

Leuning, R., Zhang, Y.Q., Rajaud, A., Cleugh, H., Tu, K., 2008. A simple surface conductance model to estimate regional evaporation using MODIS leaf area index and the Penman-Monteith equation. Water Resour. Res. 44.

Li, X., He, Y., Zeng, Z., Lian, X., Wang, X., Du, M., Jia, G., Li, Y., Ma, Y., Tang, Y., Wang, W., Wu, Z., Yan, J., Yao, Y., Ciais, P., Zhang, X., Zhang, Y., Zhang, Y., Zhou, G., Piao, S., 2018a. Spatiotemporal pattern of terrestrial evapotranspiration in China during the past thirty years. Agric. For. Meteorol. 259, 131–140.

Li, X., Yang, X., Ma, Y., Hu, G., Hu, X., Wu, X., Wang, P., Huang, Y., Cui, B., Wei, J., 2018b. Qinghai Lake Basin critical zone observatory on the Qinghai-Tibet Plateau. Vadose Zone J. 17, 180069.

Li, Z.-L., Tang, R., Wan, Z., Bi, Y., Zhou, C., Tang, B., Yan, G., Zhang, X., 2009. A review of current methodologies for regional evapotranspiration estimation from remotely sensed data. Sensors 9, 3801–3853.

Liang, S., Cheng, J., Jia, K., Jiang, B., Liu, Q., Xiao, Z., Yao, Y., Yuan, W., Zhang, X., Zhao, X., Zhou, J., 2020. The global LAnd Surface Satellite (GLASS) product suite. Bull. Am. Meteorol. Soc. 1–37.

Lipovetsky, S., Conklin, M., 2001. Analysis of regression in game theory approach. Appl. Stoch. Model. Bus. Ind. 17, 319–330.

Liu, J., Chai, L., Dong, J., Zheng, D., Wigneron, J.P., Liu, S., Zhou, J., Xu, T., Yang, S., Song, Y., Qu, Y., Lu, Z., 2021. Uncertainty analysis of eleven multisource soil moisture products in the third pole environment based on the three-corned hat method. Remote Sens. Environ. 255, 112225.

Liu, S., Li, X., Xu, Z., Che, T., Xiao, Q., Ma, M., Liu, Q., Jin, R., Guo, J., Wang, L., 2018. The Heihe integrated observatory network: a basin-scale land surface processes observatory in China. Vadose Zone J. 17.

Liu, X., Chen, B., 2000. Climatic warming in the Tibetan Plateau during recent decades. Int. J. Climatol. 20, 1729–1742.

Lu, X., Zhuang, Q., 2010. Evaluating evapotranspiration and water-use efficiency of terrestrial ecosystems in the conterminous United States using MODIS and AmeriFlux data. Remote Sens. Environ. 114, 1924–1939.

Lundberg, S., Erion, G.G., Lee, S., 2018. Consistent Individualized Feature Attribution for Tree Ensembles. arXiv: Learning.

Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., Johnson, B.A., 2019. Deep learning in remote sensing applications: a meta-analysis and review. ISPRS J. Photogramm. Remote Sens. 152, 166–177.

Ma, N., Zhang, Y., 2022. Increasing Tibetan Plateau terrestrial evapotranspiration primarily driven by precipitation. Agric. For. Meteorol. 317, 108887.

Ma, N., Zhang, Y., Xu, C.-Y., Szilagyi, J., 2015. Modeling actual evapotranspiration with routine meteorological variables in the data-scarce region of the Tibetan Plateau: comparisons and implications. J. Geophys.Res.Biogeosci. 120, 1638–1657.

Ma, Y., Hu, Z., Xie, Z., Ma, W., Wang, B., Chen, X., Li, M., Zhong, L., Sun, F., Gu, L., Han, C., Zhang, L., Liu, X., Ding, Z., Sun, G., Wang, S., Wang, Y., Wang, Z., 2020. A long-term (2005–2016) dataset of integrated land-atmosphere interaction observations on the Tibetan Plateau. Earth Syst. Sci. Data 12, 2937–2957.

Ma, Y.M., Kang, S.C., Zhu, L.P., Xu, B.Q., Tian, L.D., Yao, T.D., 2008. Tibetan observation and research platform- atmosphere–land interaction over a heterogeneous landscape. Bull. Am. Meteorol. Soc. 89, 1487–1492.

Mahrt, L., 2010. Computing turbulent fluxes near the surface: needed improvements. Agric. For. Meteorol. 150, 501–509.

Mallick, K., Boegh, E., Trebs, I., Alfieri, J.G., Kustas, W.P., Prueger, J.H., Niyogi, D., Das, N., Drewry, D.T., Hoffmann, L., Jarvis, A.J., 2015. Reintroducing radiometric surface temperature into the Penman-Monteith formulation. Water Resour. Res. 51, 6214–6243.

Marshall, M., Tu, K., Andreo, V., 2020. On parameterizing soil evaporation in a direct remote sensing model of ET: PT-JPL. Water Resour. Res. 56 e2019WR026290.

Martens, B., Miralles, D.G., Lievens, H., van der Schalie, R., de Jeu, R.A.M., Fernández-Prieto, D., Beck, H.E., Dorigo, W.A., Verhoest, N.E.C., 2017. GLEAM v3: satellite-based land evaporation and root-zone soil moisture. Geosci. Model Dev. 10, 1903–1925.

Medlyn, B.E., Duursma, R.A., Eamus, D., Ellsworth, D.S., Prentice, I.C., Barton, C.V.M., Crous, K.Y., de Angelis, P., Freeman, M., Wingate, L., 2011. Reconciling the optimal and empirical approaches to modelling stomatal conductance. Glob. Chang. Biol. 17, 2134–2144.

Miralles, D.G., Jiménez, C., Jung, M., Michel, D., Ershadi, A., McCabe, M.F., Hirschi, M., Martens, B., Dolman, A.J., Fisher, J.B., Mu, Q., Seneviratne, S.I., Wood, E.F., Fernández-Prieto, D., 2016. The WACMOS-ET project – part 2: evaluation of global terrestrial evaporation data sets. Hydrol. Earth Syst. Sci. 20, 823–842.

Monteith, J.L., 1965. Evaporation and environment. Symp. Soc. Exp. Biol. 19, 205–234.

Mu, Q., Heinsch, F.A., Zhao, M., Running, S.W., 2007. Development of a global evapotranspiration algorithm based on MODIS and global meteorology data. Remote Sens. Environ. 111, 519–536.

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

Nishida, K., Nemani, R.R., Running, S.W., Glassy, J.M., 2003. An operational remote sensing algorithm of land surface evaporation. J. Geophys. Res. 108.

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

Parlange, M., Eichinger, W., Albertson, J., 1995. Regional scale evaporation and the atmospheric boundary layer. Rev. Geophys. 33, 99–124.

Penman, H.L., Keen, B.A., 1948. Natural evaporation from open water, bare soil and grass. Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences 193, 120–145.

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

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., <check>Prabhat, check, 2019. Deep learning and process understanding for data-driven Earth system science. Nature 566, 195–204.

Rienecker, M.M., Suarez, M.J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M.G., Schubert, S.D., Takacs, L., Kim, G.-K., Bloom, S., Chen, J., Collins, D., Conaty, A., da Silva, A., Gu, W., Joiner, J., Koster, R.D., Lucchesi, R., Molod, A., Owens, T., Pawson, S., Pegion, P., Redder, C.R., Reichle, R., Robertson, F. R., Ruddick, A.G., Sienkiewicz, M., Woollen, J., 2011. MERRA: NASA's modern-era retrospective analysis for research and applications. J. Clim. 24, 3624–3648.

Sabaghy, S., Walker, J.P., Renzullo, L.J., Akbar, R., Chan, S., Chaubell, J., Das, N., Dunbar, R.S., Entekhabi, D., Gevaert, A., Jackson, T.J., Loew, A., Merlin, O., Moghaddam, M., Peng, J., Peng, J., Piepmeier, J., Rüdiger, C., Stefan, V., Wu, X., Ye, N., Yueh, S., 2020. Comprehensive analysis of alternative downscaled soil moisture products. Remote Sens. Environ. 239, 111586.

Schlesinger, W.H., Jasechko, S., 2014. Transpiration in the global water cycle. Agric. For. Meteorol. 189–190, 115–117.

Shang, K., Yao, Y., Li, Y., Yang, J., Jia, K., Zhang, X., Chen, X., Bei, X., Guo, X., 2020. Fusion of five satellite-derived products using extremely randomized trees to estimate terrestrial latent heat flux over Europe. Remote Sens. 12, 687.

Shang, K., Yao, Y., Liang, S., Zhang, Y., Fisher, J.B., Chen, J., Liu, S., Xu, Z., Zhang, Y., Jia, K., Zhang, X., Yang, J., Bei, X., Guo, X., Yu, R., Xie, Z., Zhang, L., 2021. DNN-MET: A deep neural networks method to integrate satellite-derived evapotranspiration products, eddy covariance observations and ancillary information. Agric. For. Meteorol. 308-309, 108582.

Shuttleworth, W.J., 2007. Putting the "vap" into evaporation. Hydrol. Earth Syst. Sci. Discuss. 11, 210–244.

Smith, W.K., Dannenberg, M.P., Yan, D., Herrmann, S., Barnes, M.L., Barron-Gafford, G. A., Biederman, J.A., Ferrenberg, S., Fox, A.M., Hudson, A., Knowles, J.F., MacBean, N., Moore, D.J.P., Nagler, P.L., Reed, S.C., Rutherford, W.A., Scott, R.L., Wang, X., Yang, J., 2019. Remote sensing of dryland ecosystem structure and function: progress, challenges, and opportunities. Remote Sens. Environ. 233, 111401.

Stoy, P.C., El-Madany, T.S., Fisher, J.B., Gentine, P., Gerken, T., Good, S.P., Klosterhalfen, A., Liu, S., Miralles, D.G., Perez-Priego, O., Rigden, A.J., Skaggs, T.H., Wohlfahrt, G., Anderson, R.G., Coenders-Gerrits, A.M.J., Jung, M., Maes, W.H., Mammarella, I., Mauder, M., Migliavacca, M., Nelson, J.A., Poyatos, R., Reichstein, M., Scott, R.L., Wolf, S., 2019. Reviews and syntheses: turning the challenges of partitioning ecosystem evaporation and transpiration into opportunities. Biogeosciences 16, 3747–3775.

Su, Z., 2002. The Surface Energy Balance System (SEBS) for estimation of turbulent heat fluxes. Hydrol. Earth Syst. Sci. 6, 85–100.

Talsma, C.J., Good, S.P., Jimenez, C., Martens, B., Fisher, J.B., Miralles, D.G., McCabe, M. F., Purdy, A.J., 2018. Partitioning of evapotranspiration in remote sensing-based models. Agric. For. Meteorol. 260–261, 131–143.

Trajkovic, S., Todorovic, B., Stankovic, M., 2003. Forecasting of reference evapotranspiration by artificial neural networks. J. Irrig. Drain. Eng. 129, 454–457.

Trenberth, K.E., Fasullo, J.T., Kiehl, J., 2009. Earth's global energy budget. Bull. Am. Meteorol. Soc. 90, 311–324.

Twine, T.E., Kustas, W.P., Norman, J.M., Cook, D.R., Houser, P.R., Meyers, T.P., Prueger, J.H., Starks, P.J., Wesely, M.L., 2000. Correcting eddy-covariance flux underestimates over a grassland. Agric. For. Meteorol. 103, 279–300.

Wang, J., Zuo, W., 2009. Geographic atlas of China. SinoMaps Press.

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

Wang, K., Dickinson, R.E., Wild, M., Liang, S., 2010a. Evidence for decadal variation in global terrestrial evapotranspiration between 1982 and 2002: 1.Model development. J. Geophys. Res.: Atmos. 115.

Wang, K., Dickinson, R.E., Wild, M., Liang, S., 2010b. Evidence for decadal variation in global terrestrial evapotranspiration between 1982 and 2002: 2.Results. J. Geophys. Res. 115.

Wang, L., Zhang, Y., Yao, Y., Xiao, Z., Shang, K., Guo, X., Yang, J., Xue, S., Wang, J., 2021. GBRT-based estimation of terrestrial latent heat flux in the Haihe River basin from satellite and reanalysis datasets. Remote Sens. 13, 1054.

Wang, W., Li, J., Yu, Z., Ding, Y., Xing, W., Lu, W., 2018. Satellite retrieval of actual evapotranspiration in the Tibetan Plateau: components partitioning, multidecadal trends and dominated factors identifying. J. Hydrol. 559, 471–485.

Xiao, Z., Liang, S., Sun, R., Wang, J., Jiang, B., 2015. Estimating the fraction of absorbed photosynthetically active radiation from the MODIS data based GLASS leaf area index product. Remote Sens. Environ. 171, 105–117.

Xiao, Z., Liang, S., Wang, J., Chen, P., Yin, X., Zhang, L., Song, J., 2014. Use of general regression neural networks for generating the GLASS leaf area index product from time-series MODIS surface reflectance. IEEE Trans. Geosci. Remote Sens. 52, 209–223.

Xu, Z., Liu, S., Zhu, Z., Zhou, J., Shi, W., Xu, T., Yang, X., Zhang, Y., He, X., 2020. Exploring evapotranspiration changes in a typical endorheic basin through the integrated observatory network. Agric. For. Meteorol. 290, 1–14.

Yang, H., Yang, D., Lei, Z., 2013. Seasonal variability of the complementary relationship in the Asian monsoon region. Hydrol. Process. 27, 2736–2741.

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, G., Li, Y., Zhao, S., 2013. MODIS-driven estimation of terrestrial latent heat flux in China based on a modified priestley-Taylor algorithm. Agric. For. Meteorol. 171–172, 187–202.

Yao, Y., Liang, S., Li, X., Chen, J., Liu, S., Jia, K., Zhang, X., Xiao, Z., Fisher, J.B., Mu, Q., Pan, M., Liu, M., Cheng, J., Jiang, B., Xie, X., Grünwald, T., Bernhofer, C., Roupsard, O., 2017. Improving global terrestrial evapotranspiration estimation using support vector machine by integrating three process-based algorithms. Agric. For. Meteorol. 242, 55–74.

Yao, Y., Liang, S., Li, X., Hong, Y., Fisher, J.B., Zhang, N., Chen, J., Cheng, J., Zhao, S., Zhang, X., 2014. Bayesian multimodel estimation of global terrestrial latent heat flux from eddy covariance, meteorological, and satellite observations. J. Geophys. Res. 119, 4521–4545.

Yao, Y., Zhang, Y., Liu, Q., Liu, S., Jia, K., Zhang, X., Xu, Z., Xu, T., Chen, J., Fisher, J.B., 2019. Evaluation of a satellite-derived model parameterized by three soil moisture constraints to estimate terrestrial latent heat flux in the Heihe River basin of Northwest China. Sci. Total Environ. 695, 133787.

Yuan, L., Ma, Y., Chen, X., Wang, Y., Li, Z., 2021. An enhanced MOD16 evapotranspiration model for the Tibetan Plateau during the unfrozen season. J. Geophys.Res.Atmos. 126 e2020JD032787.

Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., Xu, H., Tan, W., Yang, Q., Wang, J., Gao, J., Zhang, L., 2020. Deep learning in environmental remote sensing: achievements and challenges. Remote Sens. Environ. 241, 111716.

Zhang, L., Chen, Z., Zhang, X., Zhao, L., Li, Q., Chen, D., Tang, Y., Gu, S., 2021a. Water 13, 2061.

Zhang, Q., Kong, D., Shi, P., Singh, V.P., Sun, P., 2018. Vegetation phenology on the Qinghai-Tibetan plateau and its response to climate change (1982–2013). Agric. For. Meteorol. 248, 408–417.

Zhang, X., Zhou, J., Liang, S., Wang, D., 2021b. A practical reanalysis data and thermal infrared remote sensing data merging (RTM) method for reconstruction of a 1-km all-weather land surface temperature. Remote Sens. Environ. 260, 112437.

Zhao, W., Liu, B., Chang, X., Yang, Q., Yang, Y., Liu, Z., Cleverly, J., Eamus, D., 2016. Evapotranspiration partitioning, stomatal conductance, and components of the water balance: a special case of a desert ecosystem in China. J. Hydrol. 538, 374–386.

Zhao, W.L., Gentine, P., Reichstein, M., Zhang, Y., Zhou, S., Wen, Y., Lin, C., Li, X., Qiu, G.Y., 2019. Physics-constrained machine learning of evapotranspiration. Geophys. Res. Lett. 46, 14496–14507.

Zheng, D., 1999. Eco-Geographical Regionalization of China. SinoMaps Press. In.