Agricultural and Forest Meteorology 372 (2025) 110645

Contents lists available at ScienceDirect

Agricultural and Forest Meteorology

journal homepage: www.elsevier.com/locate/agrformet

A multimodal machine learning fused global 0.1° daily evapotranspiration dataset from 1950-2022

Qingchen Xu [a,1], Lu Li [a,1], Zhongwang Wei [a,*], Xingjie Lu [a], Nan Wei [a], Xuhui Lee [b], Yongjiu Dai [a]

[a] Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Guangdong Province Key Laboratory for Climate Change and Natural Disaster Studies, School of Atmospheric Sciences, Sun Yat-Sen University, Zhuhai, China
[b] Yale School of the Environment, Yale University, New Haven, CT, USA

ARTICLE INFO

Keywords:
Evapotranspiration
Automated machine learning
Multimodal
Prolonged data reconstruction

ABSTRACT

Evapotranspiration (ET) is the second largest hydrological flux over the land surface and connects water, energy, and carbon cycles. However, large uncertainties exist among current ET products due to their coarse spatial resolutions, short temporal coverages, and reliance on assumptions. This study introduces a multimodal machine learning framework to generate a high-resolution (0.1°, daily), long-term (1950–2022) global ET dataset by fusing 13 state-of-the-art ET products encompassing remote sensing, machine learning, land surface models, and reanalysis data relying on extensive flux tower observations (462 sites). The framework reconstructs the individual ET products to consistent spatiotemporal resolutions and time ranges using Light Gradient Boosting Machine (LightGBM) models, and the Automated Machine Learning (AutoML) technique was used to fuse ET using 13 reconstructed ET products, ERA5-land atmospheric forcings and ancillary data as predictors. In-situ observations are utilized for model training and validation. Results demonstrate significant improvements over existing datasets, with our product achieving the highest accuracy (KGE = 0.857, RMSE = 0.726 mm/day) against in situ measurements across ecosystems and regions. The fused ET dataset realistically captures spatiotemporal variability and corrects the systematic underestimation bias prevalent in other datasets, particularly in wet regions. This novel high spatial-temporal ET dataset enables more robust assessments for water, energy, and carbon cycle applications on regional hydrology and ecology. The introduced data integration methodology also provides a valuable framework for fusing multiple geoscience datasets with disparate properties.

1. Introduction

Evapotranspiration (ET) is the movement of water from the land surface to the atmosphere as vapor (Yang et al., 2023). ET plays a crucial role in the water, energy, and carbon nexus (Brutsaert, 1982; Jasechko et al., 2013; Koppa et al., 2022; Miralles et al., 2016; Pan et al., 2020). As the second largest hydrological flux on land after precipitation, ET returns over 60 % of initial precipitation back into the atmosphere (Oki and Kanae, 2006) and consumes 50 % of the total net radiation received by the land surface on average (Trenberth et al., 2009). Moreover, ET is closely linked to the carbon cycle through the regulation of plant stomata, which simultaneously control transpiration and photosynthesis (Jasechko et al., 2013; Wong et al., 1979). Given its critical role in these processes, accurately estimating ET globally is of great importance across various scientific disciplines.

ET can be quantified using various measurement techniques at different scales (Wang and Dickinson, 2012). At the plot scale, methods like the Bowen ratio, lysimeter and eddy covariance are widely used for continuous measurements (Bodesheim et al., 2018; Ma et al., 2020; Wei et al., 2018), providing insights into ecosystem-scale water cycles. However, these methods have limitations in spatial and temporal coverage. At larger scales, such as river basins or regions, ET can be estimated using surface water budget or atmospheric water balance methods (Li et al., 2019b; Zhang et al., 2023). Still, their performance decreases when applied to finer spatial-temporal scales. Recent advancements in high-resolution remote sensing have led to multiple methodologies for estimating ET spatially (e.g., Koppa et al., 2022; Martens et al., 2017; Zhang et al., 2019). Despite these developments,

* Corresponding author.
E-mail address: weizhw6@mail.sysu.edu.cn (Z. Wei).
[1] Qingchen Xu and Lu Li contributed equally to this work.

https://doi.org/10.1016/j.agrformet.2025.110645
Received 13 September 2024; Received in revised form 15 May 2025; Accepted 21 May 2025

uncertainties in ET estimation using different methods remain substantial (McCabe et al., 2016; Miralles et al., 2016; Wantong et al., 2022). The main challenge in simulating ET lies in its significant variation across different types of land cover and its substantial temporal fluctuations. These variations are driven by atmospheric conditions such as radiation, wind speed, temperature, humidity, and carbon dioxide concentration, as well as surface attributes including land use, soil properties, and vegetation structure (ElGhawi et al., 2023; Pan et al., 2020; Shang et al., 2023). Furthermore, ET is inherently a nonlinear and complex process, which introduces significant uncertainty in the parameterization of models attempting to represent it accurately (Pan et al., 2020).

There are very few ET products currently available that provide daily ET estimates for the period 1950–2022 globally (Lu et al., 2021; Muñoz-Sabater et al., 2021), and most existing long-term global ET products with daily temporal resolution are available at coarse spatial resolutions ranging from 0.25° to 0.5° (Pan et al., 2020; Piao et al., 2019; Xie et al., 2022; Yang et al., 2023). Consequently, substantial knowledge gaps exist in understanding how ET responds to climate change, such as global greening, and model-based quantifications of these responses are highly uncertain (Yang et al., 2023). Therefore, it is imperative to develop a high-resolution global ET dataset that provides continuous spatial and temporal coverage to assess land surface energy, water, and carbon cycles under climate change.

To develop advanced high-resolution, prolonged time ranges, and high-quality ET data, several studies (Bodesheim et al., 2018; Hao et al., 2019; Lu et al., 2021; Lu and Zhuang, 2010; Shang et al., 2021; Xie et al., 2022; Xu et al., 2018; Yang et al., 2006) have utilized statistical-based fusion method to merge multimodal data, including station-based, remote sensing, ML-based, land surface models (LSMs), and reanalysis data, aiming to mitigate the high uncertainty of multisource ET datasets. These fusion methods could be categorized into two pathways. The first approach utilized statistical techniques, such as simple averaging (Mueller et al., 2013), Empirical Orthogonal Functions (EOF) (Feng et al., 2016), Reliability Ensemble Averaging (Lu et al., 2021; Yoo et al., 2020), Bayesian Model Averaging (BMA) (Hao et al., 2019; Chen et al., 2015b), Taylor skill fusion (Yao et al., 2017b), and triple-collocation (Changming Li et al., 2022), to calculate the fusion weight of individual ET datasets according to field observation databases (e.g., FLUXNET) (Pastorello et al., 2020). The second approach is the direct upscaling of site observations to a global scale using various machine learning (ML) methods and atmospheric forcing and ancillary datasets (Bodesheim et al., 2018; Jung et al., 2019, 2010; Koppa et al., 2022; Shang et al., 2023). These approaches offer advantages by relaxing certain underlying assumptions and leveraging the strength of multimodal ET data (Amani and Shafizadeh-Moghadam, 2023). As a result, various fusion ET datasets have been generated and widely utilized in diverse applications.

While these fusion products have been widely used in various research fields, they face several limitations due to the methods and data employed to preprocess and fuse multisource data. The first constraint is the need to select a common time range among multiple input datasets, which restricts the temporal coverage of the fused data. Consequently, most fused ET products are limited in their time span, typically spanning only around 20–40 years (e.g., Martens et al., 2017; Yu et al., 2022; Zhang et al., 2019). This limited temporal coverage hinders the application of these products in long-term studies and climate change analyses. Secondly, another issue arises from the inconsistencies in the spatial and temporal resolutions of the products used for fusion. A common approach to address this issue is to employ linear interpolation (e.g., Lu et al., 2021). For example, interpolating from 0.5° to 0.25° spatial resolution or from 8-day to daily temporal resolution. However, this introduces substantial errors and uncertainties in the resulting fused ET estimates. The commonly used linear interpolation assumes a linear relationship between data points, which may not accurately capture the complex nonlinear spatial and temporal dynamics of ET. Thirdly, previous studies have often fused few ET datasets using a small number of in-situ observations (e.g., Shang et al., 2021), limiting the quality and reliability of fused products. However, with the availability of extensive multisource ET data and long-term EC flux tower networks established at regional and global scales, assessing data has become much easier recently (Baldocchi et al., 2001; Ma et al., 2020). Finally, as numerous ML approaches have been increasingly used to fuse multimodal ET data, the proposed methods heavily rely on human intervention, such as model design and hyperparameters tuning, which increase the time cost and bring the error derived from expert bias, limiting the generalizability of the proposed ML methods.

In recent years, the emergence of Automated Machine Learning (AutoML) has had a profound impact on ML applications in the meteorology field, particularly in the domain of non-linear hydrological process modeling (Mangalath Ravindran et al., 2022). AutoML has revolutionized geographic data science applications by automating and streamlining the arduous tasks associated with the development of ML pipelines. Compared to other ML methods, AutoML offers an enticing alternative for practitioners to enhance the quality of ET models by automatically selecting optimal hyperparameters for a chosen model or an ensemble of models, with minimal user intervention (He et al., 2021). Moreover, AutoML algorithms have been specifically designed to handle large datasets and exhibit scalability (LeDell and Poirier, 2020), making them well-suited for accommodating the increasing volumes of data required for global high-resolution ET estimation. Consequently, the utilization of AutoML holds the potential to improve the accuracy of ET estimation.

To address the above limitations, we introduce an advanced multimodal ML framework that integrates information from various sources to generate a high-resolution and long-term dataset of ET. Our approach focuses on two main aspects: the data itself and the techniques used. Regarding the data, we have incorporated 13 diverse and high-quality ET products, including remote sensing, ML, fusion products, LSMs, and reanalysis. We also incorporate a larger number of target sites for training compared to previous studies, enhancing the reliability of the fusion results. Consequently, our fusion product exhibits a significantly improved spatiotemporal resolution (0.1°, daily) and extended temporal coverage (from 1950 to 2022) compared to existing datasets. To address the challenges of integrating multimodal datasets, our approach utilizes advanced techniques. We employ nonlinear ML algorithms to reconstruct the disparate spatiotemporal resolutions and time ranges of the original multisource ET data instead of using traditional linear interpolation methods. We then extensively utilize AutoML to improve upon the traditional manual parameter tuning and single-model approach. In summary, our study presents a novel approach for estimating ET that overcomes the limitations of previous datasets and incorporates advanced techniques for data integration. The resulting fusion product offers improved data quality, enhanced spatiotemporal resolution, and extended temporal coverage, making it a valuable resource for various applications requiring accurate ET estimation.

2. Data sources

2.1. Evapotranspiration (ET) data

We combined 13 different ET datasets, covering different time lengths, temporal resolutions, and spatial resolutions, including Global Land Evaporation Amsterdam Model datasets (GLEAM v3.6a, GLEAM v3.6b) (Martens et al., 2017), ML-based ET products (FLUXCOM in 0.25° and 0.0833° resolutions, GLEAM_hybrid) (Jung et al., 2019; Koppa et al., 2022), surface energy balance-based global land ET (EB-ET) products (Chen et al., 2021), the three-temperature model-based global terrestrial ET products (ET-3T) (Yu et al., 2022), the reliability ensemble averaging method-based fusion products (REA) (Lu et al., 2021), the coupled diagnostic biophysical model product (PML-V2) (Zhang et al., 2019), ECMWF's global reanalysis product (ERA5) and for land applications global reanalysis product (ERA5-Land) (Hersbach et al., 2020; Muñoz-Sabater et al., 2021), the Global Land Data Assimilation NASA's System Catchment LSM (CLSM version2.2) and Noah LSM (Noah version2.1) datasets (Li et al., 2019a; Rodell et al., 2004).

These ET products are utilized because of the following reasons: 1) they are generated by a number of different methods, such as remote sensing-based methods e.g., (PML-V2, EB-ET, ET-3T), Priestley-Taylor algorithms (e.g., GLEAM), surface energy balance system models (e.g., EB-ET), reanalysis products e.g., (ERA5, ERA5-Land, GLDAS), ML-based products (e.g., FLUXCOM, GLEAM_hybrid); 2) these products perform well globally and regionally; 3) these products encompass different resolutions and time scales, and use various forcing data as inputs.

Building on this, considering that the accuracy of machine learning models highly depends on the quality of input data and the representativeness of samples (Amani and Shafizadeh-Moghadam, 2023), our fusion model takes full advantage of these diverse ET products, each with its own unique strengths. For example, the PML-V2 product couples GPP and ET estimation (Zhang et al., 2019), satellite-based products capture seasonal vegetation changes and spatial variability (Xie et al., 2022), Penman-Monteith-based estimates rely on wind speed and VPD dataset (Yao et al., 2013), and LSM simulations reflect various biophysical processes. Our approach incorporates model differences, allowing the algorithm to learn optimal information combinations and capture complex interactions between ET components. Building on previous fusion efforts like GLASS (Xie et al., 2022), our method combines the advantages of various models and input mechanisms to improve ET estimation accuracy and reduce uncertainty across diverse environmental conditions (Lu et al., 2021; Shang et al., 2021; Xie et al., 2022).

2.2. ERA5-land meteorological forcing data

ERA5-Land is a reanalysis dataset that provides high-resolution information on various meteorological forcing and land surface variables (Muñoz-Sabater et al., 2021). It is produced by the European Centre for Medium-Range Weather Forecasts (ECMWF) and offers several advantages for our study. Firstly, ERA5-Land has a grid spacing of approximately 9 km, enabling the capture of fine-scale spatial patterns in ET and its drivers. Secondly, the dataset spans from 1950 to the present, allowing for expanding the temporal coverage of fused ET data. Lastly, ERA5-Land contains over 50 variables describing water and energy cycles, allowing for a comprehensive set of predictors for ET modeling. For our study, we selected a subset of eight meteorological forcing variables and two land surface variables commonly used in ET modeling (Table 1), including 10 m U/V component of wind, 2 m temperature, total precipitation, specific humidity, surface pressure, surface solar radiation downwards, surface thermal radiation downwards, and two layers of soil moisture data.

2.3. Ancillary data

In addition to multisource ET and ERA5-Land forcing datasets, we incorporated several ancillary datasets to enhance the spatial representation and modeling of ET. These datasets provide information on various land surface properties that influence ET processes (Shang et al., 2021). The ancillary variables used in our study include:

Land cover: We used a static land cover map from the MODIS MCD12C1 V006 product for the year 2006 (Fig. S1, https://lpdaac.usgs.gov/products/mcd12c1v006/) (Friedl et al., 2010). This dataset, derived from MODIS imagery, provides a high-resolution characterization of land cover types, with an overall accuracy of about 75 % (Friedl et al., 2010).

Soil properties: We incorporated the 10 km SoilGrids dataset developed by Hengl et al. (2017) (Fig. S2). This dataset offers improved spatial detail and attribute accuracy compared to previous soil property products.

Soil hydraulic parameters: We used the 1 km soil hydraulic parameters dataset from Zhang et al. (2018) (Fig. S3). This dataset includes six parameters (field capacity, saturated hydraulic conductivity (K_s), plant available water, residual water (θ_r), and saturated water (θ_s)), which derived from hierarchical pedotransfer functions for the Kosugi water retention model.

Climate zones: We incorporated the Köppen-Geiger climate classification dataset from Beck et al. (2018) (Fig. S4). This high-resolution (1

Table 1
Summary of the data sources used in this study to generate an improved global daily ET dataset with 0.1° from 1950 to 2022.

Category	Variable			Spatial Resolution	Time Resolution	Period
ERA5-Land	Forcing	10 m u/v component of wind		0.1°	daily	1950–2022
		2 m temperature		0.1°	daily	1950–2022
		total precipitation		0.1°	daily	1950–2022
		specific humidity		0.1°	daily	1950–2022
		surface pressure		0.1°	daily	1950–2022
		surface solar radiation downwards		0.1°	daily	1950–2022
		surface thermal radiation downwards		0.1°	daily	1950–2022
	Land surface variables	volumetric soil water layer 1		0.1°	daily	1950–2022
		volumetric soil water layer 2		0.1°	daily	1950–2022
Ancillary	Soil properties (SoilGrids)			10km	–	–
	Soil hydraulic parameters (Kosugi)			1km	–	–
	Land cover (MCD12C1 V006, 2006)			0.05°	–	2006
	Köppen-Geiger climate classification			0.083°	–	1980–2016
	MERIT DEM			90m	–	–
	Leaf area index			500m	8-daily	2000–2021
ET	Ensembled	REA		0.25°	daily	1980–2017
	Remote sensing	GLEAM	V3.6a	0.25°	daily	2000–2021
			V3.6b	0.25°	daily	2003–2021
		PMLV2		0.1°	8-day	2001–2020
		EB-ET		0.1°	daily	2001–2016
		ET-3T		0.25°	daily	2001–2020
	Reanalysis	ERA5		0.25°	daily	2000–2021
		GLDAS	CLSM-2.2	0.25°	daily	2004–2021
			Noah-2.1	0.25°	daily	2000–2020
		ERA5-Land		0.1°	daily	1950–2022
	Machine Learning	FLUXCOM		0.25°	daily	2001–2015
				0.0833°	8-day	2001–2015
		GLEAM_hybrid		0.25°	daily	2003–2019

km) dataset provides an updated characterization of global climate zones based on an ensemble of climatological datasets.

DEM: We used the 90 m Multi-Error-Removed Improved-Terrain Digital Elevation Model (MERIT DEM) dataset from Yamazaki et al. (2017) (Fig. S5). This high-accuracy global digital elevation model has been processed to remove multiple error components, resulting in improved landscape representation.

Vegetation dynamics: In order to meet our high-resolution needs, we included the revised 1 km resolution leaf area index (LAI) dataset from Lin et al. (2023) (Fig. S6). This dataset, derived from MODIS imagery, incorporates quality control information to improve its representation of vegetation dynamics. Due to the limited availability of remote sensing data before 2000, we calculated a single mean LAI value for each day of the year for each pixel by averaging MODIS LAI data from 2000–2021. This static LAI representation was then used for the entire study period (1950–2022). This single climatological mean was then used as a static representation of vegetation for all years in our evapotranspiration model. Importantly, this assumption will introduce some limitations that we will discuss in Section 5.4.

To ensure consistency with the target ET dataset, we resampled all ancillary datasets to a common 0.1° spatial resolution. The incorporation of these diverse ancillary datasets enables our modeling framework to capture the complex interactions between land surface properties and ET processes across global scales.

Since ET changes are influenced by specific humidity, radiation, temperature, wind speed, precipitation, and more (Kalma et al., 2008; Mueller et al., 2011; Wang and Dickinson, 2012), understanding these interactions is crucial. All of these factors are related to surface properties such as topography, land cover, climate zone, soil moisture, and vegetation distribution. For example, ET is impacted by altitude gradients and meteorological variables, with ET decreasing significantly at higher elevations (Yang et al., 2019). Studies have also demonstrated that estimates of soil moisture constraints on ET are more accurate in semiarid regions (Purdy et al., 2018). The use of machine learning models can be effective in assessing the significance of all parameters in the model to improve the merging process. For example, the incorporation of auxiliary variables in the merging process of the DNN-MET model improves the spatial correlation between ET from various sources and yields more accurate spatial ET predictions (Shang et al., 2021).

2.4. In-situ data

To train and validate our ET modeling framework and product, we collected in-situ measurements from 462 EC flux tower sites worldwide (Fig. 1). These sites span a wide range of climatic and ecological conditions, providing a robust basis for model training and validation. The EC technique is a widely accepted method for measuring water and energy fluxes at the ecosystem scale (Mueller et al., 2011). The flux tower data were obtained from multiple sources, including FLUXNET, AmeriFlux, European Fluxes Database Cluster, the National Tibetan Plateau/Third Pole Environment Data Center, and the Large-Scale Biosphere-Atmosphere Experiment in Amazonia (LBA). We applied a rigorous quality control procedure to ensure measurement data quality: (1) anomalous sites and data points were identified and removed; (2) the days with missing values of the collected half-hour data exceeding 25 % per day and the flux data with an EC energy closure ratio of <0.8 were excluded. For sites with available data (i.e., net radiation, sensible heat, latent heat, and ground heat are available), we adjusted ET by forcing energy balance using the Bowen ratio closure method; (3) daily ET values were calculated from the half-hourly data, and outliers over 8 mm/day were removed; (4) sites with insufficient data points (<120 daily values) were excluded. The resulting flux tower dataset spans a wide range of ecosystem types, with data records ranging from 1 to 21 years in duration. The represented ecosystem land cover types include Barren (BSV), Croplands (CRO), Closed Shrublands (CSH), Cropland Vegetation Mosaics (CVM), Savannas (SAV), Deciduous Broadleaf Forests (DBF), Grasslands (GRA), Deciduous Needleleaf Forests (DNF), Evergreen Broadleaf Forests (EBF), Mixed Forests (MF), Evergreen Needleleaf Forests (ENF), Open Shrublands (OSH), Permanent Wetlands (WET), and Woody Savannas (WSA). The detailed description of selected sites is shown in Table S1.

To further enhance the representation of the flux tower footprint, for each EC Flux site, we calculated the minimal distance between the gridded cell and the flux tower to better represent the flux tower footprint (Jung et al., 2019; Koppa et al., 2022; Lu et al., 2021; Xie et al., 2022). Although empirical machine learning models have proven effective for upscaling site-scale EC flux measurements (Chen et al., 2010; Lu and Zhuang, 2010; McNicol et al., 2023; Zhang et al., 2022; Zhu et al., 2024), their accuracy remains unclear due to misalignments between the footprints of different data sources used in ET estimation (Barcza et al., 2009; Chu et al., 2021; Ma et al., 2020b). These factors can lead to increased uncertainty, especially in regions with large heterogeneity (Kalma et al., 2008).

2.5. Basin water balance-based (ET_{wb}) data

Long-term basin-scale water-balance evapotranspiration data (ET_{wb}) can provide valuable independent validation to complement flux tower measurements (Hirschi and Seneviratne, 2017; Liu et al., 2016; Ma et al., 2024; Ruhoff et al., 2022; Senay et al., 2011; Xiong et al., 2023). To evaluate our AutoML model's performance at larger spatial and temporal scales, we utilized ET_{wb} data from (Ma et al., 2024), covering 56 large river basins ($> 10^5 km^2$) from 1983 to 2016. This dataset overcomes the shortcomings of eddy covariance flux measurements and is particularly suitable for assessing large-scale and long-term ET models. The ETwb time series from Ma et al. (2024) was optimally merged using a Bayesian-based three-cornered hat method, providing a robust benchmark for our model evaluation. The detailed information of

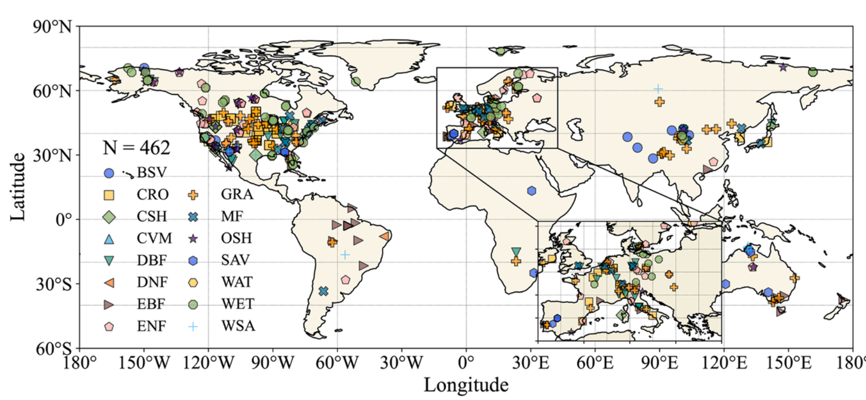

Fig. 1. The spatial distribution of 462 selected in-situ flux EC sites worldwide. See Section 2.4 for definitions of the abbreviations shown in the figure.

selected basins is shown in Table S9.

3. Methodology

We aim to generate high-resolution (1-day, 9 km) and long-term (1950–2022) global ET data by fusing multimodal high-quality ET data. Our workflow consists of two parts, including multisource ET data reconstruction and multimodal ET data fusion. The analytical framework is shown in Fig. 2 and described in detail as follows.

3.1. Multisource data reconstruction

The multisource ET datasets used in this study have varying spatial and temporal resolutions and coverage periods (Table 1). To effectively fuse these datasets, we first need to reconcile their disparate properties. We achieved this by using an ML approach based on the Light Gradient Boosting Machine (LightGBM) model (Ke et al., 2017) to reconstruct each dataset to a common spatial resolution (i.e., 0.1°), temporal resolution (i.e., daily), and time range (i.e., 1950–2022).

The LightGBM model is a gradient-boosting decision tree (GBDT) algorithm that can handle large-scale data and offers better accuracy, faster training speed, and lower memory usage compared to traditional GBDT models (Ke et al., 2017). Meteorological forcing from ERA5-Land, soil moisture, land cover types, climate zones, DEM, LAI, soil hydraulic parameters and soil properties were used as predictors for LightGBM models to estimate ET. We utilized the climatological mean of LAI data in 2000–2021 for each year. Each ET dataset (Table 1) except for ERA5-Land was used as the learning target, and therefore, 12 LightGBM models were established.

We first resampled the input features to the same spatial and temporal resolutions of the target ET data. Then, we performed two independent experiments to investigate the estimated accuracy of the developed LightGBM model (spatial and temporal cross-validation experiments). In the temporal cross-validation experiment, we trained the model with data consisting of the first 80 % of the time series at all grids. We tested the temporal generalizability of models with the remaining 20 % of data at all grids. In the spatial cross-validation experiment, the data from randomly selected 80 % of the grids were used to train the LightGBM model, and the data from the remaining grids were used to test the spatial generalizability of models. Finally, we reconstructed each ET dataset by the developed LightGBM model at high resolutions (1-day, 9 km) and extended the time ranges to 1950–2022 (i.e., the resolutions and time ranges of ERA5-Land forcing data).

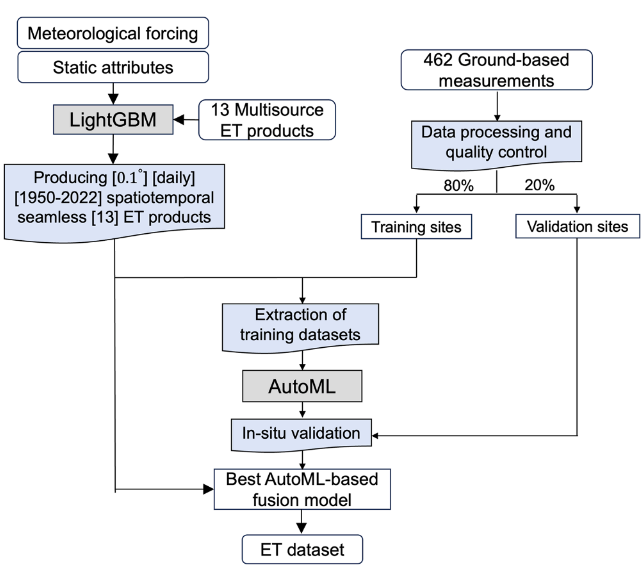

Fig. 2. Generation process for our ET product.

We set the same hyperparameters for LightGBM models trained on different ET data, and the hyperparameters are shown in Table S2. The reasons we did not tune the parameters of reconstruction models were offered as follows. (1) The training data for each LightGBM model is large enough to effectively avoid over-fitting problems. (2) We did the preliminary experiments with GLEAM v3.6a data and found that tuning the two vital hyperparameters of LightGBM (The number of estimators and leaves) did not bring much benefit, only increasing the mean KGE from 0.591 to 0.609, but significantly increased nearly three times computational cost. (3) We aimed to reconstruct data according to the main features of each ET data, rather than pursuing the absolute performance of models.

By applying the trained LightGBM models, we reconstructed each ET dataset to a common spatiotemporal resolution and time range, resulting in a set of harmonized datasets suitable for fusion. This approach preserves the unique characteristics of each dataset while enabling their integration into a unified framework. Notably, unlike some previous studies (e.g., Bai et al., 2022), we did not designate a single ET dataset as the reference for calibrating the others. Instead, we reconstructed each dataset independently, allowing us to capture and leverage the diverse information content of the individual datasets in the subsequent fusion process.

3.2. Multimodal data fusion by AutoML

After reconstructing the multimodal ET datasets to a common spatiotemporal resolution and time range, we employed the AutoML approach to fuse the datasets and generate an enhanced ET product (Fig. 2). AutoML is a software framework that could automate the complex process of building ML models (Hutter et al., 2019), such as feature engineering, ML model selection, and hyperparameters tuning. In our study, we investigated the advantage of AutoML in multimodal data fusion. We utilized FLAML, a lightweight Python library for AutoML tasks, which outperforms other libraries with significantly lower computational costs (Wang et al., 2021).

The 13 reconstructed ET data (including ERA5-Land), 8 meteorological forcing and 5 ancillary data were used as predictors (Table 1) to train the AutoML-based fusion model. The training target was derived from the 462 flux tower sites (Fig. 1, Table S1). We extracted the dataset of predictors and targets at the observation locations and performed spatial and temporal cross-validation to evaluate the accuracy of the developed AutoML model. In the temporal cross-validation experiment, we used firstly 80 % of the data from the time series at all stations to train the AutoML model, and the remaining 20 % of data at all stations was used to validate the temporal generalizability of the fusing model. In the spatial cross-validation experiment, the data from randomly selected 80 % of the stations (Table S1, 372 stations) were used to train the AutoML model, and the data from the rest of the stations (Table S1, 90 stations) were used to test the accuracy of the AutoML model at the spatial scale. Notably, we ensure the validation test covers all continents and all ecosystem types to accurately test the predictability of the fused model. Finally, we fused multisource ET data by the trained AutoML model at high resolutions (1-day, 9 km) with the time ranges from 1950 to 2022 globally. We used the trained model in the spatial cross-validation experiment because the performance of upscaling ET site data relies on the spatial extrapolation ability of the ML model. Due to the low accuracy of atmospheric forcing and ET data and lack of flux tower data in water bodies, extremely arid and permafrost regions (Ma et al., 2020; Pan et al., 2020), we excluded the ET data at these poorly quality regions by applying the land mask derived from FLUXCOM dataset (Jung et al., 2019).

We configured the AutoML to minimize the RMSE metric by assigning a time budget of 3600 s. The "auto" scheme of the ML estimator models library includes tree-based approaches such as LightGBM, XGBoost (eXtreme Gradient Boosting, Chen and Guestrin, 2016), CatBoost (Categorical Boosting, Ostroumova et al., 2017), RF (Random

Forst, Breiman, 2001), and Extra-Trees (Extremely randomized trees, Geurts et al., 2006). To avoid over-fitting, we determined the optimal value of hyperparameters in the AutoML model based on 5-fold cross-validation. We also configured the AutoML to "ensemble", i.e., find the best combination of models and corresponding hyper-parameters rather than only single ML models.

Due to limited training data and potential input dataset quality, the accuracy of the fused ET product may be lower before 2000. It is important to note that for ET products with limited temporal coverage (e.g., those starting from 2000), the reconstruction to earlier periods (1950–1999) relies primarily on relationships established between ET and ERA5-Land meteorological variables during the available data period. We assume the quantitative relationships between meteorological drivers and ET patterns identified in the post-2000 period remain valid for earlier decades. Moreover, the quality of ERA5-Land forcing data, particularly precipitation, is sufficient for historical ET estimation, and the changes in other factors affecting ET, such as CO_2 fertilization effects and land use change, should not substantially alter these relationships. Therefore, the reliability of our reconstructed ET is directly linked to the accuracy of the ERA5-Land precipitation data outside of the specified period. However, by incorporating multiple data sources and leveraging the advanced AutoML techniques, we expect our fused product to still provide an improvement over existing ET estimates for this period.

3.3. Validation

The modeling performance and quality of the ET product were evaluated using four metrics: Root Mean Squared Error (RMSE), normalized Root Mean Squared Error (nRMSE), correlation coefficient (R), and Kling-Gupta Efficiency (KGE). RMSE was applied for quantifying the model's ability of capturing the magnitude of variability in ET, including seasonal fluctuations (Kenney and Keeping, 1962). nRMSE is a normalized version of RMSE that is used to normalize the error for comparison between different data sets or variables (Stephen and Kazemi, 2014). R measures the correlation between simulations and observations (Pearson, 1920). KGE is a statistical measure that evaluates the performance of hydrological models by combining three statistical metrics: correlation coefficient, variability ratio, and bias ratio into a single measure. The KGE metric was computed as follows:

$$\mathrm{KGE} = 1 - \sqrt{(R-1)^2 + \left(\frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(S_i-\overline{S})^2}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(O_i-\overline{O})^2}} - 1\right)^2 + \left(\frac{\frac{1}{n}\sum_{i=1}^{n}S_i}{\frac{1}{n}\sum_{i=1}^{n}O_i} - 1\right)^2} \quad (1)$$

where S_i and O_i are the ith estimates and observation of ET. $\overline{S}$ and $\overline{O}$ are the mean value of the estimates and observation of ET, respectively. N denotes the number of samples.

3.4. Interpretation

To interpret the model, we applied Shapley values (Erik and Kononenko, 2014; Shapley, 1953) to quantify the contribution of each input variable to the predictions. Shapley values provide model-agnostic insights based on game theory to distribute the effect of explanatory variables. Specifically, the Shapley value for an input variable represents its average marginal contribution to the prediction across all possible variable combinations. We computed the Shapley values using the input variables from the 372-site training dataset. The Shapley values were aggregated across the site training dataset to gain an interpretable summary of how much each input variable influences the predicted outputs.

4. Results

4.1. Evaluation of the fusion model

To assess the performance of our AutoML-based fusion model, we conducted spatial and temporal cross-validation experiments using the in-situ ET measurements from the 462 flux tower sites. These experiments aimed to evaluate the model's generality to capture the variability of ET across different locations and periods. The spatial cross-validation results (Fig. 3a, i.e., performance in all periods at selected 90 validation sites) show that the fusion model accurately captures the spatial patterns of ET, with a high KGE of 0.778, a low RMSE of 0.897 mm/day and nRMSE of 0.082. The temporal cross-validation results (Fig. 3b, i.e., performance in validation periods at all 198 sites) demonstrate the model's ability to reproduce the temporal dynamics of ET, achieving a KGE of 0.822, an RMSE of 0.801 mm/day and nRMSE of 0.099. The result of the spatial cross-validation experiment was inferior to that of the temporal cross-validation experiment, which was consistent with previous studies (Ploton et al., 2020). Generally, these findings indicate that the fusion model can reliably estimate ET at locations and times not included in the training dataset.

4.2. Comparison with existing ET products

We then compared our fused ET product with 12 existing gridded ET datasets (except for the 0.25° resolution of FLUXCOM data) at daily and monthly timescales at all 462 sites. Fig. 4 presents the scatter plots of the different ET products against the in-situ measurements at the daily timescale. Our fused product exhibits the highest accuracy and precision, with a KGE of 0.857 and an RMSE (nRMSE) of 0.726 (0.062) mm/day. Notably, the fused product effectively corrects the systematic underestimation bias evident in most of the existing datasets, as indicated by the points being more concentrated around the 1:1 line, reaching the largest slope value (0.809) and the lowest intercept value (0.317) among all ET datasets, demonstrating the fused data could significantly correct this systematic bias against ground-measured ET data. At the monthly timescale (Fig. S7), the performance of all ET products improves compared to the daily timescale, and our fused product still outperforms the others, achieving a KGE of 0.915, an RMSE (nRMSE) of 0.524 (0.065) mm/day.

The Taylor diagrams in Fig. 5 further illustrate the superior performance of our fused product, showing the lowest RMSE and highest correlation and KGE values among all datasets at both daily and monthly timescales at all 462 stations. Our data has a very low RMSE and nRMSE, nearly 0, and both R and KGE are the largest. This indicates that our data is very similar to the station-observed data and fluctuates the least among all datasets considered for both daily and monthly temporal scales. Moreover, the proposed ET data is significantly improved through the fusion processes; for example, the R of benchmarking datasets ranges from 0.6–0.8, while our data achieved nearly 0.95. The rank of different data in terms of different criteria values of other data is nearly the same; specifically, Ours > FLUXCOM > REA > ERA5-Land > others. We also showed the Taylor diagrams of spatial cross-validation experiments (verified at all periods at 90 validation stations) (Fig. S8) and demonstrated the superiority of fused data compared other data at both daily and monthly scale, further reinforcing the result of Fig. 5.

The time series of observations, our data and FLUXCOM at selected six stations over different continents (AU-Otw for Oceania, BR-Sp1 for South America, CN-Qia for Asia, DE-Hai for Europe, CA-SF3 for North America, SD-Dem for Africa, Table S3) are shown in Fig. S9. The fused product yielded comparable or better performance than FLUXCOM at all stations and provided nearly the same temporal variation with observations. Moreover, FLUXCOM data could provide accurate estimates (except for Amazon sites) of ET magnitude but provides a smoother variation that cannot capture the small fluctuation of ET.

We evaluated the performance of the ET products in the Amazon

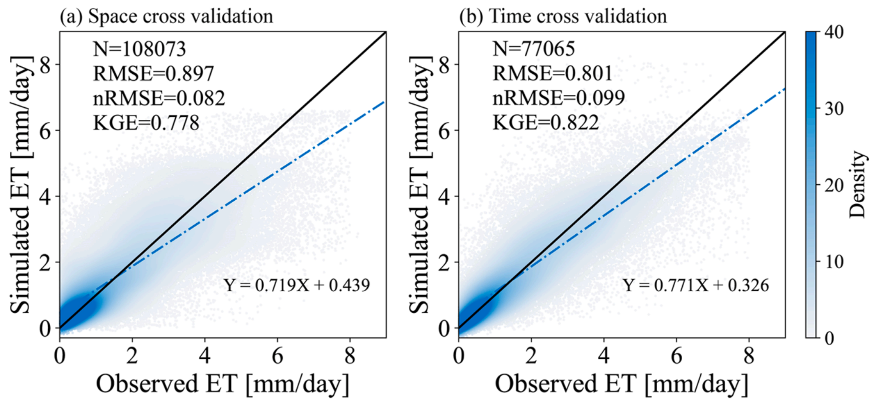

Fig. 3. The scatter plot between our ET data and in-situ ET data in the spatial cross-validation experiment (a, i.e., performance in all periods at 90 validation sites) and temporal cross-validation experiment (b, i.e., performance in validation periods at all 198 sites). N denotes the validation samples in each experiment.

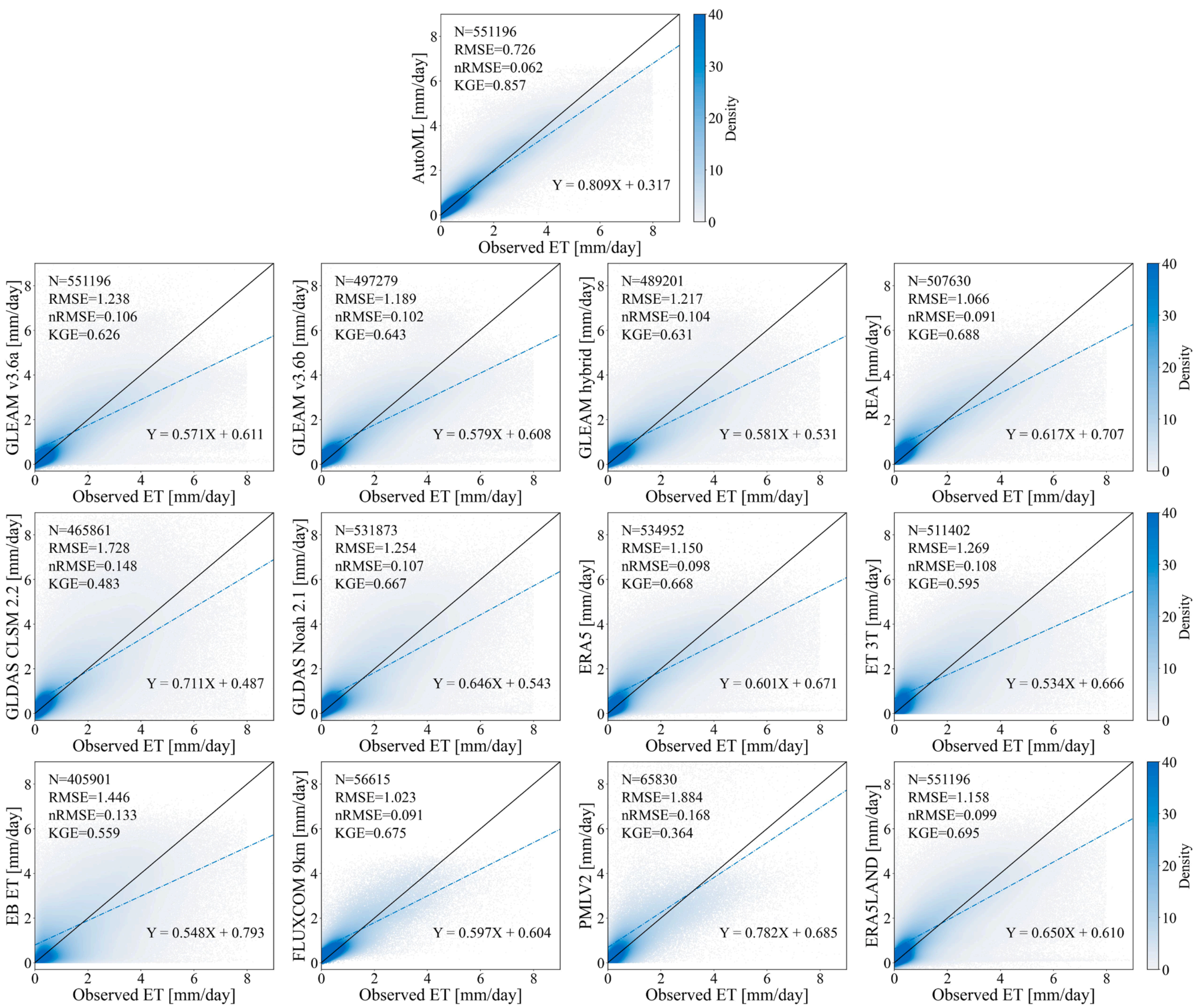

Fig. 4. Scatterplots of daily ground-measured ET and ET from different products at all 462 stations. Linear fits are plotted in blue, and the 1:1 line is depicted. N denotes the validation samples in each experiment.

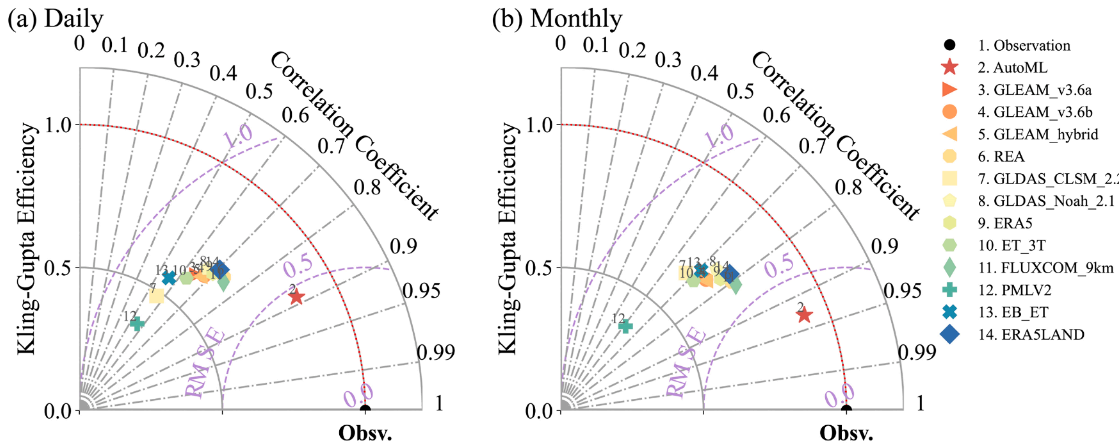

Fig. 5. Taylor diagrams of daily (a) and monthly (b) ground-measurement ET and ET from different products verified by all available validation samples from 462 station-based observations.

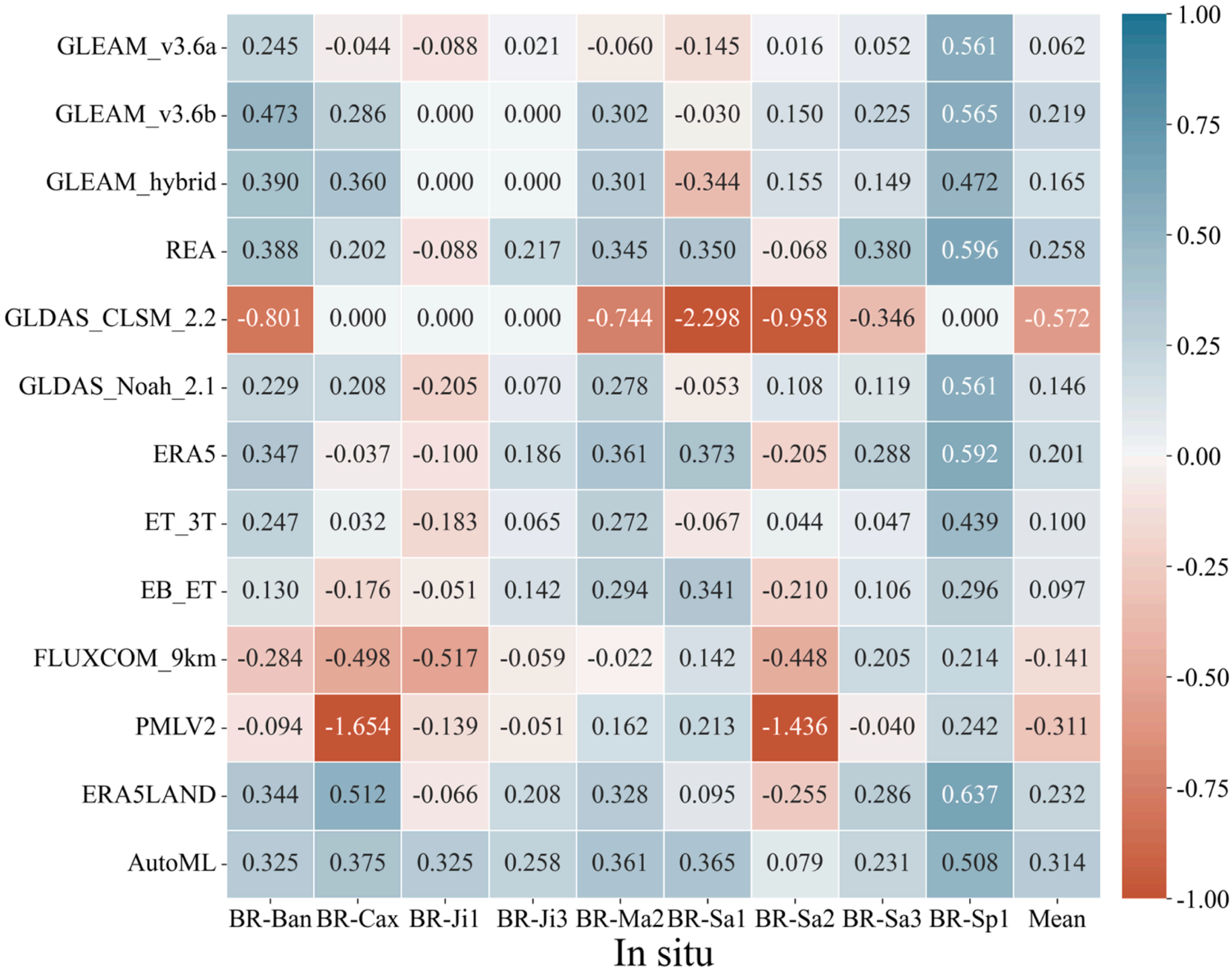

Fig. 6. The KGE value between different ET products with in-situ observations in the Amazon region. The last column shows the mean values of KGE over nine sites.

region specifically, which is known to be challenging for ET estimation. Fig. 6 shows the KGE values of the different products at nine flux tower sites in the Amazon. Our fused product achieved the highest average KGE of 0.314, outperforming the other datasets. Moreover, the fused product maintained positive KGE values at all nine sites, while all other benchmarking datasets did not, showing the robustness and reliability of our data in this critical region.

4.3. Performance across different ecosystem types

To assess the performance of the ET products across different ecosystems, we grouped the flux tower sites by their respective ecosystem types and calculated the evaluation metrics for each group. Table 2 summarizes the results for four selected advanced ET products (FLUXCOM, REA, ERA5-Land, GLEAM v3.6b) and our fused product at the daily timescale as verified by in-situ evaluation. We used fusion models to analyze data from various ecosystem types. However, some ecosystem types had fewer study sites, which could lead to varied climatic and land surface conditions. This may have caused some vegetation types, such as EBF and BSV, to have poor performance. Additionally, none of the four benchmarking data sets consistently outperformed the others across all ecosystem types. For example, although the average performance of FLUXCOM is better than the other three products (Fig. 5), the KGE of FLUXCOM is larger than the other three products only in 4 of 12 ecosystem types (DBF, EBF, MF, OSH). Interestingly, the results demonstrate that our ET products performed optimally across all ecosystems with R, KGE values increased by 5 %−31.3 %, and 17.6 %−77.8 %, respectively, RMSE and nRMSE values reduced by 0.154–0.421 mm/day, 23 %−42.6 %, separately. These results highlight the importance of developing a comprehensive fusion product that can accurately estimate ET across diverse ecosystems. The performance of different products in different ecosystem types on monthly scale (Table S4**)** further confirmed our conclusion. We also showed the performance of ET products across different ecosystems at 90 validation sites in the spatial cross-validation experiment. While the accuracy of the fusion product degraded compared to validation at all sites, it still outperformed the other three products at 9 out of 13 ecosystems and showed the best average performance. It only slightly underperformed compared to other products at four ecosystems with few validation sites (2, 4, 3, 2 sites for BSV, EBF, SAV, and WSA).

4.4. Spatial and temporal variability of global ET

Fig. 7 presents the spatial distribution of mean annual ET from five selected products spanning 2004–2015 (i.e., a shared time range of selected data). All datasets show generally consistent patterns. Regions with high ET are mainly located near the Equator, in regions with high rainfall, such as the Amazon regions, the Congo Basin, and Southeast Asia, where the yearly rainfall usually exceeds 1000 mm. Conversely, extremely low ET is found in deserts, such as the Sahara, Arabian, and Taklamakan deserts, as well as in permafrost regions in Eurasia and the north of North America. Compared to our dataset, the ET volume of GLEAM v3.6b and FLUXCOM 9 km datasets are notably higher in very wet regions near the Equator, particularly in the Amazon region. However, in arid regions like permafrost regions, these two datasets are both slightly lower than others. Moreover, REA data are found to be higher in high-latitude regions compared to the other four datasets. The fused data tends to provide relatively high value in wet regions like REA while giving relatively low value in arid and permafrost regions like FLUXCOM.

Fig. S10 shows the multi-year spatial average of the existing ET product and the fused one. General consistency in long-term trends is shown, but a large difference appears in the magnitude of products, particularly in summer. The discrepancy of mean ET in July among these datasets is nearly 100 %, ranging from 1.28 mm/day to 2.51 mm/day, indicating the high uncertainty of multisource ET products. Interestingly, our product gives more average results across months, showing the effect of using ensemble ML models rather than a single ML model in AutoML.

Table 2
The verification results including R, RMSE (values in mm/day), nRMSE and KGE between daily ground-measured ET and four selected state-of-the-art ET products in different ecosystems. The value in bold indicates the highest quality.

LC	GLEAM_v3.6b				REA				FLUXCOM				ERA5LAND				AutoML			
	R	RMSE	nRMSE	KGE	R	RMSE	nRMSE	KGE	R	RMSE	nRMSE	KGE	R	RMSE	nRMSE	KGE	R	RMSE	nRMSE	KGE
BSV (11)	0.627	1.043	0.099	0.275	0.608	1.125	0.107	0.364	0.481	1.248	0.119	0.205	0.682	0.945	0.090	0.463	**0.85**	**0.677**	**0.064**	**0.823**
CRO (61)	0.694	1.306	0.132	0.517	0.721	1.230	0.124	0.665	0.746	1.150	0.142	0.637	0.656	1.372	0.138	0.634	**0.873**	**0.85**	**0.086**	**0.82**
CSH (12)	0.398	2.096	0.263	−0.02	0.510	1.207	0.151	0.491	0.619	0.998	0.140	0.596	0.660	1.023	0.128	0.658	**0.863**	**0.631**	**0.079**	**0.817**
DBF (45)	0.708	1.135	0.142	0.625	0.767	1.065	0.133	0.636	0.748	1.077	0.136	0.674	0.761	1.123	0.140	0.661	**0.904**	**0.688**	**0.086**	**0.893**
EBF (22)	0.605	1.548	0.194	0.537	0.629	1.183	0.147	0.590	0.659	1.147	0.148	0.617	0.576	1.322	0.164	0.565	**0.821**	**0.849**	**0.106**	**0.731**
ENF (91)	0.719	0.984	0.115	0.696	0.721	0.968	0.120	0.702	0.739	0.919	0.117	0.669	0.702	1.038	0.122	0.700	**0.857**	**0.709**	**0.083**	**0.827**
GRA (87)	0.769	0.986	0.118	0.646	0.774	0.953	0.114	0.717	0.753	0.979	0.121	0.657	0.749	1.025	0.123	0.714	**0.888**	**0.696**	**0.084**	**0.843**
MF (23)	0.611	1.207	0.107	0.574	0.715	1.086	0.093	0.616	0.777	0.898	0.080	0.664	0.685	1.239	0.106	0.570	**0.908**	**0.616**	**0.053**	**0.878**
OSH (31)	0.433	1.023	0.097	0.433	0.566	0.807	0.077	0.534	0.672	0.617	0.077	0.549	0.490	0.965	0.092	0.486	**0.882**	**0.463**	**0.044**	**0.8**
SAV (14)	0.812	1.140	0.143	0.581	0.746	1.138	0.142	0.606	0.744	1.116	0.142	0.625	0.765	1.150	0.144	0.696	**0.905**	**0.716**	**0.09**	**0.899**
WET (51)	0.577	1.579	0.192	0.495	0.670	1.337	0.162	0.571	0.712	1.306	0.164	0.506	0.608	1.539	0.187	0.569	**0.882**	**0.885**	**0.107**	**0.799**
WSA (11)	0.868	0.976	0.123	0.655	0.829	0.980	0.124	0.701	0.834	0.937	0.119	0.677	0.813	1.057	0.134	0.666	**0.908**	**0.725**	**0.092**	**0.896**

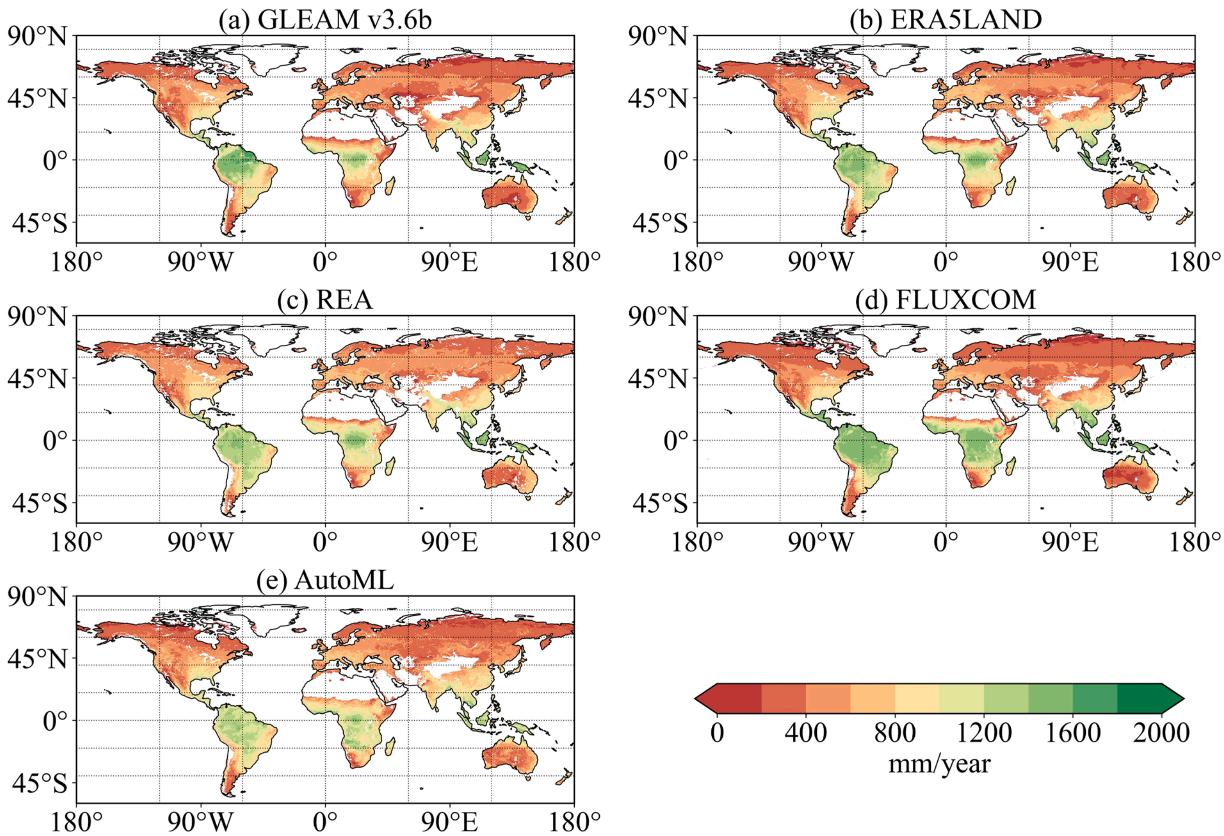

Fig. 7. Spatial distribution of annual mean ET for 2004–2015 (the shared time ranges of selected data).

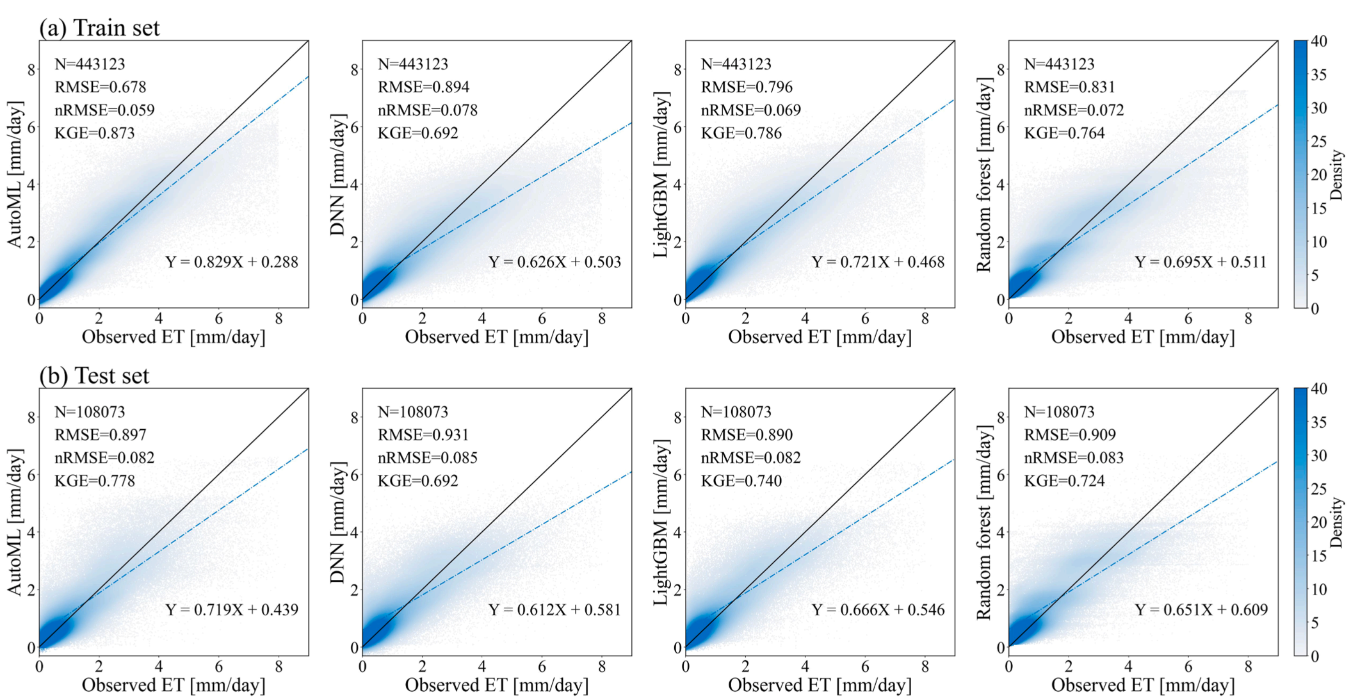

Fig. 8. Scatterplots of daily ground-measured ET and ET from different ensemble Machine Learning methods in the spatial cross-validation experiment (a and b represent the performance in the 372 training stations and the 90 validation stations, respectively). Linear fits are plotted in blue, and the 1:1 line is depicted. N denotes the validation samples in each experiment.

5. Discussion

5.1. The benefits of AutoML-assisted fusion model

Merging multiple ET datasets using ML algorithms improves the accuracy and robustness of ET estimates by mitigating the inherent biases and limitations of individual models (Parrish et al., 2012; Zhu et al., 2016), such as DNN (Xie et al., 2022), RF (Shang et al., 2021), and LightGBM (Fan et al., 2019). These ML models have shown great success in out-of-sample prediction (Ball et al., 2017), high accuracy of different land cover types (Xie et al., 2022), and relatively high computational efficiency (Shang et al., 2020). However, their performance largely depends on their parameter tuning (Chollet, 2017; Hinton and Salakhutdinov 2006), which needs lots of expert knowledge and computational cost. For DNN, we adopted the same parameters setting with the DNN model used in an advanced fusion ET product (GLASS v5.0, Xie et al., 2022), which is proven to be more efficient than BMA methods used in GLASS v4.0 (Yao et al., 2014). For LightGBM and RF, we tuned the important hyperparameters by grid search methods, and the detailed description is shown in Table S6.

It is encouraging to note that even without any human intervention, the AutoML-assisted model generally outperforms the aforementioned manually tuned ML models (i.e., RF, DNN, LightGBM) as verified by the evaluation on 462 flux tower sites in terms of all statistical metrics (Fig. 8). AutoML model showed the best performance among 46.7 % (42/90) of test sites and 82.3 % (306/372) of train sites (Table S7), and the KGE of the training and validation set for spatial cross-validation reached 0.873 and 0.778, respectively. Additionally, the RMSE reached 0.678mm/day and 0.897mm/day and the nRMSE reached 0.059 and 0.082, respectively (Fig. 8). The rank of different models in terms of different criteria values is nearly the same, specifically, AutoML > LightGBM > RF > DNN (Fig. 8, Table S7). Although RF methods have demonstrated superiority in reducing variance and preventing overfitting to sets of multiple subtrees (Elith et al., 2008), stiffness occurred during site prediction (Fig. S11), while other models did not present this abnormal prediction (figure omitted). LightGBM models utilize data examples with large gradients to estimate information gain, and propose a greedy algorithm to scan the approximate ratio, effectively reducing the dimensionality of features and ensuring the accuracy of equinox determination (Shang et al., 2023), thereby LightGBM showed promising results, providing greater accuracy than RF models, which is consistent with previous studies (Fan et al., 2019; Machado et al., 2019; Thongthammachart et al., 2022). Interestingly, DNN showed the same predictability between train and validate datasets, highly demonstrating generalization ability, but provided the poorest performance among all ML models.

To sum up, AutoML-assisted fusion models bring two obvious advantages over other ML-based fusion models. (1) It frees the data scientists from heavy and time-cost manual tunning and engineering processes in complex ML pipelines. (2) It can more accurately learn complex nonlinear relationships to improve land ET estimation based on the advantages of ensemble learning and automatic pipelines. Despite the significant benefits of AutoML, only a few studies used it to solve problems in meteorology fields, e.g., for PM2.5 estimation (Zheng et al., 2023) and soil hydraulic parameters mapping (Chen et al., 2023). We recall that AutoML is a welcome and widely used tool for ML applications in meteorology fields, particularly for fusion processes.

5.2. The effectiveness of data reconstruction

Integrating multiple ET datasets with varying spatial and temporal resolutions and coverage periods is a challenging task. Inconsistencies among the datasets can lead to significant uncertainties in the fused product (Bai et al., 2022; Chen et al., 2021; Yuan, 2020). This uncertainty can be reduced by integrating multiple satellite-based products (Jung et al., 2010; Mueller et al., 2013) and employing more complex merging methods (Yao et al., 2014). For instance, the GLASS v5.0 ET product amalgamates five traditional satellite ET products (MOD16, SW, PT-JPL, MS-PT, and SIM) (Xie et al., 2022), ensuring uniform spatiotemporal resolution across all integrated products. The REA method, on the other hand, merges three widely used land ET datasets: ERA5 (Hersbach et al., 2020), MERRA-2 (Gelaro et al., 2017), and GLDAS2 (Sheffield and Wood, 2007). Although the REA method combines a smaller number of ET products compared to the GLASS, it includes ET data at various spatial and temporal scales and adjusts these scales to a unified range. Currently, most methods for achieving unified spatiotemporal scales involve linear interpolation or utilizing data of the same spatiotemporal scale (Lu et al., 2021; Shang et al., 2021; Xie et al., 2022; Yao et al., 2017a). However, these methods could involve interpolation errors and limit the time ranges and spatiotemporal resolutions of fused products.

Our data reconstruction approach aims to harmonize the resolution to a fine scale (daily and 0.1°) and extend the time ranges of multisource ET data with persisting characteristics. It is encouraging to see that LightGBM models could handle huge amounts of data and efficiently learn complex nonlinear relationships existing in multisource ET data. Fig. 9 depicts the spatial mean of multisource ET data before and after reconstruction during 2012–2021. The reconstructed ET datasets show reduced interannual variability compared to the original data, which can be partially attributed to the use of climatological mean LAI (2000–2021) as a predictor. While this approach allows us to extend the dataset back to 1950, it may limit the model's ability to capture year-to-year variations in vegetation dynamics that influence ET. Fig. 10 shows the similarity of before and after reconstructed ET data defined by KGE in the validation set of the temporal cross-validation experiment. Obviously, the reconstructed data successfully reproduces the time variance and magnitude of the original data with finer spatiotemporal resolution and longer time ranges. The result of spatial cross-validation experiments also supports the effectiveness of the proposed methods (Table S8). All reconstructed models accurately capture the spatial patterns of original data, with the values of R all larger than 0.5, KGE all larger than 0.3, RMSE all <0.97mm/day and nRMSE all <0.1. Although the result of the spatial cross-validation experiment was inferior to that of the temporal cross-validation experiment, both results demonstrate the robustness of reconstructed models, and could reproduce the spatiotemporal variations of original data.

One significant concern of data fusion is the quality of original data, particularly for regions where estimates of ET are challenging due to the complex land-atmosphere feedback, notably in the Amazon region. Assessments conducted by Pan et al. (2020) suggested increased uncertainty across multiple ET products in the Amazon. To ascertain the validity of LightGBM refactoring in mitigating uncertainty in the region, we utilized nine Amazon site datasets to conduct a month-to-month comparison of data before and after the reconstruction process (Fig. 11). Our reconstruction method not only harmonizes the multisource ET datasets but also significantly reduces uncertainties and improves the overall quality of the data. Fig. 11 illustrates the disparity in KGE evaluation indicators before and after reconstruction, with at most 95.8 % improvement of KGE (GLDAS_CLSM). 8 of 11 datasets showed larger than 5 % improvement compared to original data, indicating a noticeable improvement effect on data in the region. This underscores the necessity for data reconstruction and the effectiveness of the proposed methods. By effectively addressing the inconsistencies among the multisource ET datasets, our LightGBM-based reconstruction approach lays a solid foundation for the subsequent fusion process. The reconstructed datasets enable the AutoML-based fusion model to better integrate the information from multiple sources and generate a global ET product with high spatiotemporal resolution and over long time periods.

5.3. Basin water balance-based estimates

The basin-scale validation results (Fig. 12) demonstrated that our

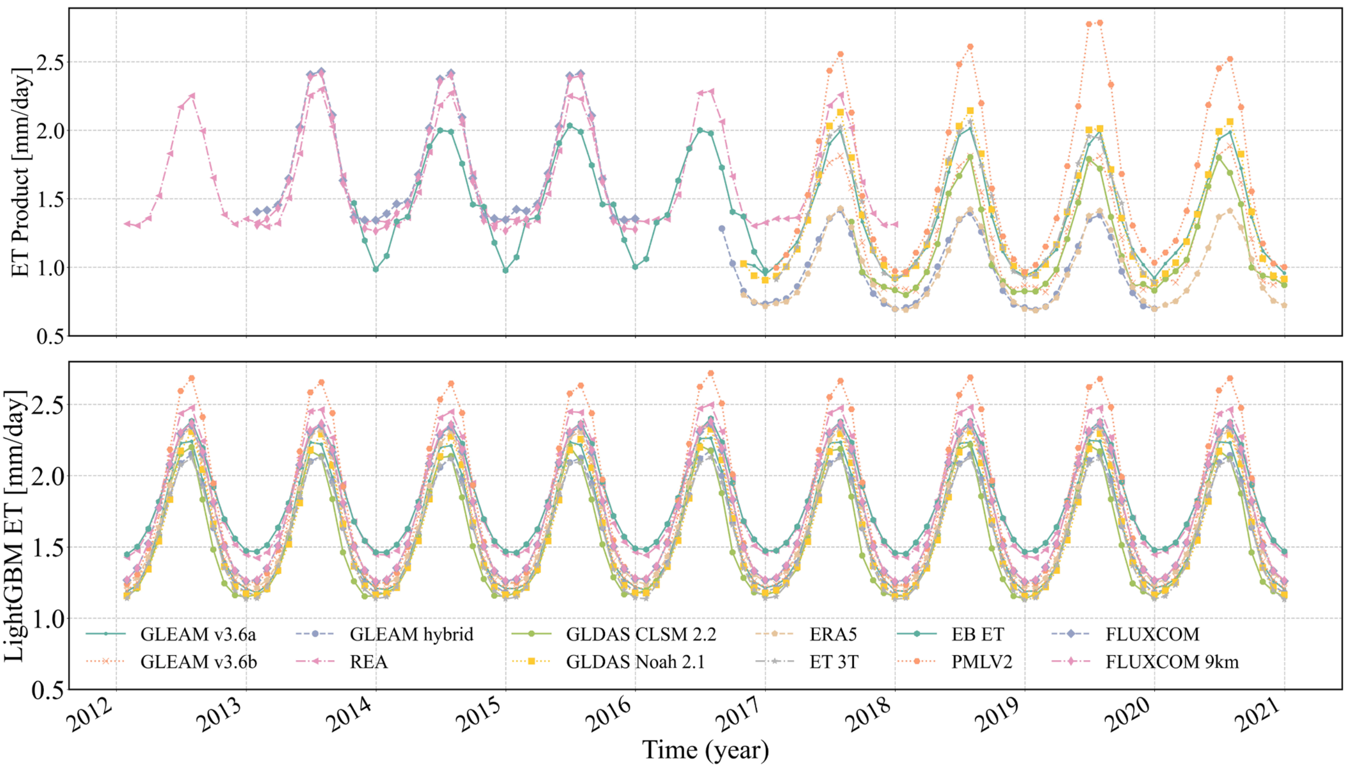

Fig. 9. The performance of multisource data reconstruction before the last 20 percent of the time range. (a) Original data for the last 20 % of each dataset's time range (used to test the accuracy of the LightGBM model); (b) is the ET data time series after the reconstruction of multi-source for the full period (2000–2021).

AutoML model-generated dataset exhibits lower bias compared to existing products. Our model achieved an RMSE of 116.42 mm/year and KGE of 0.853, and a percent bias (PBIAS) of −3.074 %, outperforming GLEAM 3.6a (RMSE: 147.48 mm/year, KGE: 0.817, PBIAS: −11.997 %). The significantly lower percent bias (PBIAS) of our AutoML model (−3.074 %) compared to GLEAM 3.6a (−11.997 %) further confirms its improved accuracy in capturing the long-term mean observed ET values across diverse river basins. These improvements are particularly noticeable in large basins, such as the Amazon. This is in line with our findings in Sections 4.2 and 5.2, where the LightGBM model reconstruction significantly reduced uncertainty in this region. It is observed that both AutoML and GLEAM v3.6a tend to overestimate ET in specific basins. GLEAM tends to underestimate ET in most basins where ET_{wb} exceeds 500 mm/year (Ma et al., 2024), especially in the Amazon region (Fig. 12). In contrast, AutoML often overestimates in drier areas like southern Africa, highlighting a limitation of machine learning with extreme values. The larger discrepancies in smaller basins are likely due to GRACE retrieval errors (Scanlon et al., 2018).

In summary, our basin-scale evaluation addresses the limitations of site-level validation. The fusion product demonstrates closer alignment with reference basin data, maintaining strong performance even in challenging prediction areas like the Amazon. Nevertheless, our product not only excels in flux tower site validation (Fig. 8) but also shows robust performance at the basin scale, providing a comprehensive validation of our approach.

While flux tower coverage is limited in some regions, particularly Africa, our fusion approach leverages physical relationships learned from similar ecosystems globally. The differences from existing products in these regions reflect our model's ability to correct known biases by integrating multiple data sources and ancillary information. Basin-scale validation results Sort the reliability of our estimates in these regions.

5.4. *The contribution of input variables*

Analyzing the Shapley values allowed us to provide a fair, quantitative interpretation of how the different explanatory variables interact and contribute to the ET predictions. Fig. 13 presents the SHAP values of the main input variables, ranked by their importance in the AutoML-based fusion model. A SHAP value of zero indicates that the feature provides little to no improvement in the model's performance (e.g., R^2). Positive SHAP values indicate contributions that increase the predicted ET compared to the expected (typically mean) prediction, while negative values indicate contributions that decrease the predicted ET. For example, higher values of influential ET datasets (ET_ERA5-Land, ET_ERA5, ET_PMLV2, ET_REA, and ET_FLUXCOM_9km) typically lead to higher predicted ET values, as shown by the red color in Fig. 13. The bimodal distribution of SHAP values for ERA5-Land ET, with both strong positive and negative values, suggests that this predictor's influence varies depending on conditions. It may have a strong positive influence in regions or periods where it performs well, while contributing negatively in areas where it tends to misestimate ET. Among all the input variables, $ET_{ERA5\text{-}Land}$, t2m, ET_{ERA5}, ET_{PMLV2}, ET_{REA}, theta_s, DEM and $ET_{FLUXCOM\_9km}$ are the eight most important features for model prediction. Among 12 ET products, $ET_{ERA5\text{-}Land}$, ET_{ERA5}, ET_{PMLV2}, ET_{REA}, and $ET_{FLUXCOM\_9km}$ are identified as the most influential, likely due to their high spatial and/or temporal resolutions, and well-known high accuracy. Other important input variables include air temperature, soil saturation, and DEM. The SHAP values also reveal the directionality of the input variables' contributions. For example, higher values of the influential ET datasets ($ET_{ERA5\text{-}Land}$, ET_{ERA5}, ET_{PMLV2}, ET_{REA}, and $ET_{FLUXCOM\_9km}$) generally contribute to higher predicted ET values, as indicated by the red color in Fig. 13. Similarly, air temperature shows a positive association with ET, except in drought-prone areas where high temperatures may not always lead to increased ET due to water limitations. The contributions of soil saturation and DEM are more complex, as their impacts on ET depend on other factors such as soil properties,

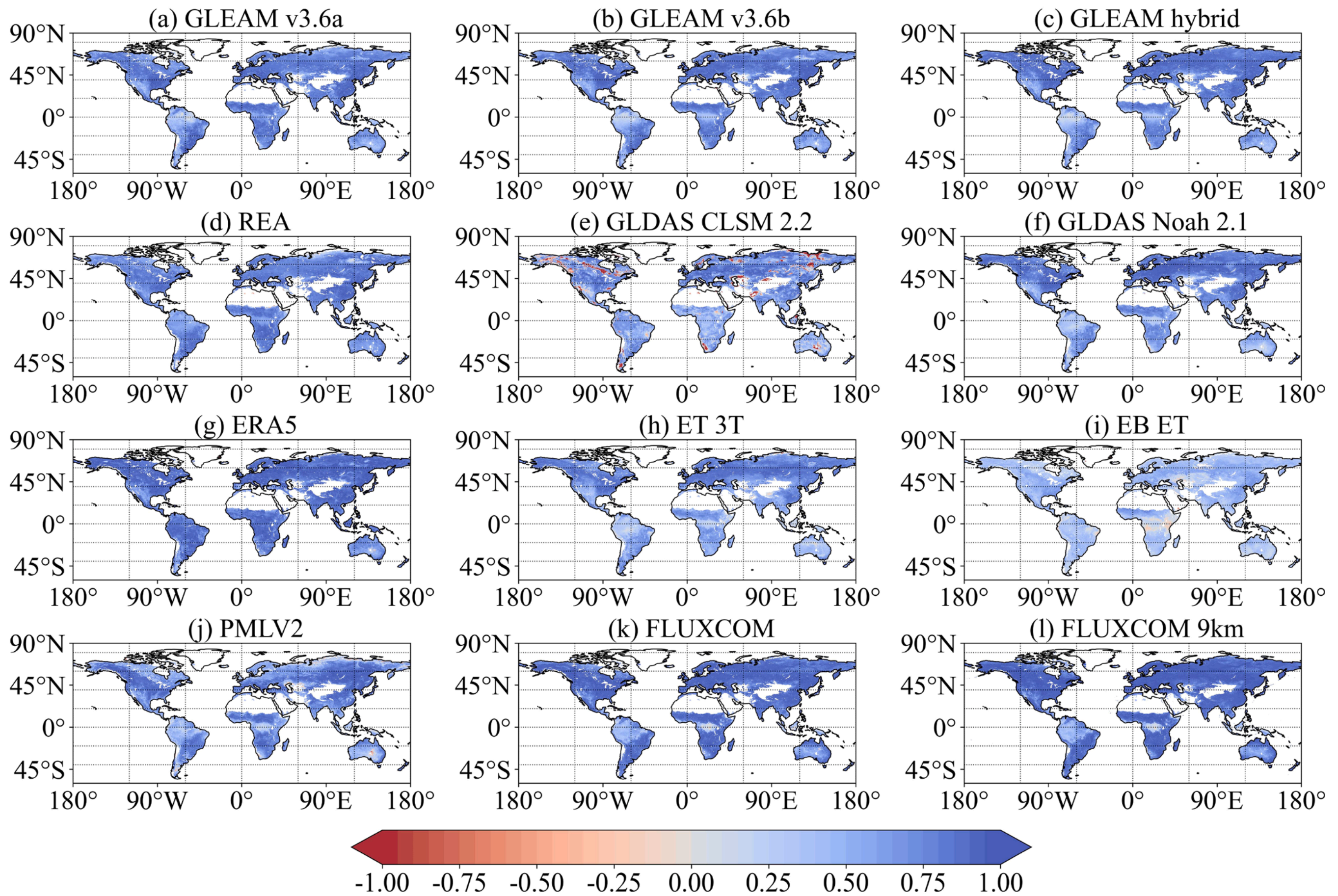

Fig. 10. The KGE value between before and after multisource data reconstruction in different ET products of temporal cross-validation experiments. All datasets are applied with a mask derived from FLUXCOM.

vegetation characteristics, and topographic effects on solar radiation and water redistribution. Furthermore, the LAI and other statistics are crucial for land surface processes and significantly impact evapotranspiration (Chen et al., 2015). Reprocessed statistical variables offer higher-resolution details that are often underrepresented in machine learning importance measures (Li et al., 2022). AutoML confirms this in SHAP analysis, reflecting the challenges of integrating these variables effectively. The SHAP analysis, in turn, captures these complex relationships and provides valuable insights into the model's behavior.

5.5. Potential applications and future work

Our ET dataset stands out due to its combination of fine spatial resolution, extended temporal coverage, and validated high accuracy, providing an unprecedented tool for hydrological and climatological research. This dataset will be continuously updated to provide open and easily accessible ET data. In general, it can greatly improve the accuracy of water resource assessments, inform drought monitoring systems, facilitate the improvement of climate models, and provide important insights into ecological conservation efforts. For example, the dataset's fine spatial resolution suits water management applications needing local-scale water balance calculations, overcoming limitations currently available (Jahromi et al., 2022). It also serves as a valuable benchmark for evaluating hydrological models, enabling assessments over longer timescales than existing coarse-resolution datasets used in model intercomparison projects like ILAMB (Collier et al., 2018). The dataset's prolonged 1950–2022 temporal coverage and validated accuracy make it ideal for studying long-term trends in land-atmosphere interactions, assessing the regional hydrological impacts of climate change. Specific potential analyses include assessing the regional hydrological impacts of climate change drivers, such as the tradeoff between the El Niño–La Niña cycle and continental evaporation (Miralles et al., 2014), and examining spatial patterns and variability in ET (Fleischmann et al., 2023). At local scales, the data can also elucidate how plant-level processes influence broader ET responses and ecosystem resilience to water availability changes over time (Knighton and Berghuijs, 2023). To clarify, the limited coverage of flux towers in certain regions, such as Africa and the Amazon, is evident, with only 4 sites in Africa and 9 in the Amazon. The scarcity of observation sites and their clustered distribution—often near research stations—results in weak constraints for modeling evapotranspiration. This lack of coverage is particularly critical for local-scale estimates, where the fusion method may introduce uncertainties.

While this study generates a global ET product at an unprecedented 0.1° spatial resolution, higher resolution data may be required for certain precision agriculture applications. For example, Martens et al. (2018) executed models at a 100 m resolution over select European regions using GLEAM's high-resolution simulation capability. Additionally, OpenET (Melton et al., 2022) provides 30 m resolution monthly ET over the continental United States by integrating multiple satellite-driven models. However, such ultra-high-resolution simulations are not yet available on a global scale. As part of future work, we aim to leverage our framework to produce a 1 km, hourly global ET dataset. This involves generating corresponding high spatiotemporal resolution atmospheric forcing by assimilating satellite data and machine learning. Additional observational constraints from data sources like solar-induced chlorophyll fluorescence (Bu et al., 2024) and atmospheric CO_2 measurements (Upton et al., 2023) can further inform

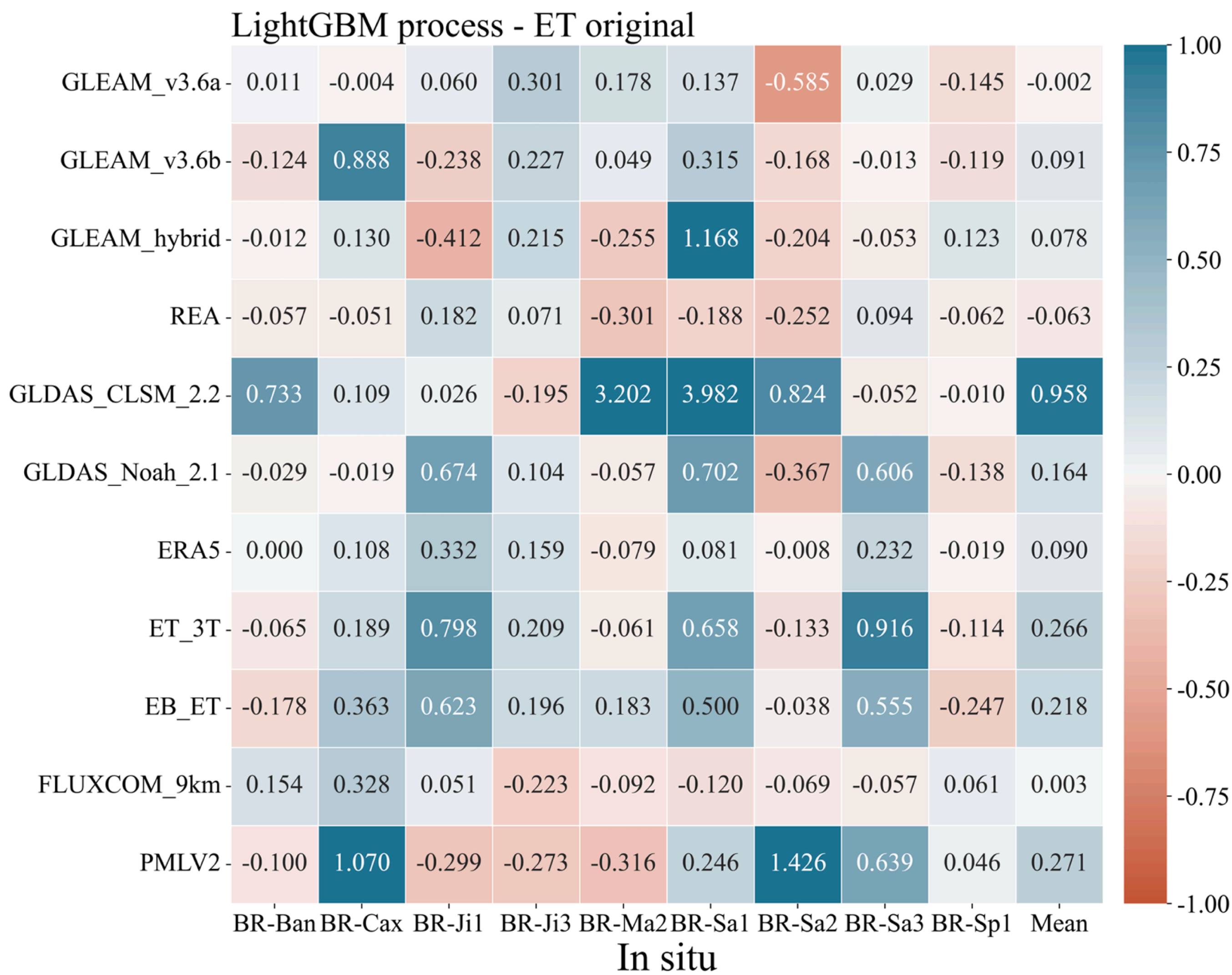

Fig. 11. The difference between the KGE values of different ET products and in-situ observation data and the KGE values of multi-source product reconstruction data and in-situ observation data. The last column shows the mean KGE values over 9 sites (the data are on a monthly scale).

process representations. Recently developed hybrid physically based machine learning techniques (ElGhawi et al., 2023; Reichstein et al., 2019; Zhao et al., 2019) and deep learning approaches (Camps-Valls et al., 2021; Giardina et al., 2023; Koppa et al., 2022) also hold promise for integrating theory and data. Overall, advancing ET simulations across scales requires coordinated efforts to improve models, reference datasets, and synergy between bottom-up and top-down strategies. It is also noted that we used the climatological land use and leaf area index dataset, which may not reflect changes in ET due to land use or shifts in vegetation driven by climate that have occurred. We intend to explore methods for integrating time-varying vegetation data, such as GIMMS LAI4g (Cao et al., 2023), or reconstructing historical vegetation patterns in our future work. This could potentially enhance the accuracy of our ET estimates, particularly for the earlier period of our dataset. It should be noted that the systematic evaluation of our product's ability to capture interannual variability, especially during extreme wet and dry conditions, is a key aspect for further study. While our current validation demonstrates good performance for mean conditions, the potential smoothing of extremes in our fusion approach requires careful assessment.

While our AutoML model has demonstrated improved ET performance, we recognize that significant uncertainties remain, partly due to this spatial scale mismatch (Barcza et al., 2009; Chu et al., 2021). In the next version of our ET product, the strategic selection of EC sites situated in regions characterized by relatively homogeneous land cover and topography, coupled with the implementation of a Plant Functional Type or Plant Community sub-grid scheme, and the execution of a comprehensive uncertainty analysis, will help us to reduce the potential errors introduced by this scale mismatch.

6. Conclusions

This study introduced a novel multimodal machine learning framework to generate a high-resolution (0.1°, daily), long-term (1950–2022) high-accuracy global terrestrial ET dataset by fusing 13 ET products from diverse sources. The framework overcomes the limitations of existing products in terms of coarse resolution, short-term coverage, and reliance on assumptions that ignore the complexity of the ET process. The framework leverages the strengths of multiple ET estimation methodologies by integrating multisource state-of-art ET encompassing remote sensing, machine learning, land surface models and reanalysis data, ET from in-situ observation (462 sites), ERA5-land atmospheric forcings and ancillary data, such as land cover type, DEM, soil property, climate zones, soil hydraulic parameters, and LAI. We use LightGBM models to reconstruct the individual input ET products to consistent spatiotemporal resolutions and time ranges. Subsequently, an Automated Machine Learning technique fused the reconstructed datasets in an optimal manner. Extensive flux tower observations were utilized for

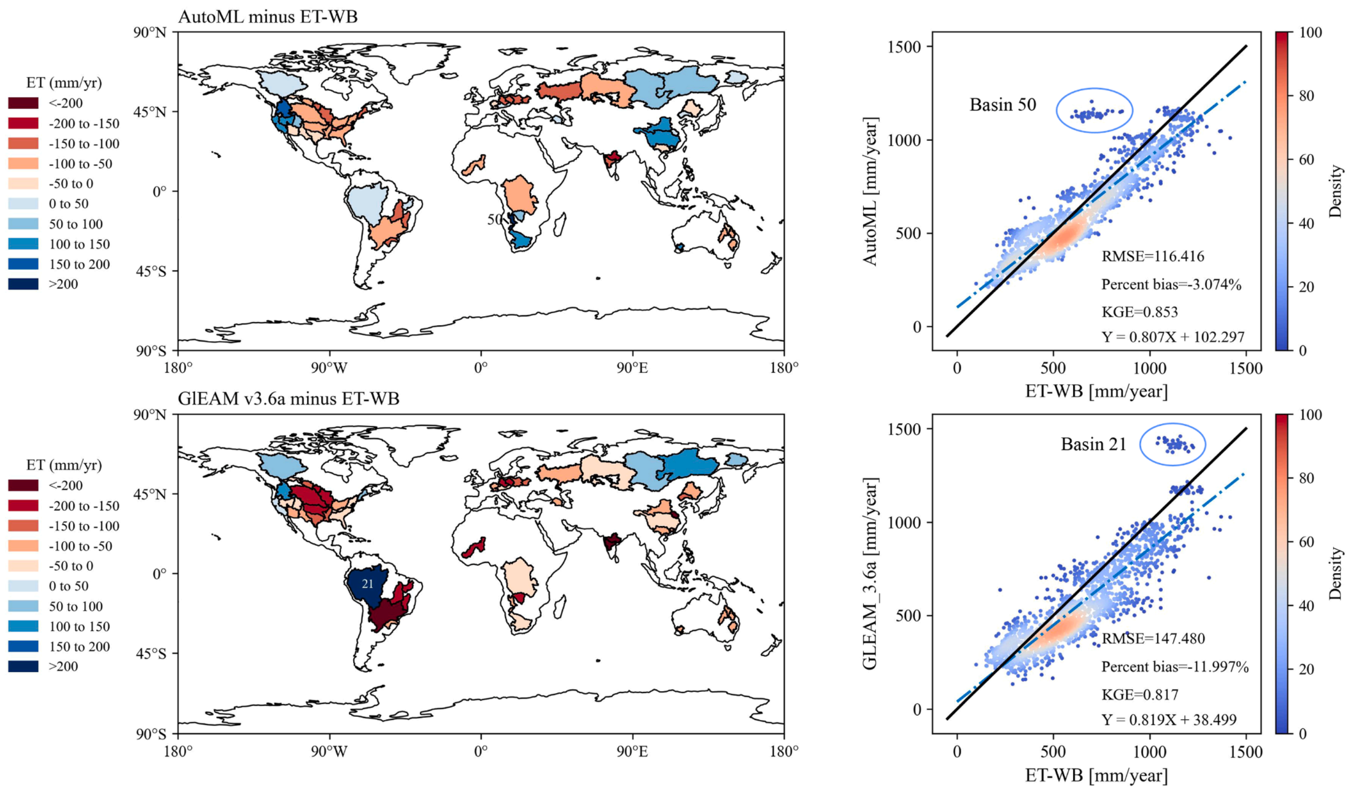

Fig. 12. Evaluation AutoML and GLEAM_v3.6a of the spatial distribution difference of the multiyear (1983–2016) mean BTCH-merged ET_{wb} values with, and the annual ET estimates against the optimally merged ET_{wb} values of the 56 basins for 1983–2016, the total sample size is 1904 (=56 × 34). Basin 21 and Basin50 represent Amazon and Okavango basin separately.

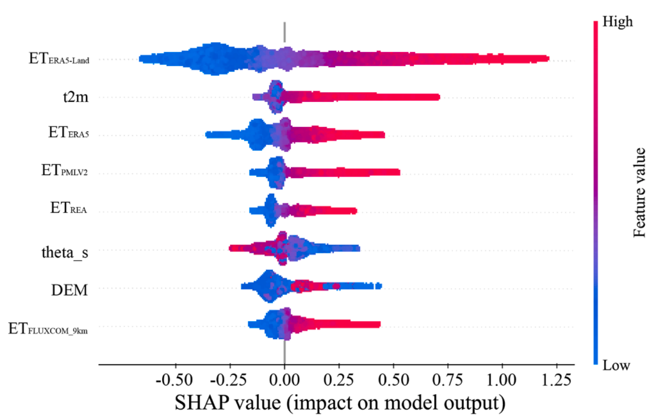

Fig. 13. The contribution (Shapely values) of the main explanatory input variables in site training.

training and multi-faceted validation. Results demonstrated that the fused ET product significantly outperforms widely used benchmark datasets across ecosystems and regions, including critical areas like the Amazon. It realistically captures fine-scale spatiotemporal variability in ET while correcting the prevalent underestimation bias in other products. Interpretability analysis confirmed that the data-driven model effectively captures the underlying physical processes driving ET. Overall, this high-fidelity global ET dataset provides an unprecedented resource to robustly assess climate change impacts on regional hydrology, agriculture, and ecology. The introduced spatiotemporal data fusion methodology also serves as a valuable framework for integrating multidimensional geoscience data. Future research could explore applying this approach to generate consistent other key variable datasets in the hydrometeorology field by fusing reanalysis, satellite, and station data.

Data availability

The CD12Q1 Version 6 land cover data is archived at the LP DAAC - MCD12Q1 (https://lpdaac.usgs.gov/products/mcd12c1v006/).

The 10 km SoilGrids dataset can be downloaded from www.soilgrids.org and/or ftp.soilgrids.org.

The 1 km soil hydraulic parameters data can be downloaded from https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/UI5LCE.

The climate region data are available via www.gloh2o.org/koppen. The 90 m high-accuracy global digital elevation model (DEM) dataset can be downloaded from http://hydro.iis.u-tokyo.ac.jp/~yamadai/MERIT_DEM/.

Some flux towers data were compiled from FLUXNET (https://fluxnet.org/data/fluxnet2015-dataset/), AmeriFlux (https://ameriflux.lbl.gov/), European Eddy Fluxes Database Cluster (http://www.europe-fluxdata.eu/). Part EC site data is provided by the National Tibetan Plateau / Third Pole Environment Data Center (http://data.tpdc.ac.cn). Amazon sites are downloaded from LBA-ECO CD-32 Flux Tower Network Data Compilation, Brazilian Amazon: 1999–2006, V2 (ornl.gov).

The water-balance-based evapotranspiration (ETwb) dataset is publicly available (https://data.tpdc.ac.cn/en/data/e010cd0d-0881-4e7e-9d63-d36992750b04).

Our data are provided in the following three Zenodo links:

Qingchen Xu, & Lu Li. (2024). Data for "A multimodal machine learning fused global 0.1° daily evapotranspiration dataset from 1950 to 2022" (1950–1974) [Data set]. Zenodo. 10.5281/zenodo.10906121

Qingchen Xu, & Lu Li. (2024). Data for "A multimodal machine learning fused global 0.1° daily evapotranspiration dataset from 1950 to 2022" (1975–1999) [Data set]. Zenodo. 10.5281/zenodo.10906126

Qingchen Xu, & Lu Li. (2024). Data for "A multimodal machine learning fused global 0.1° daily evapotranspiration dataset from 1950 to 2022″ (2000–2022) [Data set]. Zenodo. 10.5281/zenodo.10906128

CRediT authorship contribution statement

Qingchen Xu: Writing – review & editing, Writing – original draft, Visualization, Validation. **Lu Li:** Writing – review & editing, Writing – original draft, Validation, Project administration, Methodology, Conceptualization. **Zhongwang Wei:** Writing – review & editing, Writing – original draft, Project administration, Methodology. **Xingjie Lu:** Writing – review & editing. **Nan Wei:** Writing – review & editing. **Xuhui Lee:** Writing – review & editing. **Yongjiu Dai:** Writing – review & editing.

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.

Acknowledgments

This research was supported by Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai) (No. SML2023SP216), the National Natural Science Foundation of China (under Grants 42088101, 42205149, 42475172, 42375166, 42375164, and 42175158), the Guangdong Major Project of Basic and Applied Basic Research (2021B0301030007). This work is also supported by the specific research fund of The Innovation Platform for Academicians of Hainan Province (YSPTZX202143). Zhongwang Wei is supported by Guangdong Pearl River Talent Program (Young Talents) No. 2021QN02G307. The code used to reproduce the results is shown at https://github.com/leelew/DataML.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.agrformet.2025.110645.

References

Amani, S., Shafizadeh-Moghadam, H., 2023. A review of machine learning models and influential factors for estimating evapotranspiration using remote sensing and ground-based data. Agric. Water. Manag. 284, 108324. https://doi.org/10.1016/j.agwat.2023.108324.

Bai, K., Li, Ke, Ma, M., Li, K., Li, Z., Guo, J., Chang, N.B., Tan, Z., Han, D., 2022. LGHAP: the long-term gap-free high-resolution air pollutant concentration dataset, derived via tensor-flow-based multimodal data fusion. Earth Syst. Sci. Data 14, 907–927. https://doi.org/10.5194/essd-14-907-2022.

Li, B., Rodell, M., Kumar, S., Beaudoing, H.K., Getirana, A., Zaitchik, B.F., De Goncalves, L.G., Cossetin, C., Bhanja, S., Mukherjee, A., Tian, S., Tangdamrongsub, N., Long, D., Nanteza, J., Lee, J., Policelli, F., Goni, I.B., Daira, D., Bila, M., De Lannoy, G., Mocko, D., Steele-Dunne, S.C., Save, H., Bettadpur, S., 2019a. Global GRACE Data assimilation for groundwater and drought monitoring: advances and challenges. Water Resour. Res. 55, 7564–7586. https://doi.org/10.1029/2018WR024618.

Baldocchi, D., Falge, E., Gu, L., Olson, R., Hollinger, D., Running, S., Anthoni, P., Bernhofer, C., Davis, K., Evans, R., Fuentes, J., Goldstein, A., Katul, G., Law, B., Lee, X., Malhi, Y., Meyers, T., Munger, W., Oechel, W., Paw, K.T., Pilegaard, K., Schmid, H.P., Valentini, R., Verma, S., Vesala, T., Wilson, K., Wofsy, S., 2001. FLUXNET: a new tool to study the temporal and spatial variability of ecosystem–Scale carbon dioxide, water vapor, and energy flux densities. Bull. Am. Meteor. Soc. 82, 2415–2434. https://doi.org/10.1175/1520-0477(2001)082<2415:FANTTS>2.3.CO;2.

Ball, J.E., Anderson, D.T., Chan, C.S., 2017. Comprehensive survey of deep learning in remote sensing: theories, tools, and challenges for the community. J. Appl. Remote Sens. 11, 1. https://doi.org/10.1117/1.JRS.11.042609.

Barcza, Z., Kern, A., Haszpra, L., Kljun, N., 2009. Spatial representativeness of tall tower eddy covariance measurements using remote sensing and footprint analysis. Agric. For. Meteorol. 149, 795–807. https://doi.org/10.1016/j.agrformet.2008.10.021.

Beck, H.E., Zimmermann, N.E., McVicar, T.R., Vergopolan, N., Berg, A., Wood, E.F., 2018. Present and future Köppen-Geiger climate classification maps at 1-km resolution. Sci. Data 5, 180214. https://doi.org/10.1038/sdata.2018.214.

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. https://doi.org/10.5194/essd-10-1327-2018.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Brutsaert, W. (viaf)12584656, 1982. Evaporation into the atmosphere: theory, history, and applications. Dordrecht.

Bu, J., Gan, G., Chen, J., Su, Y., Yuan, M., Gao, Y., Domingo, F., López-Ballesteros, A., Migliavacca, M., El-Madany, T.S., Gentine, P., Xiao, J., Garcia, M., 2024. Dryland evapotranspiration from remote sensing solar-induced chlorophyll fluorescence: constraining an optimal stomatal model within a two-source energy balance model. Remote Sens. Environ. 303, 113999. https://doi.org/10.1016/j.rse.2024.113999.

Camps-Valls, G., Tuia, D., Zhu, X.X., Reichstein, M., 2021. Deep learning for the earth sciences: a comprehensive approach to remote sensing, climate science and geosciences. 10.1002/9781119646181.

Cao, S., Li, M., Zhu, Z., Wang, Z., Zha, J., Zhao, W., Duanmu, Z., Chen, J., Zheng, Y., Chen, Y., Myneni, R.B., Piao, S., 2023. Spatiotemporally consistent global dataset of the GIMMS leaf area index (GIMMS LAI4g) from 1982 to 2020. Earth Syst. Sci. Data 15, 4877–4899. https://doi.org/10.5194/essd-15-4877-2023.

Changming, Li, Yang, H., Yang, W., Liu, Z., Jia, Y., Li, S., Yang, D., 2022. CAMELE: collocation-analyzed multi-source ensembled land evapotranspiration data. Earth Syst. Sci. Data Discussions 2022, 1–45. https://doi.org/10.5194/essd-2021-456.

Chen, B., Ge, Q., Fu, D., Yu, G., Sun, X., Wang, S., Wang, H., 2010. A data-model fusion approach for upscaling gross ecosystem productivity to the landscape scale based on remote sensing and flux footprint modelling. Biogeosciences 7, 2943–2958. https://doi.org/10.5194/bg-7-2943-2010.

Chen, H., Wang, T., Zhang, Y., Bai, Y., Chen, X., 2023. Dynamically weighted ensemble of geoscientific models via automated machine-learning-based classification. Geosci. Model. Dev. 16, 5685–5701. https://doi.org/10.5194/gmd-16-5685-2023.

Chen, T., Guestrin, C., 2016. XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.

Chen, X., Su, Z., Ma, Y., Trigo, I., Gentine, P., 2021a. Remote sensing of global daily evapotranspiration based on a surface energy balance method and reanalysis data. JGR Atmos. 126. https://doi.org/10.1029/2020JD032873 e2020JD032873.

Chollet, F., 2017. Deep learning with python. Manning.

Chu, H., Luo, X., Ouyang, Z., Chan, W.S., Dengel, S., Biraud, S.C., Torn, M.S., Metzger, S., Kumar, J., Arain, M.A., Arkebauer, T.J., Baldocchi, D., Bernacchi, C., Billesbach, D., Black, T.A., Blanken, P.D., Bohrer, G., Bracho, R., Brown, S., Brunsell, N.A., Chen, J., Chen, X., Clark, K., Desai, A.R., Duman, T., Durden, D., Fares, S., Forbrich, I., Gamon, J.A., Gough, C.M., Griffis, T., Helbig, M., Hollinger, D., Humphreys, E., Ikawa, H., Iwata, H., Ju, Y., Knowles, J.F., Knox, S.H., Kobayashi, H., Kolb, T., Law, B., Lee, X., Litvak, M., Liu, H., Munger, J.W., Noormets, A., Novick, K., Oberbauer, S.F., Oechel, W., Oikawa, P., Papuga, S.A., Pendall, E., Prajapati, P., Prueger, J., Quinton, W.L., Richardson, A.D., Russell, E.S., Scott, R.L., Starr, G., Staebler, R., Stoy, P.C., Stuart-Haëntjens, E., Sonnentag, O., Sullivan, R.C., Suyker, A., Ueyama, M., Vargas, R., Wood, J.D., Zona, D., 2021. Representativeness of eddy-covariance flux footprints for areas surrounding AmeriFlux sites. Agric. For. Meteorol. 301–302, 108350. https://doi.org/10.1016/j.agrformet.2021.108350.

Collier, N., Hoffman, F.M., Lawrence, D.M., Keppel-Aleks, G., Koven, C.D., Riley, W.J., Mu, M., Randerson, J.T., 2018. The International Land Model Benchmarking (ILAMB) System: design, theory, and implementation. J. Adv. Model. Earth Syst. 10, 2731–2754. https://doi.org/10.1029/2018MS001354.

ElGhawi, R., Kraft, B., Reimers, C., Reichstein, M., Körner, M., Gentine, P., Winkler, A.J., 2023. Hybrid modeling of evapotranspiration: inferring stomatal and aerodynamic resistances using combined physics-based and machine learning. Environ. Res. Lett. 18, 034039. https://doi.org/10.1088/1748-9326/acbbe0.

Elith, J., Leathwick, J.R., Hastie, T., 2008. A working guide to boosted regression trees. J. Anim. Ecol. 77, 802–813. https://doi.org/10.1111/j.1365-2656.2008.01390.x.

Erik, S., Kononenko, I., 2014. Explaining prediction models and individual predictions with feature contributions. Knowl. Inf. Syst. 41, 647–665.

Fan, J., Ma, X., Wu, L., Zhang, F., Yu, X., Zeng, W., 2019. Light gradient boosting machine: an efficient soft computing model for estimating daily reference evapotranspiration with local and external meteorological data. Agric. Water. Manag. 225, 105758. https://doi.org/10.1016/j.agwat.2019.105758.

Feng, F., Li, X., Yao, Y., Liang, S., Chen, J., Zhao, X., Jia, K., Pintér, K., McCaughey, J.H., 2016. An empirical orthogonal function-based algorithm for estimating terrestrial latent heat flux from Eddy covariance, meteorological and satellite observations. PLoS One 11, e0160150. https://doi.org/10.1371/journal.pone.0160150.

Fleischmann, A.S., Laipelt, L., Papa, F., Paiva, R.C.D.D., De Andrade, B.C., Collischonn, W., Biudes, M.S., Kayser, R., Prigent, C., Cosio, E., Machado, N.G., Ruhoff, A., 2023. Patterns and drivers of evapotranspiration in South American wetlands. Nat. Commun. 14, 6656. https://doi.org/10.1038/s41467-023-42467-0.

Friedl, M.A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., Huang, X., 2010. MODIS collection 5 global land cover: algorithm refinements and characterization of new datasets. Remote Sens. Environ. 114, 168–182. https://doi.org/10.1016/j.rse.2009.08.016.

Gelaro, R., McCarty, W., Suárez, M.J., Todling, R., Molod, A., Takacs, L., Randles, C.A., Darmenov, A., Bosilovich, M.G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., Da Silva, A.M., Gu, W., Kim, G.K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J.E., Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert, S.D., Sienkiewicz, M., Zhao, B., 2017. The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). J. Clim. 30, 5419–5454. https://doi.org/10.1175/JCLI-D-16-0758.1.

Geurts, P., Ernst, D., Wehenkel, L., 2006. Extremely randomized trees. Mach. Learn. 63, 3–42. https://doi.org/10.1007/s10994-006-6226-1.
Giardina, F., Gentine, P., Konings, A.G., Seneviratne, S.I., Stocker, B.D., 2023. Diagnosing evapotranspiration responses to water deficit across biomes using deep learning. New Phytol. 240, 968–983. https://doi.org/10.1111/nph.19197.
Hao, Y., Baik, J., Choi, M., 2019. Combining generalized complementary relationship models with the Bayesian model averaging method to estimate actual evapotranspiration over China. Agric. For. Meteorol. 279, 107759. https://doi.org/10.1016/j.agrformet.2019.107759.
He, X., Zhao, K., Chu, X., 2021. AutoML: a survey of the state-of-the-art. Knowl. Based Syst. 212, 106622. https://doi.org/10.1016/j.knosys.2020.106622.
Hengl, T., Mendes De Jesus, J., Heuvelink, G.B.M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M.N., Geng, X., Bauer-Marschallinger, B., Guevara, M.A., Vargas, R., MacMillan, R.A., Batjes, N.H., Leenaars, J.G.B., Ribeiro, E., Wheeler, I., Mantel, S., Kempen, B., 2017. SoilGrids250m: global gridded soil information based on machine learning. PLoS One 12, e0169748. https://doi.org/10.1371/journal.pone.0169748.
Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R.J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., De Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., Thépaut, J., 2020. The ERA5 global reanalysis. Q. J. R. Meteorol. Soc. 146, 1999–2049. https://doi.org/10.1002/qj.3803.
Hinton, G.E., Salakhutdinov, R.R., 2006. Reducing the dimensionality of data with neural networks. Science 313, 504–507. https://doi.org/10.1126/science.1127647 (1979).
Hirschi, M., Seneviratne, S.I., 2017. Basin-scale water-balance dataset (BSWB): an update. Earth Syst. Sci. Data 9, 251–258. https://doi.org/10.5194/essd-9-251-2017.
Hutter, F., Kotthoff, L., Vanschoren, J., 2019. Automated machine learning: methods, systems, challenges. Automated Machine Learning.
Jahromi, M.N., Miralles, D., Koppa, A., Rains, D., Zand-Parsa, S., Mosaffa, H., Jamshidi, S., 2022. Ten years of GLEAM: a review of scientific advances and applications. Comp. Intell. Water Environ. Sci. 525–540.
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. https://doi.org/10.1038/nature11983.
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. https://doi.org/10.1038/s41597-019-0076-8.
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. https://doi.org/10.1038/nature09396.
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. https://doi.org/10.1007/s10712-008-9037-z.
Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.Y., 2017. LightGBM: a highly efficient gradient boosting decision tree, in: neural Information processing systems.
Kenney, J.F., Keeping, E.S., 1962. Mathematics of statistics, Mathematics of Statistics. Van Nostrand Company.
Knighton, J., Berghuijs, W.R., 2023. Water ages explain tradeoffs between long-term evapotranspiration and ecosystem drought resilience. Geophys. Res. Lett. 50. https://doi.org/10.1029/2023GL103649 e2023GL103649.
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. https://doi.org/10.1038/s41467-022-29543-7.
LeDell, E., Poirier, S., 2020. H2O AutoML: scalable automatic machine learning.
Lin, W., Yuan, H., Dong, W., Zhang, S., Liu, S., Wei, N., Lu, X., Wei, Z., Hu, Y., Dai, Y., 2023. Reprocessed MODIS version 6.1 leaf area index dataset and its evaluation for land surface and climate modeling. Remote Sens. 15, 1780. https://doi.org/10.3390/rs15071780 (Basel).
Liu, W., Wang, L., Zhou, J., Li, Y., Sun, F., Fu, G., Li, X., Sang, Y.F., 2016. A worldwide evaluation of basin-scale evapotranspiration estimates against the water balance method. J. Hydrol. 538, 82–95. https://doi.org/10.1016/j.jhydrol.2016.04.006.
Lu, J., Wang, G., Chen, T., Li, S., Hagan, D.F.T., Kattel, G., Peng, J., Jiang, T., Su, B., 2021. A harmonized global land evaporation dataset from model-based products covering 1980–2017. Earth Syst. Sci. Data 13, 5879–5898. https://doi.org/10.5194/essd-13-5879-2021.
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. https://doi.org/10.1016/j.rse.2010.04.001.
Ma, N., Zhang, Y., Szilagyi, J., 2024. Water-balance-based evapotranspiration for 56 large river basins: a benchmarking dataset for global terrestrial evapotranspiration modeling. J. Hydrol. 630, 130607. https://doi.org/10.1016/j.jhydrol.2024.130607.
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., 2020a. A long-term (2005–2016) dataset of hourly integrated land–atmosphere interaction observations on the Tibetan Plateau. Earth Syst. Sci. Data 12, 2937–2957. https://doi.org/10.5194/essd-12-2937-2020.
Machado, M.R., Karray, S., De Sousa, I.T., 2019. LightGBM: an effective decision tree gradient boosting method to predict customer loyalty in the finance industry. In: Proceedings of the 2019 14th International Conference on Computer Science & Education (ICCSE), pp. 1111–1116. https://doi.org/10.1109/ICCSE.2019.8845529.
Mangalath Ravindran, S., Moorakkal Bhaskaran, S.K., K Ambat, S., Balakrishnan, K., Manguttathil Gopalakrishnan, M., 2022. An automated machine learning methodology for the improved prediction of reference evapotranspiration using minimal input parameters. Hydrol. Process. 36, e14571. https://doi.org/10.1002/hyp.14571.
Martens, B., De Jeu, R.A.M., Verhoest, N.E.C., Schuurmans, H., Kleijer, J., Miralles, D.G., 2018. Towards estimating land evaporation at field scales using GLEAM. Remote Sens. 10, 1720. https://doi.org/10.3390/rs10111720 (Basel).
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. https://doi.org/10.5194/gmd-10-1903-2017.
McCabe, M.F., Ershadi, A., Jimenez, C., Miralles, D.G., Michel, D., Wood, E.F., 2016. The GEWEX LandFlux project: evaluation of model evaporation using tower-based and globally gridded forcing data. Geosci. Model. Dev. 9, 283–305. https://doi.org/10.5194/gmd-9-283-2016.
McNicol, G., Fluet-Chouinard, E., Ouyang, Z., Knox, S., Zhang, Z., Aalto, T., Bansal, S., Chang, K., Chen, M., Delwiche, K., Feron, S., Goeckede, M., Liu, J., Malhotra, A., Melton, J.R., Riley, W., Vargas, R., Yuan, K., Ying, Q., Zhu, Q., Alekseychik, P., Aurela, M., Billesbach, D.P., Campbell, D.I., Chen, J., Chu, H., Desai, A.R., Euskirchen, E., Goodrich, J., Griffis, T., Helbig, M., Hirano, T., Iwata, H., Jurasinski, G., King, J., Koebsch, F., Kolka, R., Krauss, K., Lohila, A., Mammarella, I., Nilson, M., Noormets, A., Oechel, W., Peichl, M., Sachs, T., Sakabe, A., Schulze, C., Sonnentag, O., Sullivan, R.C., Tuittila, E., Ueyama, M., Vesala, T., Ward, E., Wille, C., Wong, G.X., Zona, D., Windham-Myers, L., Poulter, B., Jackson, R.B., 2023. Upscaling wetland methane emissions from the FLUXNET-CH4 Eddy Covariance Network (UpCH4 v1.0): model development, network assessment, and budget comparison. AGU Adv. 4, e2023AV000956. https://doi.org/10.1029/2023AV000956.
Melton, F.S., Huntington, J., Grimm, R., Herring, J., Hall, M., Rollison, D., Erickson, T., Allen, R., Anderson, M., Fisher, J.B., Kilic, A., Senay, G.B., Volk, J., Hain, C., Johnson, L., Ruhoff, A., Blankenau, P., Bromley, M., Carrara, W., Daudert, B., Doherty, C., Dunkerly, C., Friedrichs, M., Guzman, A., Halverson, G., Hansen, J., Harding, J., Kang, Y., Ketchum, D., Minor, B., Morton, C., Ortega-Salazar, S., Ott, T., Ozdogan, M., ReVelle, P.M., Schull, M., Wang, C., Yang, Y., Anderson, R.G., 2022. OpenET: filling a critical data gap in water management for the western United States. J. Am. Water Resour. Assoc. 58, 971–994. https://doi.org/10.1111/1752-1688.12956.
Chen, M., Willgoose, G.R., Saco, P.M., 2015a. Investigating the impact of leaf area index temporal variability on soil moisture predictions using remote sensing vegetation data. J. Hydrol. 522, 274–284. https://doi.org/10.1016/j.jhydrol.2014.12.027.
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. https://doi.org/10.5194/hess-20-823-2016.
Miralles, D.G., Teuling, A.J., Van Heerwaarden, C.C., Vilà-Guerau De Arellano, J., 2014. Mega-heatwave temperatures due to combined soil desiccation and atmospheric heat accumulation. Nat. Geosci. 7, 345–349. https://doi.org/10.1038/ngeo2141.
Mueller, B., Hirschi, M., Jimenez, C., Ciais, P., Dirmeyer, P.A., Dolman, A.J., Fisher, J.B., Jung, M., Ludwig, F., Maignan, F., Miralles, D.G., McCabe, M.F., Reichstein, M., Sheffield, J., Wang, K., Wood, E.F., Zhang, Y., Seneviratne, S.I., 2013. Benchmark products for land evapotranspiration: landFlux-EVAL multi-data set synthesis. Hydrol. Earth Syst. Sci. 17, 3707–3720. https://doi.org/10.5194/hess-17-3707-2013.
Mueller, B., Seneviratne, S.I., Jimenez, C., Corti, T., Hirschi, M., Balsamo, G., Ciais, P., Dirmeyer, P.A., Fisher, J.B., Guo, Z., Jung, M., Maignan, F., Mccabe, M.F., Reichle, R. H., Reichstein, M., Rodell, M., Sheffield, J., Teuling, A.J., Wang, K., Wood, E.F., Zhang, Y., 2011. Evaluation of global observations-based evapotranspiration datasets and IPCC AR4 simulations. Geophys. Res. Lett. 38.
Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles, D.G., Piles, M., Rodríguez-Fernández, N.J., Zsoter, E., Buontempo, C., Thépaut, J.N., 2021. ERA5-Land: a state-of-the-art global reanalysis dataset for land applications. Earth Syst. Sci. Data 13, 4349–4383. https://doi.org/10.5194/essd-13-4349-2021.
Ma, N., Szilagyi, J., Jozsa, J., 2020b. Benchmarking large-scale evapotranspiration estimates: a perspective from a calibration-free complementary relationship approach and FLUXCOM. J. Hydrol. https://doi.org/10.1016/j.jhydrol.2020.125221.
Oki, T., Kanae, S., 2006. Global hydrological cycles and world water resources. Science 313, 1068–1072. https://doi.org/10.1126/science.1128845 (1979).
Ostroumova, L., Gusev, G., Vorobev, A., Dorogush, A.V., Gulin, A., 2017. CatBoost: unbiased boosting with categorical features, in: neural information processing systems.
Pan, S., Pan, N., Tian, H., Friedlingstein, P., Sitch, S., Shi, H., Arora, V.K., Haverd, V., Jain, A.K., Kato, E., Lienert, S., Lombardozzi, D., Nabel, J.E.M.S., Ottlé, C., Poulter, B., Zaehle, S., Running, S.W., 2020. Evaluation of global terrestrial evapotranspiration using state-of-the-art approaches in remote sensing, machine learning and land surface modeling. Hydrol. Earth Syst. Sci. 24, 1485–1509. https://doi.org/10.5194/hess-24-1485-2020.

Parrish, M.A., Moradkhani, H., DeChant, C.M., 2012. Toward reduction of model uncertainty: integration of Bayesian model averaging and data assimilation. Water Resour. Res. 48, 2011WR011116. https://doi.org/10.1029/2011WR011116.

Pastorello, G., Trotta, C., Canfora, E., Chu, H., Christianson, D., Cheah, Y.W., Poindexter, C., Chen, J., Elbashandy, A., Humphrey, M., Isaac, P., Polidori, D., Reichstein, M., Ribeca, A., Van Ingen, C., Vuichard, N., Zhang, L., Amiro, B., Ammann, C., Arain, M.A., Ardö, J., Arkebauer, T., Arndt, S.K., Arriga, N., Aubinet, M., Aurela, M., Baldocchi, D., Barr, A., Beamesderfer, E., Marchesini, L.B., Bergeron, O., Beringer, J., Bernhofer, C., Berveiller, D., Billesbach, D., Black, T.A., Blanken, P.D., Bohrer, G., Boike, J., Bolstad, P.V., Bonal, D., Bonnefond, J.M., Bowling, D.R., Bracho, R., Brodeur, J., Brümmer, C., Buchmann, N., Burban, B., Burns, S.P., Buysse, P., Cale, P., Cavagna, M., Cellier, P., Chen, S., Chini, I., Christensen, T.R., Cleverly, J., Collalti, A., Consalvo, C., Cook, B.D., Cook, D., Coursolle, C., Cremonese, E., Curtis, P.S., D'Andrea, E., Da Rocha, H., Dai, X., Davis, K.J., Cinti, B.D., Grandcourt, A.D., Ligne, A.D., De Oliveira, R.C., Delpierre, N., Desai, A.R., Di Bella, C.M., Tommasi, P.D., Dolman, H., Domingo, F., Dong, G., Dore, S., Duce, P., Dufrêne, E., Dunn, A., Dušek, J., Eamus, D., Eichelmann, U., ElKhidir, H.A.M., Eugster, W., Ewenz, C.M., Ewers, B., Famulari, D., Fares, S., Feigenwinter, I., Feitz, A., Fensholt, R., Filippa, G., Fischer, M., Frank, J., Galvagno, M., Gharun, M., Gianelle, D., Gielen, B., Gioli, B., Gitelson, A., Goded, I., Goeckede, M., Goldstein, A.H., Gough, C.M., Goulden, M.L., Graf, A., Griebel, A., Gruening, C., Grünwald, T., Hammerle, A., Han, S., Han, X., Hansen, B.U., Hanson, C., Hatakka, J., He, Y., Hehn, M., Heinesch, B., Hinko-Najera, N., Hörtnagl, L., Hutley, L., Ibrom, A., Ikawa, H., Jackowicz-Korczynski, M., Janouš, D., Jans, W., Jassal, R., Jiang, S., Kato, T., Khomik, M., Klatt, J., Knohl, A., Knox, S., Kobayashi, H., Koerber, G., Kolle, O., Kosugi, Y., Kotani, A., Kowalski, A., Kruijt, B., Kurbatova, J., Kutsch, W.L., Kwon, H., Launiainen, S., Laurila, T., Law, B., Leuning, R., Li, Y., Liddell, M., Limousin, J.M., Lion, M., Liska, A.J., Lohila, A., López-Ballesteros, A., López-Blanco, E., Loubet, B., Loustau, D., Lucas-Moffat, A., Lüers, J., Ma, S., Macfarlane, C., Magliulo, V., Maier, R., Mammarella, I., Manca, G., Marcolla, B., Margolis, H.A., Marras, S., Massman, W., Mastepanov, M., Matamala, R., Matthes, J.H., Mazzenga, F., McCaughey, H., McHugh, I., McMillan, A.M.S., Merbold, L., Meyer, W., Meyers, T., Miller, S.D., Minerbi, S., Moderow, U., Monson, R.K., Montagnani, L., Moore, C.E., Moors, E., Moreaux, V., Moureaux, C., Munger, J.W., Nakai, T., Neirynck, J., Nesic, Z., Nicolini, G., Noormets, A., Northwood, M., Nosetto, M., Nouvellon, Y., Novick, K., Oechel, W., Olesen, J.E., Ourcival, J.M., Papuga, S.A., Parmentier, F.J., Paul-Limoges, E., Pavelka, M., Peichl, M., Pendall, E., Phillips, R.P., Pilegaard, K., Pirk, N., Posse, G., Powell, T., Prasse, H., Prober, S.M., Rambal, S., Rannik, Ü., Raz-Yaseef, N., Rebmann, C., Reed, D., Dios, V.R.D., Restrepo-Coupe, N., Reverter, B.R., Roland, M., Sabbatini, S., Sachs, T., Saleska, S.R., Sánchez-Cañete, E.P., Sanchez-Mejia, Z.M., Schmid, H.P., Schmidt, M., Schneider, K., Schrader, F., Schroder, I., Scott, R.L., Sedlák, P., Serrano-Ortíz, P., Shao, C., Shi, P., Shironya, I., Siebicke, L., Šigut, L., Silberstein, R., Sirca, C., Spano, D., Steinbrecher, R., Stevens, R.M., Sturtevant, C., Suyker, A., Tagesson, T., Takanashi, S., Tang, Y., Tapper, N., Thom, J., Tomassucci, M., Tuovinen, J.P., Urbanski, S., Valentini, R., Van Der Molen, M., Van Gorsel, E., Van Huissteden, K., Varlagin, A., Verfaillie, J., Vesala, T., Vincke, C., Vitale, D., Vygodskaya, N., Walker, J.P., Walter-Shea, E., Wang, H., Weber, R., Westermann, S., Wille, C., Wofsy, S., Wohlfahrt, G., Wolf, S., Woodgate, W., Li, Y., Zampedri, R., Zhang, J., Zhou, G., Zona, D., Agarwal, D., Biraud, S., Torn, M., Papale, D., 2020. The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data. Sci. Data 7, 225. https://doi.org/10.1038/s41597-020-0534-3.

Pearson, K., 1920. Notes on the history of correlation. Biometrika 13, 25–45. https://doi.org/10.1093/biomet/13.1.25.

Piao, S., Wang, X., Park, T., Chen, C., Lian, X., He, Y., Bjerke, J.W., Chen, A., Ciais, P., Tømmervik, H., Nemani, R.R., Myneni, R.B., 2019. Characteristics, drivers and feedbacks of global greening. Nat. Rev. Earth Environ. 1, 14–27. https://doi.org/10.1038/s43017-019-0001-x.

Ploton, P., Mortier, F., Réjou-Méchain, M., Barbier, N., Picard, N., Rossi, V., Dormann, C., Cornu, G., Viennois, G., Bayol, N., Lyapustin, A., Gourlet-Fleury, S., Pélissier, R., 2020. Spatial validation reveals poor predictive performance of large-scale ecological mapping models. Nat. Commun. 11, 4540. https://doi.org/10.1038/s41467-020-18321-y.

Purdy, A.J., Fisher, J.B., Goulden, M.L., Colliander, A., Halverson, G., Tu, K., Famiglietti, J.S., 2018. SMAP soil moisture improves global evapotranspiration. Remote Sens. Environ. 219, 1–14. https://doi.org/10.1016/j.rse.2018.09.023.

Li, Q., Shi, G., Shangguan, W., Nourani, V., Li, J., Li, L., Huang, F., Zhang, Y., Wang, C., Wang, D., Qiu, J., Lu, X., Dai, Y., 2022. A 1 km daily soil moisture dataset over China using in situ measurement and machine learning. Earth Syst. Sci. Data 14, 5267–5286. https://doi.org/10.5194/essd-14-5267-2022.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., Prabhat, 2019. Deep learning and process understanding for data-driven Earth system science. Nature 566, 195–204. https://doi.org/10.1038/s41586-019-0912-1.

Rodell, M., Houser, P.R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C.J., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M., Entin, J.K., Walker, J.P., Lohmann, D., Toll, D., 2004. The global land data assimilation system. Bull. Am. Meteor. Soc. 85, 381–394. https://doi.org/10.1175/BAMS-85-3-381.

Ruhoff, A., De Andrade, B.C., Laipelt, L., Fleischmann, A.S., Siqueira, V.A., Moreira, A.A., Barbedo, R., Cyganski, G.L., Fernandez, G.M.R., Brêda, J.P.L.F., Paiva, R.C.D.D., Meller, A., Teixeira, A.D.A., Araújo, A.A., Fuckner, M.A., Biggs, T., 2022. Global evapotranspiration datasets assessment using water balance in South America. Remote Sens. 14, 2526. https://doi.org/10.3390/rs14112526 (Basel).

Scanlon, B.R., Zhang, Z., Save, H., Sun, A.Y., Müller Schmied, H., Van Beek, L.P.H., Wiese, D.N., Wada, Y., Long, D., Reedy, R.C., Longuevergne, L., Döll, P., Bierkens, M.F.P., 2018. Global models underestimate large decadal declining and rising water storage trends relative to GRACE satellite data. Proc. Natl. Acad. Sci. U. S. A. 115, E1080–E1089. https://doi.org/10.1073/pnas.1704665115.

Senay, G.B., Leake, S., Nagler, P.L., Artan, G., Dickinson, J., Cordova, J.T., Glenn, E.P., 2011. Estimating basin scale evapotranspiration (ET) by water balance and remote sensing methods. Hydrol. Process. 25, 4037–4049. https://doi.org/10.1002/hyp.8379.

Shang, K., Yao, Y., Di, Z., Jia, K., Zhang, X., Fisher, J.B., Chen, J., Guo, X., Yang, J., Yu, R., Xie, Z., Liu, L., Ning, J., Zhang, L., 2023. Coupling physical constraints with machine learning for satellite-derived evapotranspiration of the Tibetan Plateau. Remote Sens. Environ. 289, 113519. https://doi.org/10.1016/j.rse.2023.113519.

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. https://doi.org/10.3390/rs12040687 (Basel).

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. https://doi.org/10.1016/j.agrformet.2021.108582.

Shapley, L.S., 1953. A value for n-person games. In: Kuhn, H.W., Tucker, A.W. (Eds.), Contributions to the Theory of Games II. Princeton University Press, Princeton, pp. 307–317.

Sheffield, J., Wood, E.F., 2007. Characteristics of global and regional drought, 1950–2000: analysis of soil moisture data from off-line simulation of the terrestrial hydrologic cycle. J. Geophys. Res. 112, 2006JD008288. https://doi.org/10.1029/2006JD008288.

Stephen, K.D., Kazemi, A., 2014. Improved normalization of time-lapse seismic data using normalized root mean square repeatability data to improve automatic production and seismic history matching in the nelson field. Geophys. Prospect. 62, 1009–1027. https://doi.org/10.1111/1365-2478.12109.

Thongthammachart, T., Araki, S., Shimadera, H., Matsuo, T., Kondo, A., 2022. Incorporating Light Gradient Boosting Machine to land use regression model for estimating NO2 and PM2.5 levels in Kansai region, Japan. Environ. Model. Softw. 155, 105447. https://doi.org/10.1016/j.envsoft.2022.105447.

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

Upton, S., Reichstein, M., Gans, F., Peters, W., Kraft, B., Bastos, A., 2023. Constraining biospheric carbon dioxide fluxes by combined top-down and bottom-up approaches. EGUsphere 2023, 1–31. https://doi.org/10.5194/egusphere-2023-805.

Wang, C., Wu, Q., Weimer, M., Zhu, E., 2021. FLAML: a fast and lightweight AutoML library.

Wang, K., Dickinson, R.E., 2012. A review of global terrestrial evapotranspiration: observation, modeling, climatology, and climatic variability. Rev. Geophys. 50. https://doi.org/10.1029/2011RG000373, 2011RG000373.

Wantong, Li, Migliavacca, M., Forkel, M., Denissen, J.M.C., Reichstein, M., Yang, H., Duveiller, G., Weber, U., Orth, R., 2022. Widespread increasing vegetation sensitivity to soil moisture. Nat. Commun. 13, 3959. https://doi.org/10.1038/s41467-022-31667-9.

Wei, Z., Lee, X., Wen, X., Xiao, W., 2018. Evapotranspiration partitioning for three agro-ecosystems with contrasting moisture conditions: a comparison of an isotope method and a two-source model calculation. Agric. For. Meteorol. 252, 296–310. https://doi.org/10.1016/j.agrformet.2018.01.019.

Wong, S.C., Cowan, I.R., Farquhar, G.D., 1979. Stomatal conductance correlates with photosynthetic capacity. Nature 282, 424–426. https://doi.org/10.1038/282424a0.

Xie, Z., Yao, Y., Zhang, X., Liang, S., Fisher, J.B., Chen, J., Jia, K., Shang, K., Yang, J., Yu, R., Guo, X., Liu, L., Ning, J., Zhang, L., 2022. The global LAnd surface satellite (GLASS) evapotranspiration product version 5.0: algorithm development and preliminary validation. J. Hydrol. 610, 127990. https://doi.org/10.1016/j.jhydrol.2022.127990.

Xiong, J., Abhishek, Xu, L., Chandanpurkar, H.A., Famiglietti, J.S., Zhang, C., Ghiggi, G., Guo, S., Pan, Y., Vishwakarma, B.D., 2023. ET-WB: water-balance-based estimations of terrestrial evaporation over global land and major global basins. Earth Syst. Sci. Data 15, 4571–4597. https://doi.org/10.5194/essd-15-4571-2023.

Xu, T., Guo, Z., Liu, S., He, X., Meng, Y., Xu, Z., Xia, Y., Xiao, J., Zhang, Y., Ma, Y., Song, L., 2018. Evaluating different machine learning methods for upscaling evapotranspiration from flux towers to the regional scale. JGR Atmos. 123, 8674–8690. https://doi.org/10.1029/2018JD028447.

Li, X., Long, D., Han, Z., Scanlon, B.R., Sun, Z., Han, P., Hou, A., 2019b. Evapotranspiration estimation for Tibetan plateau headwaters using conjoint terrestrial and atmospheric water balances and multisource remote sensing. Water Resour. Res. 55, 8608–8630. https://doi.org/10.1029/2019WR025196.

Yamazaki, D., Ikeshima, D., Tawatari, R., Yamaguchi, T., O'Loughlin, F., Neal, J.C., Sampson, C.C., Kanae, S., Bates, P.D., 2017. A high-accuracy map of global terrain elevations. Geophys. Res. Lett. 44, 5844–5853. https://doi.org/10.1002/2017GL072874.

Chen, Y., Yuan, W., Xia, J., Fisher, J.B., Dong, W., Zhang, X., Liang, S., Ye, A., Cai, W., Feng, J., 2015b. Using Bayesian model averaging to estimate terrestrial evapotranspiration in China. J. Hydrol. 528, 537–549. https://doi.org/10.1016/j.jhydrol.2015.06.059.

Yang, F., White, M.A., Michaelis, A.R., Ichii, K., Hashimoto, H., Votava, P., Zhu, A.X., Nemani, R.R., 2006. Prediction of continental-scale evapotranspiration by combining MODIS and AmeriFlux data through support vector machine. IEEE Trans. Geosci. Remote Sens. 44, 3452–3461. https://doi.org/10.1109/TGRS.2006.876297.

Yang, Y., Chen, R., Song, Y., Han, C., Liu, J., Liu, Z., 2019. Sensitivity of potential evapotranspiration to meteorological factors and their elevational gradients in the

Qilian Mountains, northwestern China. J. Hydrol. 568, 147–159. https://doi.org/10.1016/j.jhydrol.2018.10.069.

Yang, Y., Roderick, M.L., Guo, H., Miralles, D.G., Zhang, L., Fatichi, S., Luo, X., Zhang, Y., McVicar, T.R., Tu, Z., Keenan, T.F., Fisher, J.B., Gan, R., Zhang, X., Piao, S., Zhang, B., Yang, D., 2023. Evapotranspiration on a greening earth. Nat. Rev. Earth Environ. 4, 626–641. https://doi.org/10.1038/s43017-023-00464-3.

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. https://doi.org/10.1016/j.agrformet.2012.11.016.

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., 2017a. Improving global terrestrial evapotranspiration estimation using support vector machine by integrating three process-based algorithms. Agric. For. Meteorol. 242, 55–74. https://doi.org/10.1016/j.agrformet.2017.04.011.

Yao, Y., Liang, S., Li, X., Hong, Y., Fisher, J.B., Zhang, N., Chen, J., Cheng, J., Zhao, S., Zhang, X., Jiang, B., Sun, L., Jia, K., Wang, K., Chen, Y., Mu, Q., Feng, F., 2014. Bayesian multimodel estimation of global terrestrial latent heat flux from eddy covariance, meteorological, and satellite observations. JGR Atmos. 119, 4521–4545. https://doi.org/10.1002/2013JD020864.

Yao, Y., Liang, S., Li, X., Zhang, Y., Chen, J., Jia, K., Zhang, X., Fisher, J.B., Wang, X., Zhang, L., Xu, J., Shao, C., Posse, G., Li, Y., Magliulo, V., Varlagin, A., Moors, E.J., Boike, J., Macfarlane, C., Kato, T., Buchmann, N., Billesbach, D.P., Beringer, J., Wolf, S., Papuga, S.A., Wohlfahrt, G., Montagnani, L., Ardö, J., Paul-Limoges, E., Emmel, C., Hörtnagl, L., Sachs, T., Gruening, C., Gioli, B., López-Ballesteros, A., Steinbrecher, R., Gielen, B., 2017b. Estimation of high-resolution terrestrial evapotranspiration from Landsat data using a simple Taylor skill fusion method. J. Hydrol. 553, 508–526. https://doi.org/10.1016/j.jhydrol.2017.08.013.

Yoo, B.H., Kim, J., Lee, B.W., Hoogenboom, G., Kim, K.S., 2020. A surrogate weighted mean ensemble method to reduce the uncertainty at a regional scale for the calculation of potential evapotranspiration. Sci. Rep. 10, 870. https://doi.org/10.1038/s41598-020-57466-0.

Yu, L., Qiu, G.Y., Yan, C., Zhao, W., Zou, Z., Ding, J., Qin, L., Xiong, Y., 2022. A global terrestrial evapotranspiration product based on the three-temperature model with fewer input parameters and no calibration requirement. Earth Syst. Sci. Data 14, 3673–3693. https://doi.org/10.5194/essd-14-3673-2022.

Yuan, Q., 2020. Deep learning in environmental remote sensing_ achievements and challenges. Remote Sens. Environ. 24.

Zhang, C., Brodylo, D., Rahman, M., Rahman, M.A., Douglas, T.A., Comas, X., 2022. Using an object-based machine learning ensemble approach to upscale evapotranspiration measured from eddy covariance towers in a subtropical wetland. Sci. Total Environ. 831, 154969. https://doi.org/10.1016/j.scitotenv.2022.154969.

Zhang, L., Marshall, M., Vrieling, A., Nelson, A., 2023. The divergence of energy- and water-balance evapotranspiration estimates in humid regions. J. Hydrol. 624, 129971. https://doi.org/10.1016/j.jhydrol.2023.129971.

Zhang, Y., Kong, D., Gan, R., Chiew, F.H.S., McVicar, T.R., Zhang, Q., Yang, Y., 2019. Coupled estimation of 500 m and 8-day resolution global evapotranspiration and gross primary production in 2002–2017. Remote Sens. Environ. 222, 165–182. https://doi.org/10.1016/j.rse.2018.12.031.

Zhang, Y., Schaap, M.G., Zha, Y., 2018. A high-resolution global map of soil hydraulic properties produced by a hierarchical parameterization of a physically based water retention model. Water Resour. Res. 54, 9774–9790.

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. https://doi.org/10.1029/2019GL085291.

Zheng, Z., Fiore, A.M., Westervelt, D.M., Milly, G.P., Goldsmith, J., Karambelas, A., Curci, G., Randles, C.A., Paiva, A.R., Wang, C., Wu, Q., Dey, S., 2023. Automated machine learning to evaluate the information content of tropospheric trace gas columns for fine particle estimates over India: a modeling testbed. J. Adv. Model. Earth Syst. 15, e2022MS003099. https://doi.org/10.1029/2022MS003099.

Zhu, G., Li, X., Zhang, K., Ding, Z., Han, T., Ma, J., Huang, C., He, J., Ma, T., 2016. Multi-model ensemble prediction of terrestrial evapotranspiration across north China using Bayesian model averaging: multi-Model prediction of terrestrial ET using BMA. Hydrol. Process. 30, 2861–2879. https://doi.org/10.1002/hyp.10832.

Zhu, S., Tristan, Q., Timothy, H., 2024. Uniform upscaling techniques for eddy covariance FLUXes (UFLUX). Int. J. Remote Sens. 45, 1450–1476. https://doi.org/10.1080/01431161.2024.2312266.