Agricultural and Forest Meteorology 308–309 (2021) 108582

Contents lists available at ScienceDirect

Agricultural and Forest Meteorology

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

# DNN-MET: A deep neural networks method to integrate satellite-derived evapotranspiration products, eddy covariance observations and ancillary information

Ke Shang [a], Yunjun Yao [a,*], Shunlin Liang [b], Yuhu Zhang [c], Joshua B. Fisher [d], Jiquan Chen [e], Shaomin Liu [f], Ziwei Xu [f], Yuan Zhang [f], Kun Jia [a], Xiaotong Zhang [a], Junming Yang [a], Xiangyi Bei [a], Xiaozheng Guo [a], Ruiyang Yu [a], Zijing Xie [a], Lilin Zhang [g]

[a] State Key Laboratory of Remote Sensing Science, Faculty of Geographical Science, Beijing Normal University, Beijing, 100875, China
[b] Department of Geographical Sciences, University of Maryland, College Park, MD 20742, USA
[c] College of Resource Environment and Tourism, Capital Normal University, Beijing 100048, China
[d] Joint Institute for Regional Earth System Science and Engineering, University of California at Los Angeles, Los Angeles, CA, 90095-7228, USA
[e] Department of Geography, Environment, and Spatial Sciences, Michigan State University, East Lansing, MI 48823, USA
[f] State Key Laboratory of Earth Surface Processes and Resource Ecology, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China
[g] Faculty of Geo-Information and Earth Observation (ITC), University of Twente, 7500 AE Enschede, The Netherlands

ARTICLE INFO

*Keywords:*
Evapotranspiration
Ancillary information
Deep neural networks
Merging method
DNN-MET

ABSTRACT

Accurate estimates of the spatiotemporal distribution of evapotranspiration (ET) are essential for understanding terrestrial energy, carbon and water cycles. Station-based observations are limited for their spatial coverage whereas satellite-derived ET products exhibit large discrepancies and uncertainties. Here we presented a Deep Neural Networks based Merging ET (DNN-MET) framework that combines information from satellite-derived ET products, eddy covariance (EC) observations and ancillary surface properties to improve the representation of the spatiotemporal distribution of ET, especially in data-sparse regions. DNN-MET was implemented over the Heihe River Basin (HRB) from 2008 to 2015, and the performance of DNN-MET and eight input state-of-the-art satellite-derived ET products (i.e., MOD16, ET-SEMI, ET-JPL, ET-MS, ET-HF, GLEAM, ETMonitor and EB-ET) was evaluated against observations from 19 EC flux tower sites. The results showed that DNN-MET improved ET estimates over HRB, and decreased the RMSE by 0.13 to 1.02 mm/day (14%-56%) when compared with eight products. DNN-MET also yielded superior performance compared to the products derived by other merging methods (i.e., Random Forest, Bayesian model averaging and a simple averaging method). When DNN-MET was validated for data-scarce regions, its performance remained better even when the training samples were decreased to 20% of the available EC sites. An innovation of our approach is by building a multivariate merging model with ancillary surface properties, DNN-MET incorporated geographical proximity effects and spatial autocorrelations into merging procedure, which can be used as a "spatial knowledge engine" to improve ET predictions. The approach can be readily and effectively applied elsewhere to improve the spatiotemporal representation of various hydrometeorological variables.

## 1. Introduction

Evapotranspiration (ET) — the loss of water from terrestrial ecosystems to the atmosphere through water evaporation and plant transpiration — is a critical nexus between the hydrological cycle and energy budgets within terrestrial ecosystems (Fisher et al. 2017; Jung et al. 2010; Kool et al. 2014; Liang et al. 2010b). More than half of the solar energy absorbed by land surfaces is used for ET (Trenberth et al. 2009), and ET returns approximately 60% of precipitation back to the atmosphere (Bonan 2015). Despite this importance, accurate ET estimations for its spatiotemporal changes remain challenging due to a combination of the complex feedback mechanisms, spatial representations of in situ observations, and heterogeneity nature of local topography and climate (Allen et al. 2011; Liang et al. 2019; Yao et al. 2015).

* Corresponding author at: Beijing Normal University, Beijing, 100875, China, Tel: 86+ 10-5880-3002
*E-mail address:* boyyunjun@163.com (Y. Yao).

https://doi.org/10.1016/j.agrformet.2021.108582
Received 27 January 2021; Received in revised form 28 July 2021; Accepted 31 July 2021

Integrating surface- and satellite-based observation systems with well-known flux equations (Fisher et al. 2008; Monteith 1965; Mu et al. 2011; Shuttleworth and Wallace 1985; Su 2002; Wang et al. 2010; Yao et al. 2013; Yao et al. 2017c) provides unprecedented opportunities for monitoring ET distribution at multiple spatiotemporal scales. These ET estimates exhibit relatively good agreement at global scale but yield substantial discrepancies regionally (Talsma et al. 2018). The discrepancies in ET volumes based on process-based models, reanalyses, upscaling estimates, and land surface models (LSM) can be 50% (Mao et al. 2015; Vinukollu et al. 2011). The spatial and temporal dynamics of each ET estimate appear highly uncertain (Maes et al. 2020), and each estimation has own advantages and limitations. For example, soil moisture (SM) constrained estimates performed better in semiarid regions, while other biophysical regulations should be more emphasized for vegetated areas (Aires 2014; Purdy et al. 2018). Nonetheless, no single individual ET product provides the most accurate estimate for all biomes, and each ET product can offer valuable information (Yao et al. 2014). This situation is more pronounced in regions where multiple climate modes coexist. A promising alternative is to blend multiple ET products and in situ observations to obtain: 1) more reliable estimates under variable climatic conditions, and 2) a benchmark for multiple ET products, i.e., a surrogate indicator for uncertainties in climate forecasts (Aires 2014; Liang et al. 2018).

Numerous efforts have been made to develop effective merging methods, ranging from simple averaging (SA) method (Mueller et al. 2013) to more complex models such as Bayesian model averaging (BMA) (Chen et al. 2015; Yao et al. 2014), Empirical Orthogonal Function (EOF) (Feng et al. 2016), the Taylor skill fusion method (Yao et al. 2017b), the weight dominated method (He et al. 2020; Jiménez et al. 2018), the water budget closure controlled integration method (Aires 2014), and the machine learning method (Shang et al. 2020; Yao et al. 2017a). Table 1 presents a summary of relevant studies on ET merging

**Table 1**
Literature review of relevant studies on merging multiple products and ground-based measurements to improve the spatiotemporal representation of ET at regional and global scales.

| Study | Merging models | Study domain | Products | Number of sites (Training/Validation) | Main conclusion |
|---|---|---|---|---|---|
| 1. Mueller et al. (2013) | Simple averaging (SA) method | Global | 40 products (7 diagnostic/29 LSMs/4 reanalysis products). | ET temporal variability comparison, without using ground measurements | The merged product reproduced a negative trend of global ET from 1998 to 2005, which supported the previous findings of the declining trend in global ET. |
| 2. Yao et al. (2014) | Bayesian model averaging (BMA), SA | Global | MOD16, RRS-PM, PT-JPL, MS-PT, UMD-SEMI. | 120 (50%)/120 (50%) | The RMSEs for five single LE products ranged from 39.4 W/m$^2$ to 45.6 W/m$^2$. The RMSEs for BMA and SA merged LE products were 35.3 W/m$^2$ and 37.4 W/m$^2$, respectively. |
| 3. Chen et al. (2015) | BMA | China | 8 products (5 empirical and 3 process-based products). | 23 sites used for validation | By examined the accuracies of 247 different possible combinations of eight products, they found that the four-product ensemble model performed best. This model increased the KGE by 4% and decreased the RMSE by 4%. |
| 4. Feng et al. (2016) | Empirical Orthogonal Function (EOF), SA | Global | MOD16, PT-JPL. | 22 sites used for validation | MOD16 and PT-JPL obtained comparable performances for different biomes. Despite the acceptable results from SA, the EOF algorithm showed notable improvement across all the biomes. |
| 5. Yao et al. (2017b) | Taylor skill fusion method | Global | RS-PM, SW, PT-JPL, MS-PT, SIM. | 103 (50%)/103 (50%) | The RMSEs of the products ranged from 27.1 W/m$^2$ to 29.8 W/m$^2$, the $R^2$ ranged from 0.55 to 0.59, and the Bias ranged from 5.5 W/m$^2$ to 10.1 W/m$^2$, while the merged product obtained an RMSE of 25.1 W/m$^2$, an $R^2$ of 0.60, and a Bias of 5.4 W/m$^2$. |
| 6. Jiménez et al. (2018) | Weight average method (WA), SA | Global | GLEAM, PT-JPL, PM-MOD. | 84 sites used for validation | WA performed better than any individual product and outperformed SA with 89/83 mm/yr (MSWEP/WorldClim) RMSDs and −64/−46 mm/yr Biases (while SA obtained 115/107 mm/yr RMSDs and −97/−80 mm/yr Biases). |
| 7. He et al. (2020) | Bayesian Three-Cornered Hat (BTCH), Ensemble Mean (EM) method | The contiguous United States | GFET, GLEAM, 4 NLDAS-2 datasets, 4 NLDAS-testbed datasets. | 15 sites used for validation | The BTCH method outperformed EM by capturing more accurate seasonal variations in ET, and BTCH obtained lower RMSD of 14.54 mm/month, while the RMSD of EM was 19.15 mm/month. |
| 8. Aires (2014) | Post Processing Filtering (PF) combined with merging methods including Simple Weighting, Constrained Linear, Optimal Interpolation and Neural Networks | Mississippi and Niger sites | The terrestrial water cycle dataset produced by Pan and Wood (2006). | 2 sites used for validation | All the four integration methods could combine with the PF step to impose water budget closures and achieve better estimation performances. The results showed that combination with PF could actually improve the calibration of original products. |
| 9. Yao et al. (2017a) | Support Vector Machine (SVM), BMA, General Regression Neural Networks (GRNNs) | Global | MOD16, PT-JPL, SEMI-PM. | 200 sites/tested by fourfold cross-validation | Three products yielded $R^2$ values of 0.66 to 0.67 and RMSEs of 23.71 to 28.11 mm/month. The BMA and GRNNs obtained $R^2$ values of 0.70 and 0.78 and RMSEs of 20.23 and 15.72 mm/month, respectively. The SVM yielded the best performance with an $R^2$ of 0.80 and RMSE of 14.71 mm/month. |
| 10. Shang et al. (2020) | Extremely Randomized Trees (ETR), Gradient Boosting Regression Tree, Random Forest, Gaussian Process Regression | Europe | RS-PM, SW, PT-JPL, MS-PT, SEMI-PM. | 39 (51%)/37 (49%) | The RMSEs of individual products ranged from 26.37 W/m$^2$ to 33.17 W/m$^2$. The RMSE of ETR was 16.87 W/m$^2$, while the RMSEs of the three other fusion products varied from 16.88 to 18.25 W/m$^2$. |

methods. Unfortunately, these merging methods depend highly on ground-based observations. In areas with sparse monitoring network (e.g., eddy covariance (EC) flux towers), it is extremely difficult to accurately model the spatial changes of ET due to the large uncertainties involved in merging procedures (Shang et al. 2020). Here both the locations and number of EC flux towers must be considered because of the important role the tower density plays during the merging process (Baez-Villanueva et al. 2020).

Long-term EC flux tower networks have been established at regional scales in the past three decades, such as AmeriFlux, AsiaFlux, BERMS (Boreal Ecosystem Research and Monitoring Sites), ChinaFlux, Watershed Allied Telemetry Experimental Research (WATER), and Heihe Watershed Allied Telemetry Experimental Research (HiWATER) (Baldocchi et al. 2001; Li et al. 2013a; Li et al. 2009a). Despite the unprecedented opportunity to measure ET directly from these ground observation networks, integrating in situ measurements with satellite-derived products to predict regional ET contains large uncertainties, especially in areas where the spatial autocorrelations and landscape heterogeneity are high (Kalma et al. 2008; Liang and Wang 2019). This is mainly because the topography directly affects the distributions of solar energy, meteorological conditions and microclimate that have direct regulations on ET (e.g., their changes with elevation, slope and aspect). Additionally, land cover type and structure directly influence microclimate, which in turn affects energy balance (Chen et al. 1999; Liang et al. 2010a). For remote sensing products, there also exist scale effects, i.e., mixed pixels from satellite observations contain complex surface information. Merging satellite-derived ET products without considering the impacts of surface properties can introduce additional difficulties in spatial predictions. Clearly, a spatial integration framework to combine all available information from satellite-derived ET products, EC observations and surface properties is needed for improving ET estimates at regional and global scales.

Deep Neural Networks (DNN) as the basis of deep learning methods have achieved great success in various applications, particularly in efficiently finding transformations that turn input data into representations for a given target (Chollet 2017). The depth of a DNN model always reaches tens/hundreds of multiple layers through parameterized nonlinear modules that can be trained by backpropagation (Zhou and Feng 2017). Substantial previous studies have demonstrated the ability of DNN in predicting a variety of ecological variables (Ienco et al. 2019; Saggi and Jain 2019; Yuan et al. 2020). We are convinced that DNN has great potential for integrating ET products, surface properties and ground observations to make more accurate spatiotemporal ET predictions despite the challenges presented by topography and climate variability.

Specifically, we developed a DNN-based merging ET (DNN-MET) framework to improve the representation of the spatiotemporal distribution of ET. Our objectives are to: (1) evaluate eight satellite-derived ET products and DNN-MET using ground measurements from 19 EC flux towers; (2) quantify the roles of ancillary surface variables, tower density and the spatial resolutions of ET products on model performances; (3) compare DNN-MET with different merging methods; and (4) implement the mapping of DNN-MET at the regional scale.

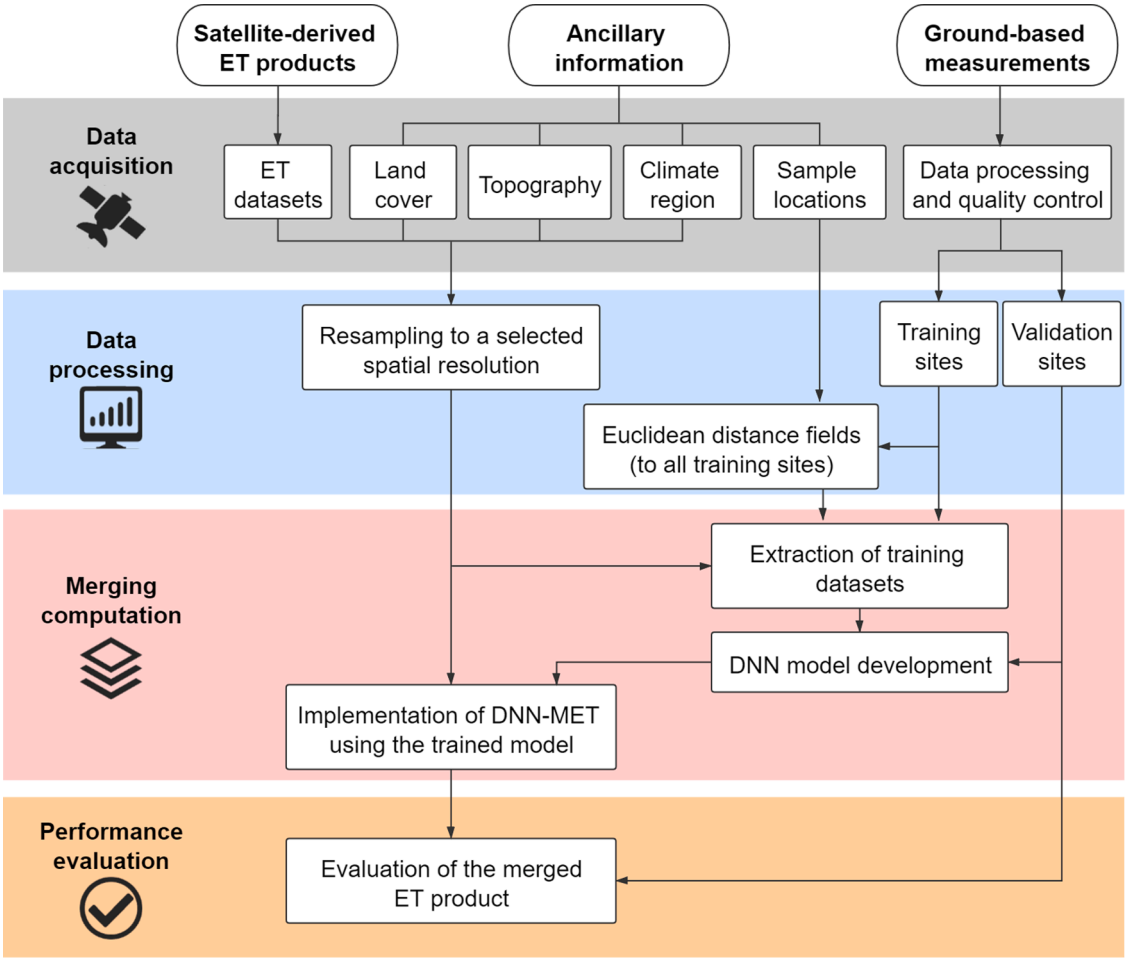


**Fig. 1.** Architectural framework and procedures of DNN-MET through integrations of satellite-derived ET products, ancillary information and ground-based measurements for an improved prediction of ET at regional scale.

## 2. DNN-MET

The traditional DNN models ignore surface properties in the ET spatial predictions, which can generate potential biases and suboptimal predictions. The under or over estimates are especially noticeable in areas where spatial autocorrelations of both predictors and dependent variables are high, or where landscape composition and structure are heterogeneous partially because the complex topography and climate (Hengl et al. 2018). To make accurate spatiotemporal ET predictions, proposed DNN-MET framework in contrast integrates information from satellite-derived ET products, ancillary surface properties and ground-based measurements (Fig. 1).

### 2.1. Data acquisition

Satellite-derived ET products and ancillary information are acquired as explanatory variables and EC measurements are acquired as predictor variables in our DNN models. The ancillary information consists of four components: land cover, topography (including elevation, slope and aspect), climate and sampling locations. The EC measurements were quality-controlled and aggregated from half-hourly observations into daily intervals.

### 2.2. Data processing

Satellite-derived ET products and ancillary surface properties were spatially resampled into a uniform resolution (0.01°) by bilinear interpolation to ensure identical raster geometry. We divided all 19 EC sites into training and validation categories to calibrate and evaluate the DNN model, respectively. We conducted two experiments in which (1) we evaluated the performance of our approach based on nineteen-fold cross validations (a.k.a. leave-one-out cross validation), which implies that each target EC site values are predicted based on the training using the remaining sites, and (2) evenly distributed in space, we selected 4 sites (approximately 20%) as independent validation sites to assess the uncertainty in spatial prediction of our approach for unknown locations.

To build a multivariate spatial ET model, DNN-MET incorporated geographical proximity effects into modeling process. The sample locations such as coordinates of samples used alone might not be sufficient to express geographical proximity effects. Thus, gridded layers of spatial Euclidean distance fields (Rosenfeld and Pfaltz 1968) to each EC training site were considered to be explicitly autocorrelated indicators and were used as explanatory variables in DNN-MET.

### 2.3. Merging computation

DNN was developed based on the shallow neural networks with multilayer perceptions (MLPs) (Ivakhnenko and Lapa 1966). DNN emphasize successive layers of representations. The number of contributing layers in a model is called the model "depth". Approaches based on shallow neural networks often employ learning at one to two layers. In contrast, a DNN model performs the input-to-target transformation via a deep sequence of layers, with the depth reaching tens to hundreds layers (Chollet 2017).

Our DNN model was set with a suite of hyperparameters in multiple layers of network, loss functions, and an optimizer (supporting information in Fig. S1). The weight for a parameter of a layer indicates the essential the specifications for each input during the transformation. To assess the performance of a DNN model, a loss function, i.e., an objective function, is required to measure how far the predictions are from the in situ observations (i.e., the targets). The optimizer uses the distance score as a feedback signal to adjust the weights of the layers in the network (Chollet 2017). This parameter optimization process is implemented by the backpropagation algorithm (Benvenuto and Piazza 1992), which is the central algorithm in deep learning. For each training loop, the loss score decreases with updated weights. A DNN model with minimal loss is selected when the outputs are as close as possible to the ideal target for the trained network.

### 2.4. Performance evaluation

The following metrics are used to evaluate the performance of our models and modeling approach. The coefficient of determination ($R^2$) characterizes how much of the variation in the ground measurements is explained by the model; the root-mean-square error (RMSE) measures the closeness between the ET predictions and observations; the Bias value quantifies the difference between the averages of the estimates and observations; the Kling-Gupta efficiency (KGE) (Gupta et al. 2009) and Taylor diagrams (Taylor 2001) are used to comprehensively assess the model performance. KGE incorporates the correlation ($r$), relative variability ratio ($\alpha$) and mean values ratio ($\beta$) into a single multi-objective criterion. KGE allows a better understanding of the origin of mismatches as:

$$KGE = 1 - ED \tag{1}$$

$$ED = \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2} \tag{2}$$

$$\alpha = \frac{\sigma_e}{\sigma_o} \tag{3}$$

$$\beta = \frac{\mu_e}{\mu_o} \tag{4}$$

where $ED$ reflects the Euclidian distance from the ideal value; $\sigma_e$ and $\sigma_o$ are the standard deviation of the estimations and observations, respectively; and $\mu_e$ and $\mu_o$ are the mean values of the estimations and observations, respectively. The ideal values of the three components, obtained when there are no simulation errors, are $r = 1$, $\alpha = 1$ and $\beta = 1$; thus, the ideal value of KGE is 1.

A Taylor diagram is a polar-style graph that provides a concise statistical summary of how well patterns match each other in terms of their correlation coefficient (r), root-mean-square difference (RMSD) and standard deviation (SD). Taylor diagram provides a graphical framework for comparing a suite of variables from one or more predictions to reference data. The variables can be the ET estimates from one or more models, whereas the reference is the EC ground measurements.

## 3. Case study

The Heihe River Basin (HRB) is the second-largest inland river basin in China and is used as our testbed to assess the performance of DNN-MET. HRB has heterogeneous land mosaics, topography and climate.

### 3.1. Study domain

The HRB (approximately 143200 $km^2$) is located between 97.1°−102.0°E and 37.7°−42.7°N in the arid and semiarid region of northwestern China (Fig. 2). The Heihe River flows through three provinces in China. The upstream originates from the Qilian Mountains in Qinghai Province, the midstream flows through the Hexi Corridor in Gansu Province, and the downstream ends in two terminal lakes in western Inner Mongolia (Xiong et al. 2015; Yao et al. 2019). With terrain elevations varying from 900 m to 5500 m, the HRB is characterized with the alternation of high mountains and basins (Cheng et al. 2014) (Fig. 2).

The upstream of the HRB is characterized with cold and mountainous with elevations ranging from 2000 m to 5500 m. Precipitation, snowmelt and glacier melt are the primary water sources of the upstream region. The dominant land cover types include grassland (alpine meadow), shrubland (valley bush), evergreen needleleaf forest (*Picea crassifolia*), wetland (swamp), and snow/ice (frozen soil, snow, glaciers). The precipitation in the middle and downstream regions is low, resulting

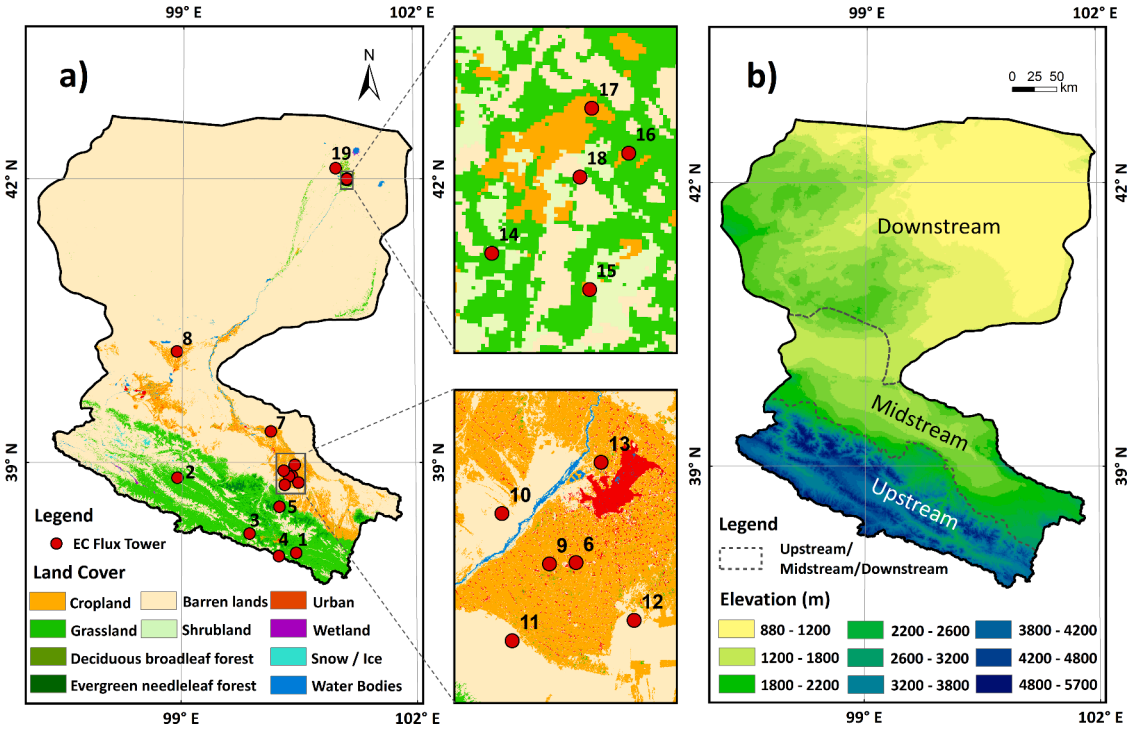


**Fig. 2.** Study domain. a) Land cover map of the Heihe River Basin (HRB) and the locations of eddy covariance (EC) flux sites. b) Elevation of the HRB. The sub-catchment boundary denote the upstream, midstream and downstream regions.

in these areas to be a net consumption for water. The midstream area is dominated by cropland (maize, spring wheat, vegetables), barren lands and urban/built-up land, with irrigation widely practiced. The downstream region is a composite of natural oases and the arid Gobi Desert that include barren lands (desert), water bodies (terminal lakes) and natural oases. The natural oases are composed of shrubland (Tamarix), cropland and deciduous broadleaf forests (Li et al. 2013a; Liu et al. 2018). The annual air temperature increases from 0.12°C in the upstream to 6.98°C in the midstream, and to 9.49°C in the downstream. The annual precipitation is 400-550 mm, 100-160 mm and 30-40 mm in the upstream, midstream, and downstream regions, respectively (Xu et al. 2020).

### 3.2. Data collection

#### 3.2.1. Eddy covariance flux measurements

Latent heat flux measurements from 19 EC flux tower sites were collected from WATER (Li et al. 2009a) and HiWATER (Li et al. 2013a) experiments over the HRB. Both experiments were designed with comprehensive hydrometeorological observatory networks across the three reaches. Five EC flux tower sites are located upstream, eight are located midstream and six are located downstream. Data from these EC sites provides us with rich information to capture a wide range of land surface conditions (see details in Table S1). They cover seven land cover types: grassland (GRA), evergreen needleleaf forest (ENF), cropland (CRO), barren lands (BRA), wetland (WET), deciduous broadleaf forest (DBF) and shrubland (SHR). The climate for these flux tower locations varies from snow to dry and steppe to desert across the whole river basin. All EC measurements were filtered and quality controlled (Liu et al. 2018; Xu et al. 2020). The daily ET observations were aggregated from half-hourly measurements for a consistent temporal resolution of ET. The energy closure imbalance (Foken 2008) of the EC observations was corrected by the Bowen ratio closure method (Twine et al. 2000).

We divided the EC sites into training and validation groups. In addition to nineteen-fold cross-validations (or leave-one-out cross-validation), we also conducted an independent validation experiment by selecting 20% of EC sites (i.e., 4 sites) to evaluate the spatial prediction performance of our approach for unknown locations. These validation sites were evenly distributed in space and represent the characterized land cover types of the three reaches in the HRB as: Hulugou (upstream, GRA), Linze (midstream, CRO), Huazhaizi (midstream, BAR) and Populus euphratica (downstream, DBF).

#### 3.2.2. Satellite-derived ET products

We merged eight available daily ET products during 2008-2015, including the MOD16 algorithm framework-based ET product (MOD16) (Mu et al. 2011), semiempirical Penman algorithm-based ET product (ET-SEMI) (Wang et al. 2010), Priestley-Taylor of Jet Propulsion Laboratory based ET product (ET-JPL) (Fisher et al. 2008), modified satellite-based Priestley-Taylor ET product (ET-MS) (Yao et al. 2013), ET product based on hybrid formulation parameterized by soil moisture (ET-HF) (Yao et al. 2019), Global Land Evaporation Amsterdam Model datasets (GLEAM) (Martens et al. 2017; Miralles et al. 2011), monitoring of ET by a hybrid remotely sensed model (ETMonitor) (Hu and Jia 2015) and surface energy balance-based global land evapotranspiration (EB-ET) (Chen et al. 2019a) products.

These global or quasi-global ET products were used because: (1) they were produced by well-known process-based or empirical flux equations, such as the Penman–Monteith equation (Monteith 1965) (e.g., MOD16, ET-SEMI), the Priestley–Taylor algorithm (Priestley and Taylor 1972) (e.g., ET-JPL, ET-MS, GLEAM), the Surface Energy Balance System model (Su 2002) (e.g., EB-ET), the Shuttleworth–Wallace Dual-Source model (Shuttleworth and Wallace 1985) (e.g., ETMonitor) and the empirical hybrid formulation (Wang et al. 2010; Yao et al. 2019) (e.g., ET-HF); (2) the input forcing data of these ET products included meteorological data, vegetation indexes, soil moisture (SM), land surface temperature (LST), vegetation optical depth (VOD) and other biophysical parameters derived from microwave and optical remote sensing observations, which are highly and intrinsically correlated with ET process; and (3) these products perform well on global and regional scales (Chen et al. 2014; Ershadi et al. 2014; Kalma et al. 2008). In addition, we used Global LAnd Surface Satellite (GLASS) ET product

(Liang et al. 2020; Yao et al. 2014) as comparison to assess the performance of DNN-MET. Information on these products is summarized in supporting information in Table S2.

**(1) MOD16.** The MOD16 product used in this study was generated based on the MODIS global ET algorithm framework proposed by Mu et al. (Mu et al. 2011). Since the latest MOD16A2 Version 6 global ET product is an 8-day composite dataset and exists too many missing values over the HRB, which limit the continuously spatial and temporal predictions of DNN-MET. Therefore, we produced a daily ET product based on the MOD16 algorithm (Text S1) over the HRB. In order to keep high consistencies with MOD16A2 ET product, we used the same input variables with the original product: the global GMAO meteorological data, and the leaf area index (LAI) and fraction of absorbed photosynthetically active radiation (FPAR) from MODIS. The difference between the MOD16 we produced and the original product is that we retained the ET estimates for non-vegetated areas. The ET product has a 0.01° spatial resolution and spans from 2008 to 2015.

**(2) ET-SEMI.** The ET-SEMI product was produced by a semiempirical Penman-based method developed by Wang et al. (Wang et al. 2010). The coefficients were determined and validated using continuous measurements collected at 64 global sites. The empirical coefficients used in the Penman equation make it possible to include dependencies on vegetation and soil moisture by using a range of climate conditions (Text S2). We generated the ET-SEMI product using downward shortwave radiation ($R_s$), $T_a$, wind speed (WS) and RH data derived from CMFD; and the normalized difference vegetation index (NDVI) from MODIS. The ET-SEMI covers the HRB during the period of 2008-2015 at a 0.01° spatial resolution and a daily temporal resolution.

**(3) ET-JPL.** The ET-JPL was calculated by a novel Priestley–Taylor equation proposed by Fisher et al. (Fisher et al. 2008). This biometeorological method translates potential ET from the Priestley–Taylor equation into actual ET without using calibration from intensive field measurements (Fisher et al. 2008) (Text S3). The input variables used to calculate ET-JPL included $R_n$, $T_a$, and RH derived from CMFD; NDVI derived from MODIS and LAI derived from GLASS. The ET-JPL is available from 2008 to 2015 over the HRB with a daily temporal resolution and a 0.01° spatial resolution.

**(4) ET-MS.** The ET-MS product was based on a modified satellite-based Priestley–Taylor algorithm proposed by Yao et al. (Yao et al. 2013). ET-MS separated ET into four components: canopy transpiration, vegetation interception evaporation, unsaturated soil evaporation, and saturated wet soil evaporation (Text S4). Avoiding the complicated computation of aerodynamic resistance parameters (Yao et al. 2013), ET-MS only requires $R_n$, $T_a$, and DT derived from CMFD, and NDVI from MODIS as inputs. The ET-MS product covering the HRB has the same spatial and temporal resolution as does ET-JPL over the period from 2008 to 2015.

**(5) ET-HF.** The ET-HF product was generated using an empirical hybrid formulation parameterized by SM (Yao et al. 2019). The SM constraint used in the ET algorithm is vital to the understanding of regional surface energy and water cycle processes in the HRB (Yao et al. 2019). We produced ET-HF using $R_n$ and $T_a$ derived from CMFD; NDVI derived from MODIS and SM derived from the European Space Agency Climate Change Initiative (ESA CCI) program (Text S5). The ET-HF is available at a 0.01° spatial resolution and a daily temporal resolution over the HRB from 2008 to 2015.

**(6) GLEAM.** The GLEAM ET product was calculated by complex land surface models including a Priestley-Taylor based potential ET module, a stress module, a soil module and a rainfall interception model (Martens et al. 2017). Transpiration, bare-soil evaporation, interception loss, open-water evaporation and sublimation were estimated separately (Text S6). The latest global daily/monthly/yearly GLEAM ET products with 0.25° spatial temporal resolution are available at www.gleam.eu. We used GLEAM v3.3 a version in this study, and the 0.25° daily GLEAM data extracted over the HRB was interpolated into the target grid size of 0.01° by the bilinear method.

**(7) ETMonitor.** The ETMonitor monitors daily ET over the HRB by combining microwave and optical remote sensing observations for all sky conditions (Hu and Jia 2015). As Text S7 shows, the soil evaporation and vegetation transpiration were calculated by the Shuttleworth–Wallace (Shuttleworth and Wallace 1985) model, and the rainfall interception was calculated based on the Gash model. ETMonitor requires Radiation (Rad), $T_a$, Precipitation (Pres), RH, WS, P, NDVI, SM, and a land cover map of the HRB as inputs. The ETMonitor used in our study, which has a 1-km spatial resolution and a daily temporal resolution, was obtained from the National Tibetan Plateau Data Center (https://data.tpdc.ac.cn/zh-hans/).

**(8) EB-ET.** EB-ET is a surface energy balance-based global ET product derived from satellite data. The global daily and monthly ET dataset is produced with the revised Surface Energy Balance System (SEBS) algorithm proposed by Chen et al. (Chen et al. 2019a; Chen et al. 2013) (Text S8). LST, NDVI, global forest height, albedo and LAI were used in this ET calculation. We obtained daily EB-ET over the HRB with a spatial resolution of 0.05° from the National Tibetan Plateau Data Center (https://data.tpdc.ac.cn/zh-hans/) and interpolated it to a 0.01° spatial resolution by the bilinear method.

**(9) GLASS ET.** GLASS ET is a merged ET product using BMA method based on five process-based ET algorithms (Text S9), which provides a benchmark for various terrestrial ET products (Liang et al. 2020; Yao et al. 2014). The forcing data for GLASS ET mainly include meteorological datasets from GMAO-MERRA, and Albedo, LAI, NDVI, FPAR from MODIS/AVHRR. We obtained 8-day GLASS ET with 0.01° resolution (MODIS driven) from the following websites: http://www.glass.umd.edu.

#### 3.2.3. Ancillary data

Ancillary data such as surface properties are essential to make a DNN more applicable for spatial predictions because they quantify the proximity and geographical connections between points in space. We selected four surface properties as ancillary variables to enhance spatial expression in the modeling procedure: land cover, topography (elevation, slope and aspect), climate region and sample locations. They were interpolated to 0.01° to have the same spatial resolution of ET products.

**(1) Land cover.** The 30-m land cover map of the HRB generated by Zhong et al. (Zhong et al. 2015) that is archived at the National Tibetan Plateau Data Center (https://data.tpdc.ac.cn/zh-hans/). The classifications are more detailed and visually better than other land cover products (Fig. 2). Based on evaluations against the high spatial resolution remote sensing data within Google Earth and data from ground campaigns, the overall accuracy was 92.2% (Zhong et al. 2015).

**(2) Topography.** The topography-related datasets were derived from the Shuttle Radar Topography Mission (STRM) DEM dataset with a spatial resolution of 250 m. The elevation over the HRB was extracted directly from the global STRM, and the slope and aspect were calculated from the elevation.

**(3) Climate region.** The climate region data were downloaded from the Köppen-Geiger climate classification (Kottek et al. 2006). The Köppen-Geiger climate map is the most frequently used climate classification map and is provided with a spatial resolution of 0.5°. The HRB covers three climate classifications: snow/dry winter/cool summer (Dwc), dry/steppe/cold arid (BSk) and dry/desert/cold arid (BWk).

**(4) Sample locations.** The sample locations were represented by the Euclidean distance fields (EDFs) to site. The EDFs of the training sites were used to generate buffer distance maps of sample locations, and one EDF map corresponds to one training site.

### 3.3. Merging procedure

There are components in our architecture of the merging procedure (Fig. 3). Ground-based measurements from the training EC sites, corresponding grid-cell values from the ET products and ancillary variables were used in the DNN training process. With the well trained DNN

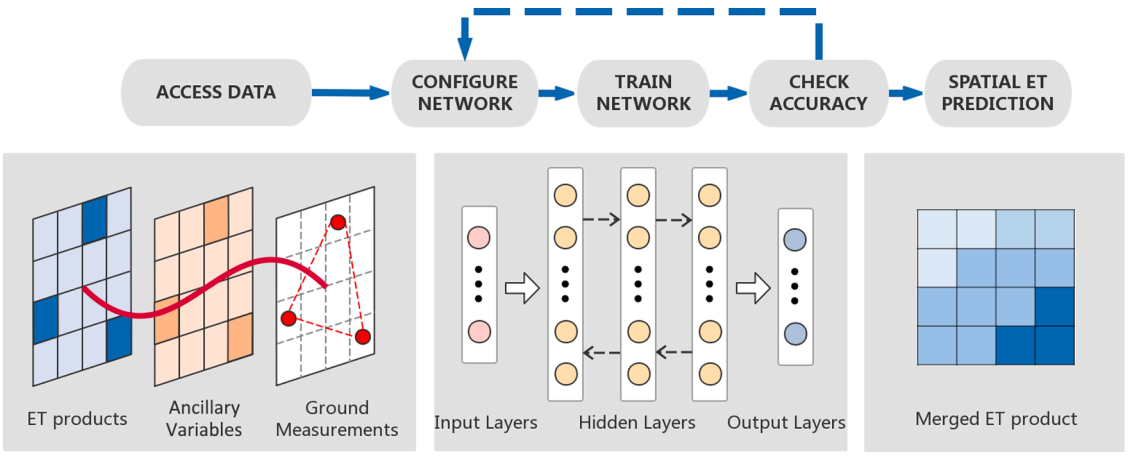


**Fig. 3.** Schematic of the merging procedure using ET products, ancillary variables and ground measurements.

model, we implemented spatiotemporal ET predictions.

We constructed our DNN model with Keras and TensorFlow backend in Python. TensorFlow is an end-to-end, open-source machine learning platform that efficiently executes tensor operations. Keras is a high-level deep learning application programming interface (API) in TensorFlow that focuses on providing essential abstractions and building blocks for enabling fast experimentation from idea to result with a high iteration velocity. We constructed a five-layer that fully connectes DNN model to implement the merging of the satellite-derived ET products, ancillary information and ground-based measurements. To reduce the risk of overfitting, we restricted the hidden layers to three (64, 32, 32) and adopted a 'rectified linear units' (relu) activation function for the first hidden layer and two 'linear' activation functions for the other two hidden layers. We adopted the 'Adam' optimizer and the mean squared error (MSE) loss function to compile the DNN model. The training terminal was set as 2000 echoes, and the RMSE was used as the cost function.

### 3.4. Evaluation methods

We assessed the accuracy of our merging method with $R^2$, RMSE, Bias, KGE and Taylor diagrams. The estimates from DNN-MET and individual ET products were directly compared with EC ground measurements at 19 flux tower sites.

To evaluate the spatial prediction performance for unknown locations in site-scarce regions, we investigated the influence of the number of EC sites in the training dataset. We generated DNN-MET estimations using 80%, 60%, 40% and 20% of the 19 sites, representing 15, 11, 8 and 4 training EC flux towers, respectively (Table 2). We chose Arou, Guantan, Daman and Sidaoqiao as the last 20% sites because they are the least site collocations that can represent the main plant function types and site spatial distribution of the whole basin. The merged ET products produced by the incorporation of these varying percentages of training sites were used to evaluate the performance of DNN-MET under different data-scarcity scenarios.

In addition to the creations of a DNN-MET at 0.01° spatial resolution, we also computed DNN-MET at 0.05°, 0.1° and 0.25° spatial resolutions to investigate the influences of spatial resolutions of the input ET products. Additionally, we merged eight ET products using three algorithms: Random Forest (RF), Bayesian Model Averaging (BMA) and a simple averaging (SA) method. These fusion ET products from different merging methods and GLASS ET product were also compared with DNN-MET. Finally, to test the spatial expression performance, we compared the maps of annual average ET over the HRB from 2012 through 2015 from DNN-MET and the eight ET products.

**Table 2**
Summary of EC flux tower sites used for the varying percentages of total available sites in the training of DNN-MET.

| Percentage of training sites | Training sites | Number of training sites over up-/mid-/downstream | Validation sites |
|---|---|---|---|
| 80% | Arou / Dashalong / Yakou / Guantan / Yingke / Jinta / Daman / Bajitan / Shenshawo / Zhangye wetland / Mixed forest / Sidaoqiao / Cropland / Barren land / Desert | 4 / 6 / 5 | Hulugou / Linze / Huazhaizi / Populus euphratica |
| 60% | Arou / Dashalong / Guantan / Yingke / Jinta / Daman / Shenshawo / Zhangye wetland / Mixed forest / Sidaoqiao / Barren land | 3 / 5 / 3 | |
| 40% | Arou / Dashalong / Guantan / Jinta / Daman / Shenshawo / Mixed forest / Sidaoqiao | 3 / 3 / 2 | |
| 20% | Arou / Guantan / Daman / Sidaoqiao | 2 / 1/ 1 | |

## 4. Results

### 4.1. Evaluation of satellite-derived ET products

The Taylor diagrams for daily ET observations and ET estimates from individual products for different land cover types indicate that ETMonitor and ET-HF estimates are closer to the observations (Fig. 4), implying that ETMonitor and ET-HF outperform the other products. ETMonitor has the highest r value of 0.83 and the lowest RMSD of 0.92 mm/day, followed by ET-HF with a r of 0.79 and a RMSD of 1.05 mm/day. The correlation coefficients for the other six ET products range from 0.58 to 0.78, and the RMSDs vary from 1.13 to 1.41 mm/day. Due to differences in precipitation and vegetation coverage, the eight satellite-derived products show large inter-biome discrepancies (Fig. 4).

ET-HF demonstrates the best performance at the GRA and DBF with the highest r values (0.90 and 0.88) and the lowest RMSD values (0.62 and 0.95 mm/day), respectively. For the GRA sites, ET-SEMI, GLEAM and ET-JPL outperform ETMonitor, MOD16 and EB-ET, with r values ranging 0.87-0.88 and RMSD of <0.79 mm/day. For the DBF sites, ETMonitor reached the second-highest r value of 0.85 and a low RMSD

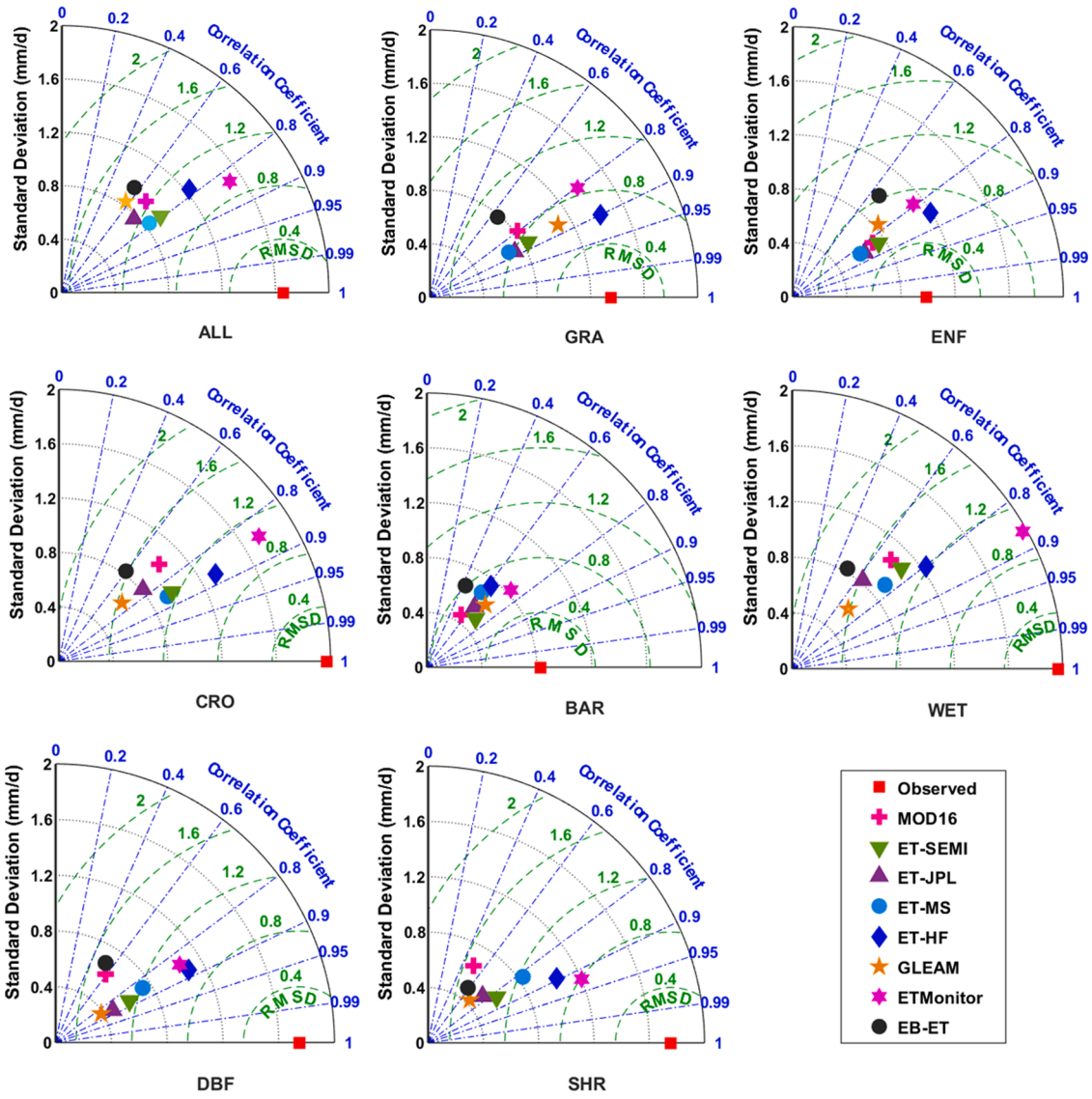


**Fig. 4.** Taylor diagrams showing the correlation coefficient (R), root mean square difference (RMSD) and standard deviation (SD) values of multiple satellite-derived ET products and ground measurements for different land cover types. ALL: all land cover types, GRA: grassland, ENF: evergreen needleleaf forest, CRO: cropland, BAR: barren land, WET: wetland, DBF: deciduous broadleaf forest and SHR: shrubland.

of 1.02 mm/day, making it superior to the other six ET products. ETMonitor exhibits the best performances at WET and SHR, with the highest r values (0.87 and 0.92) and the lowest RMSD values (1.01 and 0.79 mm/day), respectively. For WET and SHR, ET-HF outperforms the other six products, with the second-highest r values of 0.80 and 0.89, and low RMSD values of 1.21 and 0.95 mm/day, respectively.

For ENF sites, ET-SEMI, ET-JPL, ET-HF and ET-MS yield good performances, with r values ranging from 0.84 to 0.85 and RMSD values varying from 0.53 to 0.63 mm/day. For CRO sites, ET-HF and ETMonitor have higher accuracies, with higher r values (0.87 and 0.85) and lower RMSD values (1.07 and 1.06 mm/day) than the other products. For BAR sites, ETMonitor and ET-SEMI outperform the other products, with r values of 0.74 and 0.71 and RMSD values of 0.60 and 0.59 mm/day, respectively.

Overall, the SDs of all ET products are lower than those of the observations, implying that individual ET products make suboptimal predictions and have the same underestimation trend in the HRB. Additionally, the eight satellite-derived ET products show substantial discrepancies among land cover types, with none of the individual ET products providing the most accurate ET estimates over the HRB.

### 4.2. *Evaluation of the merged ET product*

To evaluate the performance of the merged ET product, we conducted spatial and temporal assessments of DNN-MET based on nineteen-fold cross-validations (i.e., leave-one-out cross-validation). Here each target estimate is predicted by the trained model based on the remaining sites.

The comparison between ground-measured and estimated ET demonstrates that our approach accurately estimates daily and seasonal ET at all 19 EC sites (Fig. 5). For daily ET estimates, the cross-validation KGE of DNN-MET is 0.83, the $R^2$ is 0.76 ($p < 0.01$), and the Bias and RMSE are -0.02 mm/day and 0.81 mm/day, respectively. DNN-MET outperforms the most accurate individual ET product (ETMonitor with KGE of 0.77). The seasonal variation from DNN-MET appears robust, with the KGE of 0.89, $R^2$ of 0.85 ($p < 0.01$), Bias of -0.47 mm/month, and RMSE of 17.12 mm/month, all better than those of the individual ET products. DNN-MET also seems better in predicting the among-site variability; the KGE of the site-averaged estimates versus the observations is 0.84, the $R^2$ value is 0.91 ($p < 0.01$), and the Bias and RMSE values are -0.05 mm/day and 0.37 mm/day. Similarity, our approach

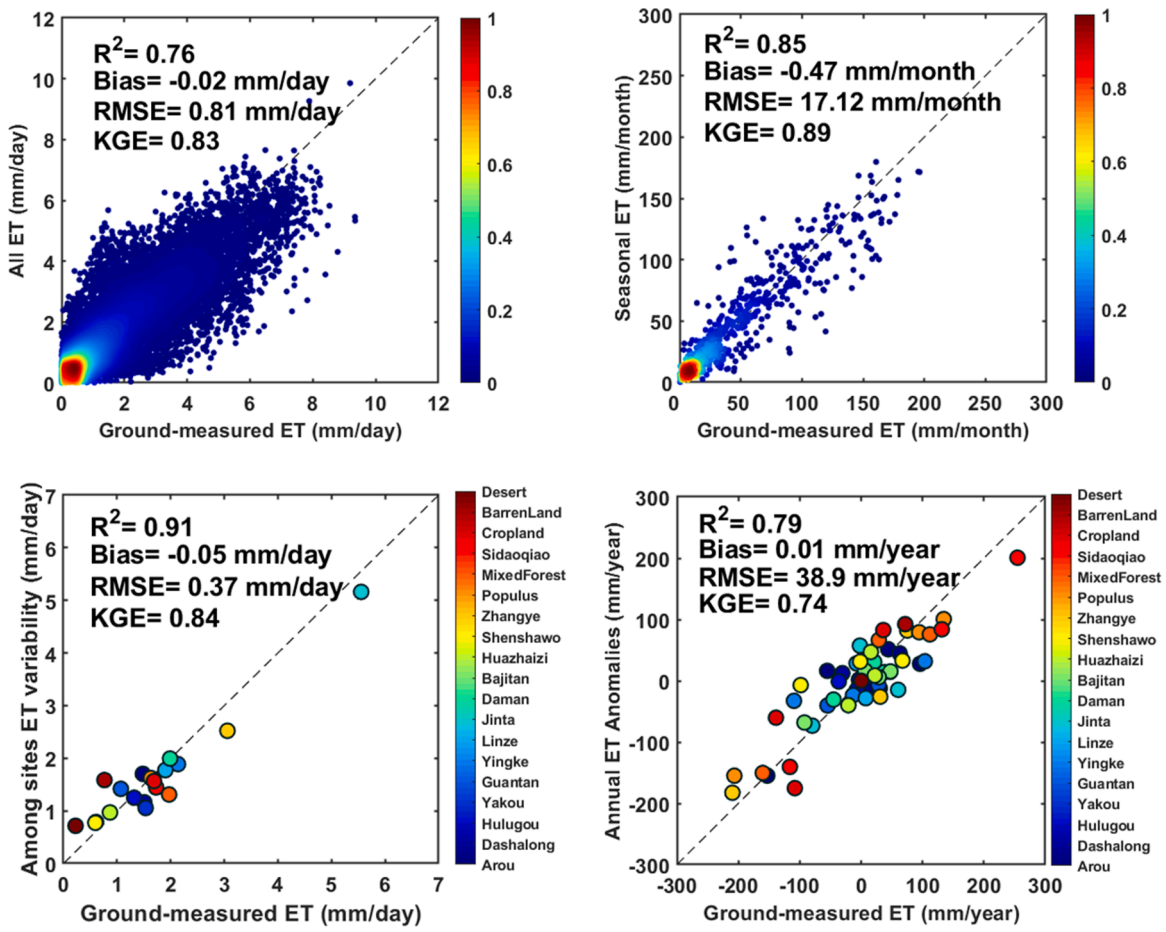


**Fig. 5.** The estimated ET (vertical axis) versus the ground-measured ET (horizontal axis) based on nineteen-fold cross-validation (i.e., leave-one-out cross-validation) for all ET, seasonal ET, among-site ET variability and annual ET anomalies.

appears more satisfactory in reproducing the interannual variability (Fig. 5). The KGE of the estimates versus the ground-measured ET annual anomalies is 0.74, the $R^2$ value is 0.79 ($p < 0.01$), and the Bias and RMSE values are as low as 0.01 mm/year and 38.9 mm/year.

Fig. 6 presents the spatial distributions of the KGE values at each EC site for all ET products over the HRB. The eight individual ET products show lower KGE values in the downstream region with similar trends. These regions are classified as dry/desert/cold arid according to the Köppen-Geiger classification, implying that the performances of the ET products in the cold, arid desert regions remain low. All the individual products exhibit increasing performance in upstream, where precipitation, snowmelt and glacier melt are much higher.

Fig. 6 shows the overall KGE values and the number of sites within each KGE range for the eight ET products and DNN-MET. DNN-MET demonstrates the best performance with the highest total KGE value of 0.83, and the overall KGE values of the eight products varied from 0.06 to 0.77 (Fig. 6). For DNN-MET, the KGE values with 9 sites vary from 0.6 to 0.8, and with 5 sites showing >0.8, indicating that more than 70% of the EC sites yield KGE values higher than 0.6. With no sites yielding KGE values lower than 0.2 (3 sites in the range of 0.4 to 0.6, 2 sites in the range of 0.2 to 0.4), DNN-MET provides the most accurate ET predictions. DNN-MET considerably outperforms all the products used in the merging procedure in the middle and downstream regions and maintains or even enhances the high accuracies from individual products in the upstream region, highlighting the advantages of combining information from ET products, ground-based measurements and ancillary variables.

### *4.3. Impact of EC flux tower density and the spatial resolution of the ET products*

To understand the influence of the tower density in the training dataset, 80%, 60%, 40% and 20% of the EC sites were used to construct our DNN-MET, and 20% of the EC sites (i.e., 4 sites) were used to evaluate the spatial prediction performance for unknown locations (Table 2). Fig. 7 presents the performance of DNN-MET at the 4 validation sites with varying numbers of EC sites used in the training set. The median KGE values of DNN-MET with varying numbers of training sites range from 0.79 to 0.87, while the median KGE value of the best available product (ETMonitor) is 0.73. The $R^2$, RMSE and Bias of DNN-MET obtained with scarce training data are also superior to all the ET products used in the computation. As expected, inclusion of more EC sites in the training of DNN-MET improves the product performance. An improvement is obtained even when only 20% (4 sites) of EC sites are used, suggesting that our approach can be effectively used in other data-scarce regions to improve the spatiotemporal representation of ET.

To examine the impacts of spatial resolution in the merging procedure, the satellite-derived ET products and ancillary information were resampled to several resolutions to generate DNN-MET by using 80% of the EC sites at 0.01°, 0.05°, 0.1° and 0.25°. Fig. 8 presents the KGE values of DNN-MET with the four spatial resolutions vary from 0.73 to 0.86, whereas those of the best-performing products (ETMonitor/ET-HF) vary from 0.46 to 0.70. The results indicate that input variables with coarser resolutions will affect the performance of the merging computation, mainly caused by uncertainties of spatially-mismatched pixels between the ground-based observations and the satellite-derived data sources. But DNN-MET yielded improvement performance with even

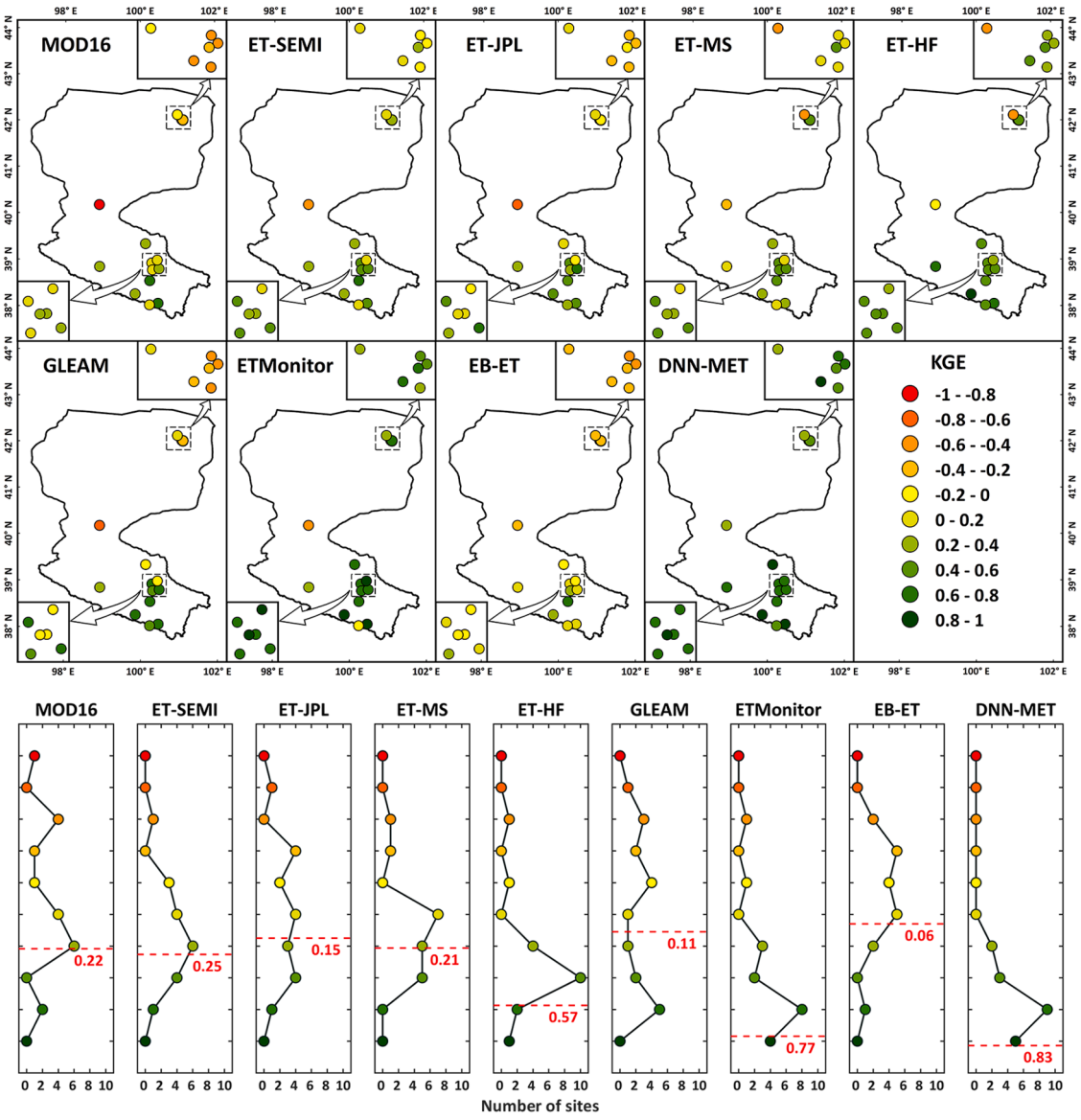


**Fig. 6.** Spatial distribution of the individual KGE values at each EC site for all ET products. The dot plots in the bottom present the number of sites within each KGE range. The red dotted lines indicate the overall KGE value for each ET product at all EC sites.

coarser input variable spatial resolutions.

### 4.4. Comparison between DNN-MET and different merging methods

Fig. 9 presents our cross-validations results of DNN-MET and different merging methods (i.e., RF, BMA and SA). The best-performing product used in the merging computation is also plotted as the benchmark for the comparison of the merging methods. ET-HF has the highest $R^2$ median value (0.78), while ETMonitor achieves the lowest RMSE (0.85 mm/day) and Bias (-0.10 mm/day) values and the highest median KGE (0.66) value (Fig. 9). In sum, DNN-MET yielded the best performance compared with the other merging methods as evidenced by the median KGE value of 0.71, the median $R^2$ value of 0.82 ($p < 0.01$), and the median RMSE and Bias values of 0.67 and -0.04 mm/day, respectively. These values are significantly superior to those of the other products. The RF-merged product performed similarly to DNN-MET, but DNN-MET showed low dispersion of the $R^2$, RMSE, Bias and KGE values at the 19 EC sites. BMA performed better than SA, with lowered RMSE and Bias values, but both methods obtained negligible improvements compared with the best-performing ET product used in the merging computation.

The comparison of performances between our approach and GLASS ET at 8-day spatial resolution shows that DNN-MET yielded superior performance over HRB (Fig. S2 in supporting information). The evaluation against all EC sites shows that for DNN-MET, the median value for KGE is 0.77, for $R^2$ is 0.90 ($p < 0.01$), for RMSE and Bias are 0.48 mm/day and -0.14 mm/day, respectively (The red dotted lines in Fig. S2). Whereas those of GLASS ET, the median values for these assessment metrics are 0.22, 0.61, 1.15 mm/day and -0.68 mm/day (The blue dotted lines in Fig. S2). Similar to other daily satellite-derived ET products, the two 8-day resolution ET products showed better performance at upstream, but decreased accuracies at midstream and downstream. At upstream, the KGE values of DNN-MET range from 0.73 to 0.95, while the values of GLASS ET vary from 0 to 0.57. At midstream, the KGE values of DNN-MET vary from 0.68 to 0.91, whereas those for GLASS ET range from -0.29 to 0.56. At downstream, the KGE values of DNN-MET are from 0.19 to 0.88, and the values for GLASS ET vary from -0.07 to 0.52.

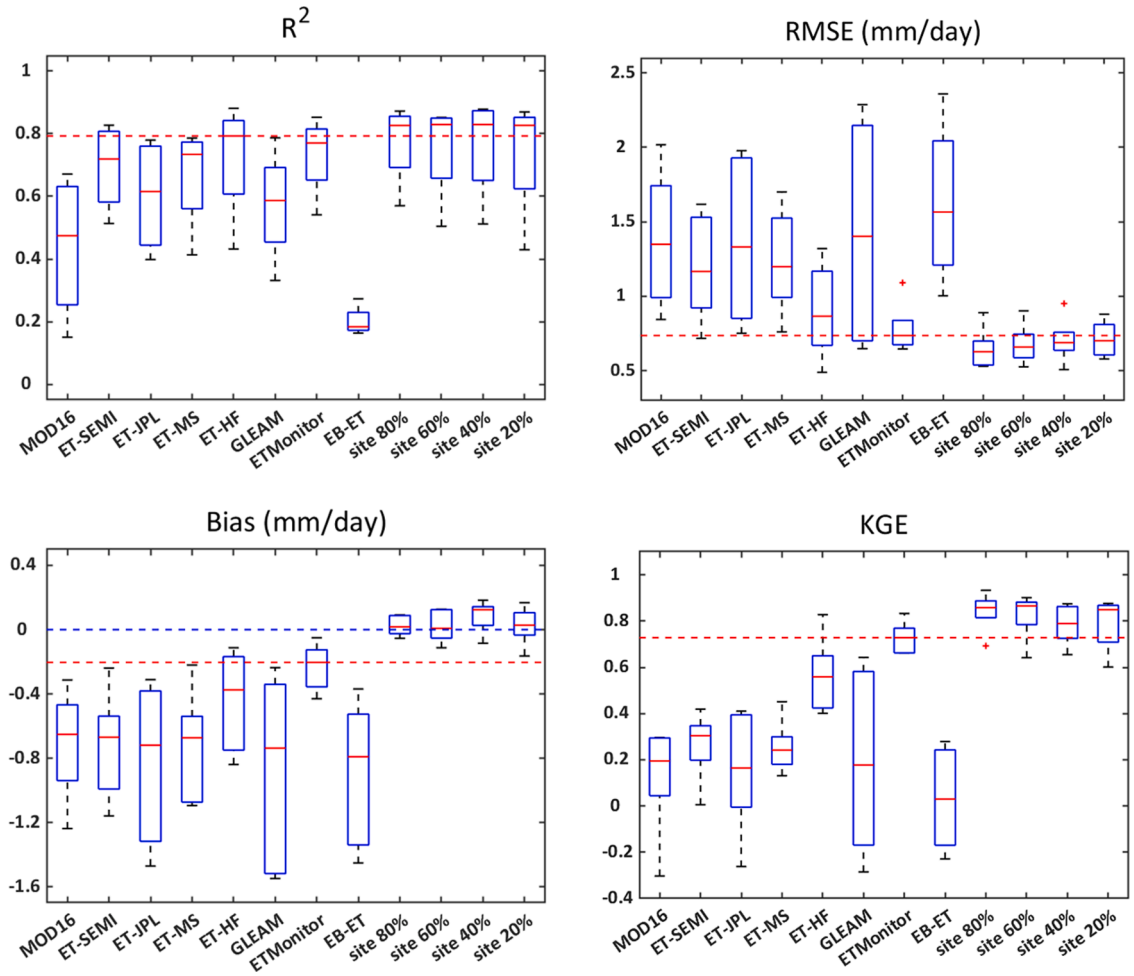


**Fig. 7.** Performance of the DNN-MET at 4 validation sites using varying percentages of flux tower sites in the training set. The red solid lines indicate the median values, the box edges indicate the first and third quartile values, the end of the whiskers indicate values no more than 1.5 times the interquartile value, and the most extreme data points beyond the whiskers indicate the extreme outliers. The red dotted lines indicate the median values of the best-performing product used in the computation of DNN-MET. The blue dotted line indicates the optimal Bias value for all ET products.

### *4.5. Mapping of DNN-MET over the Heihe River Basin*

Fig. 10 presents the spatial distribution of the multiyear (2012–2015) mean ET over the HRB. Despite the differences in spatial resolution (GLEAM with a resolution of 0.25°, EB-ET with 0.05° and other products with 0.01°) and the different algorithms used for ET estimates, all ET products showed highly-consistent spatial patterns over the basin. Significantly higher annual ET values were apparent for the region and in the oases of the mid- and downstream regions. With the increase of latitude, the annual ET decreases sharply along a gradient from south to north (i.e., upstream to downstream). The ET of the oases in the midstream region is the highest, likely due to the intensive irrigations. It is worth noting that with the varying of altitude, ET estimates showed consistent spatial distributions with the topography and landscape in the HRB.

There remain some discrepancies between the different ET product estimates (Fig. 10). The multi-year (2012–2015) mean estimate of HRB from DNN-MET is 228.79 mm/year. ET-HF is the closest estimates to those of DNN-MET, with an annual average ET of 223.46 mm/year. Seven other products yielded lower estimations, with annual average ET values of 80.29-141.86 mm/year. Based on DNN-MET, the multiyear mean ET of the upstream region is 200-500 mm/year, the annual ET in the artificial oasis in the midstream region is 500-700 mm/year, and the annual ET in the desolate desert areas in the downstream region is $< 150$ mm/year. The strong spatial variability of ET over the HRB implies the existence of control effects from topography, vegetation distribution and climate.

## 5. Discussion

### *5.1. Performance of DNN-MET*

By integrating satellite-derived ET products, ground-based measurements and ancillary surface properties, DNN-MET not only preserved the spatiotemporal information of the ET process but also yielded more accurate and robust ET estimations over the HRB (Fig. 5 and Fig. 6). Validation for 19 EC sites demonstrated that DNN-MET obtained higher KGE and $R^2$ and lower RMSE and Bias at all spatial scales compared with the individual ET products (Table S3 in supporting information), which indicates that our approach is able to provide a better representation of ET spatiotemporal patterns.

#### *5.1.1. Capacity for the DNN to simulate ET*

Fig. 11 demonstrates the relative differences in the mean annual (2008-2015) ET values between products and ground-based measurements, which is consistent with the spatial performance assessment in Fig. 6. The eight original products yielded better performances in the meadow and oasis areas in the upstream and midstream regions and generated higher relative differences in the arid desert regions. These results are consistent with those of Yao et al. (2019) and Hu and Jia

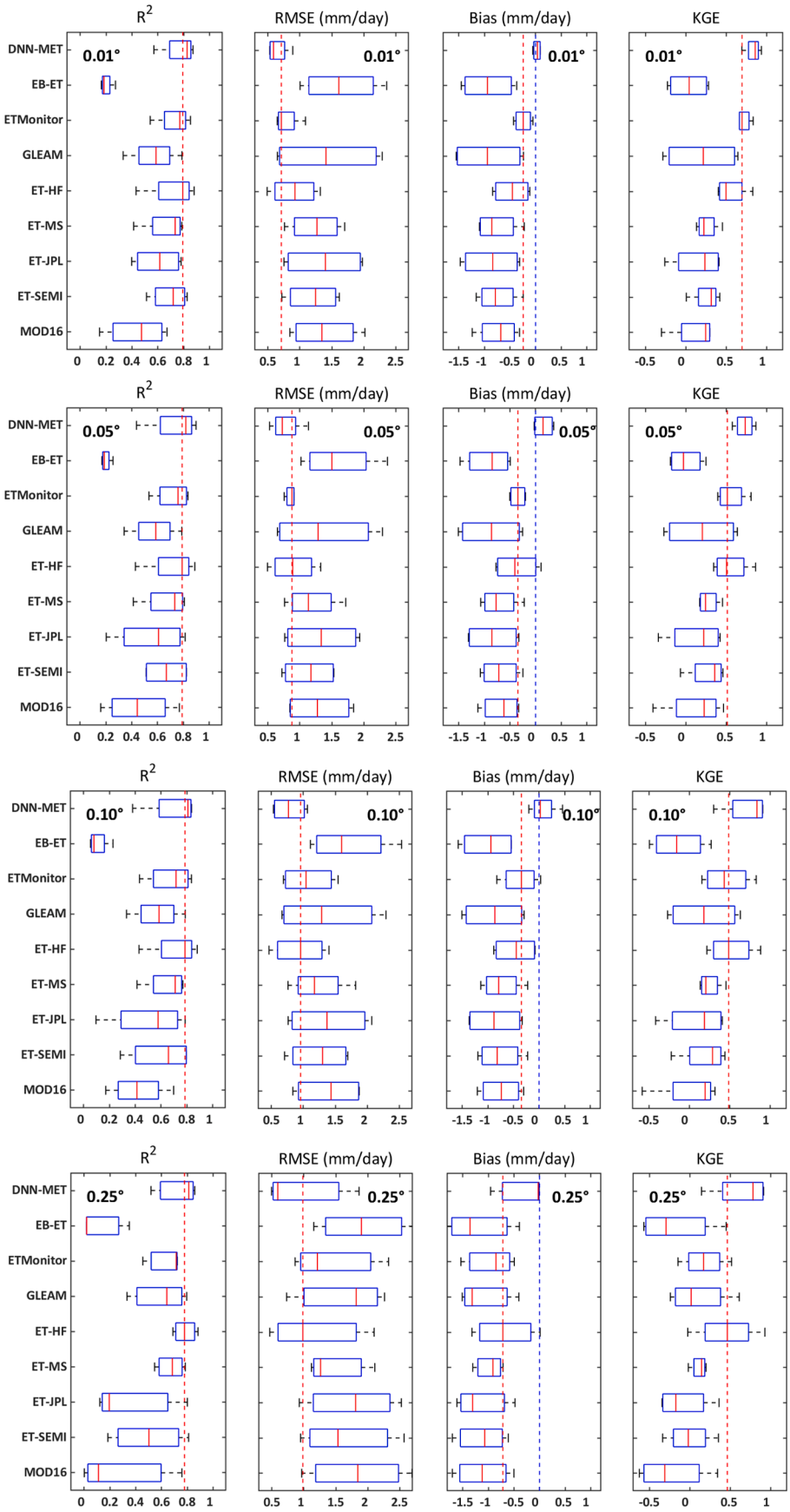


**Fig. 8.** Performance at 4 validation sites of the DNN-MET computed at four spatial resolutions: 0.01°, 0.05°, 0.10° and 0.25°. The red solid lines indicate the median values, the box edges indicate the first and third quartile values, and the end of the whiskers indicate values no greater than 1.5 times the inter-quartile value. The red dotted lines indicate the median values of the best-performing product used in the computation of DNN-MET. The blue dotted line indicates the optimal Bias value for all ET products.

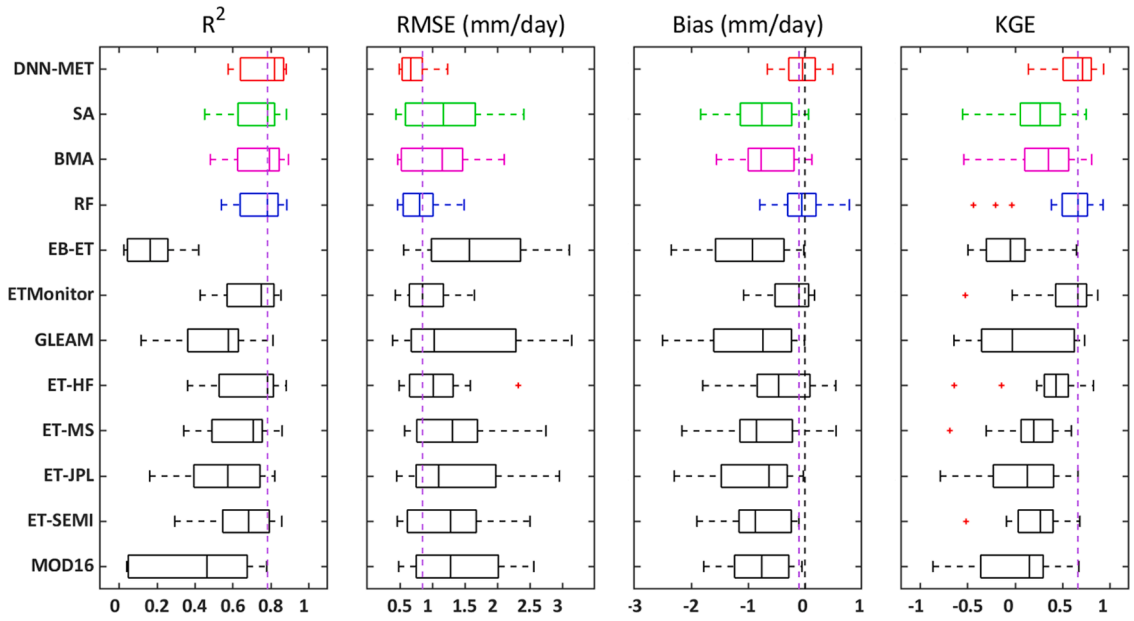


**Fig. 9.** Comparison of performances of DNN-MET and different merging methods (including Random Forest (RF), Bayesian model averaging (BMA) and a simple averaging (SA) method) using $R^2$, RMSE, Bias and KGE. The solid lines in the boxes indicate the median values, the box edges indicate the first and third quartile values, the end of the whiskers indicate values no greater than 1.5 times the interquartile value, and the most extreme data points beyond the whiskers indicate the extreme outliers. The purple dotted lines indicate the median values of the best-performing product used in the merging computation. The black dotted line indicates the optimal Bias value for all ET products.

(2015), which suggests that estimating ET in arid regions remains a challenge for existing ET products.

Substantial previous studies indicated that satellite-derived ET products may underestimate the terrestrial actual ET in the growing season, especially for irrigated artificial oases located in semiarid climate region (Kalma et al. 2008; Velpuri et al. 2013; Wang and Dickinson 2012). For MOD16, ET-SEMI, ET-JPL and ET-MS, the vapor pressure deficit (VPD), relative humidity deficit (RHD) and apparent thermal inertia (ATI) were used as indicators of water stress instead of soil moisture. Due to irrigation in the growing season and snowmelt in the spring, the soil water contents in the surface and root-zone layers are high (Liu et al. 2013), and the soil evaporation and vegetation transpiration are underestimated due to the overestimation of water stress parameterized by the VPD, RHD and ATI (Hu and Jia 2015). Conversely, ETMonitor and ET-HF used SM as soil moisture constraints, which performed better with lower mean annual relative differences (Fig. 11). The larger relative differences of GLEAM were mainly caused by uncertainties resulting from mismatched spatial scales between the pixel (0.25°) estimates and observations. These uncertainties are especially high in areas where the surface heterogeneity is notable (Kalma et al. 2008; Li et al. 2009a). EB-ET exhibited large relative differences partly due to the coarse spatial resolution (0.05°) and partly due to the influence of the LST used in the ET computation. ET computations using LST are restricted to clear sky conditions. In addition, the LST of mixed pixels cannot be directly used as the aerodynamic temperature, and this problem adversely affects ET computation, especially over heterogeneous and nonisothermal surfaces (Li et al. 2013b). These factors lead to EB-ET containing many missing values over the HRB throughout the whole year.

Although the eight global or quasi-global ET products showed substantial discrepancies when compared to ground-based measurements, they contained useful information about the spatiotemporal patterns of ET. Fig. 11 shows that DNN-MET had no significant deviations and yielded the closest ET estimates despite the challenges presented by the climate and topography variabilities. This may be attributed to the fact that DNN can learn new features by multiple layers of parameterized differentiable nonlinear modules (Zhou and Feng 2017) and perform the input-to-target transformation through these deep architectures (Saggi and Jain 2019). In particular, despite the substantial discrepancies between the original ET products, the DNN is able to collect positive information and enhance the overall accuracy to provide improved and robust spatiotemporal ET predictions.

### 5.1.2. *Comparison of different merging methods*

Fig. 9 shows that DNN-MET outperformed the other merging methods and had the highest median KGE value, 0.71, while the median KGE values for RF, BMA and SA were 0.66, 0.35 and 0.26, respectively. For the traditional SA and BMA methods, the merging performance highly depends on the accuracies of the original products (Duan and Phillips 2010). When the original estimates show substantial discrepancies, the merging accuracy will be limited by the worst-performing product (Yao et al. 2014). Machine learning methods such as RF can learn complex patterns to make predictions about unobserved data (Ball et al. 2017). Studies have demonstrated great success in terms of computation efficiency and accuracy (Breiman 2001; Geurts et al. 2006; Shang et al. 2020). However, without considering the interaction between ET process and surface properties, spatial prediction over heterogeneous underlying surfaces remains a great challenge for these nonspatial approaches (Hengl et al. 2018). DNN-MET presents a spatial prediction framework in which ancillary surface properties are used as explanatory variables, thus incorporating topography, landscape, climate and spatial autocorrelation into the prediction process. By integrating multiple products, ground-based measurements and ancillary information, DNN-MET can be used as a "knowledge engine'' for the spatial prediction of continuous ET and other climatological variables.

As Fig. S2 illustrates, when compared with a global 8-day temporal resolution ET product – GLASS ET, DNN-MET also yielded improved performance, especially at arid barren regions (downstream). At downstream, the median KGE for DNN-MET is 0.70, while the value for GLASS ET is only 0.09. This is mainly due to the combined influence from calibration sites, forcing data and merging methods. As a global ET product, GLASS ET uses 240 EC towers across the world on all continents provided by FLUXNET during merging procedure, which are not

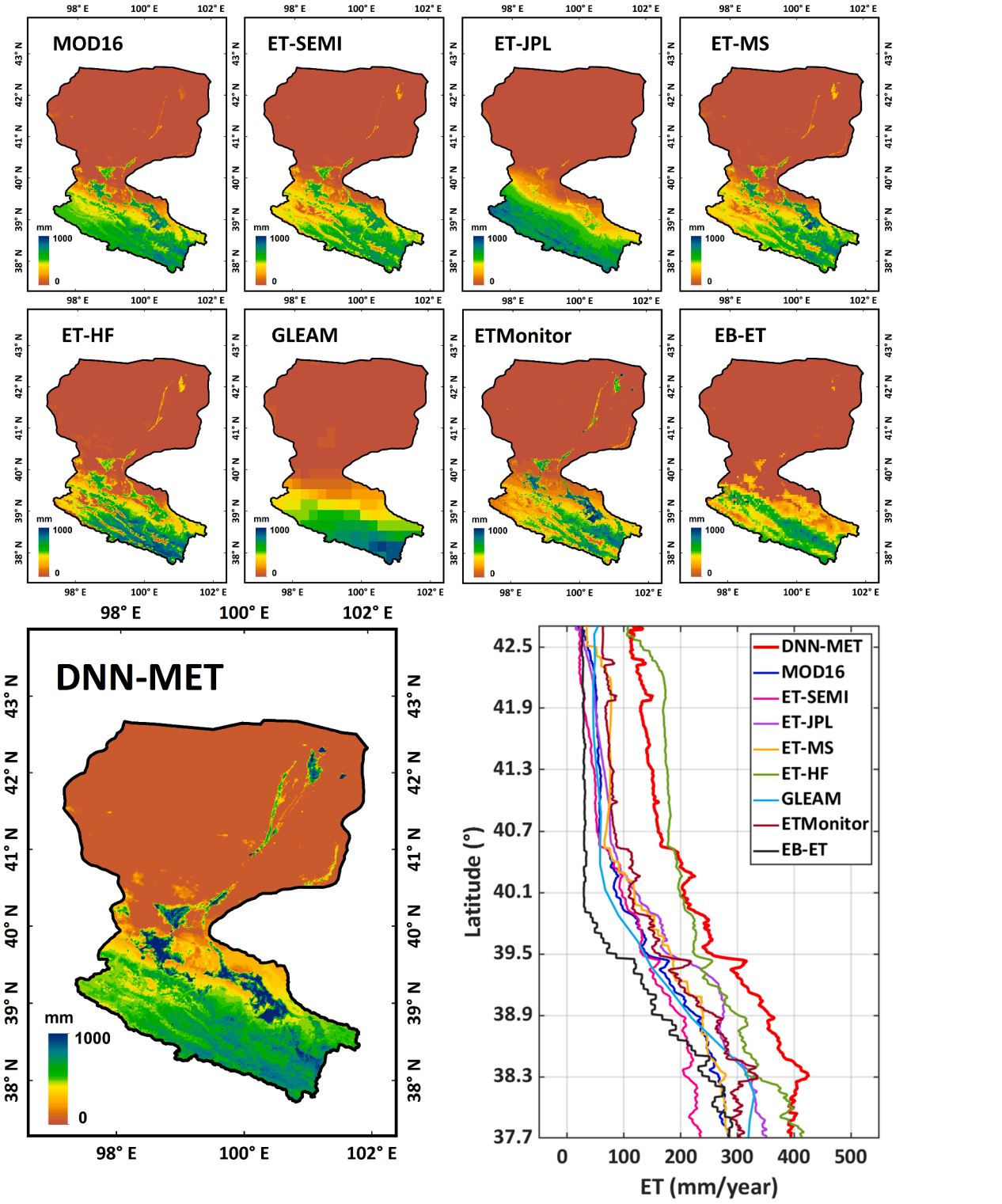


**Fig. 10.** Mapping of average annual DNN-MET in the period from 2012 to 2015 and ET variation with latitude over the Heihe River Basin.

included in the calibration and validation sites used in DNN-MET over HRB. And the meteorological forcing data of GLASS ET is derived from GMAO-MERRA with the resolution of 1/2° × 2/3°, which is much coarser than that of DNN-MET (0.1°). Additionally, as a weight dominated method, the BMA used in GLASS ET is not only limited by the worst-performing product, but also without considering the influence of surface properties, which makes it a lower accuracy ET product than DNN-MET.

*5.1.3. Generalization of merging and upscaling ET estimates*

Why did we implement the merging of multiple ET products instead of using the forcing data of these products as input variables directly to upscale ET from flux towers to the regional scale? To investigate the differences between merging ET and upscaling ET, we compared the generalization performances of these two approaches. Since it is difficult to comprehensively obtain the forcing data of ETMonitor, GLEAM and EB-ET, we chose MOD16, ET-SEMI, ET-JPL, ET-MS and ET-HF to

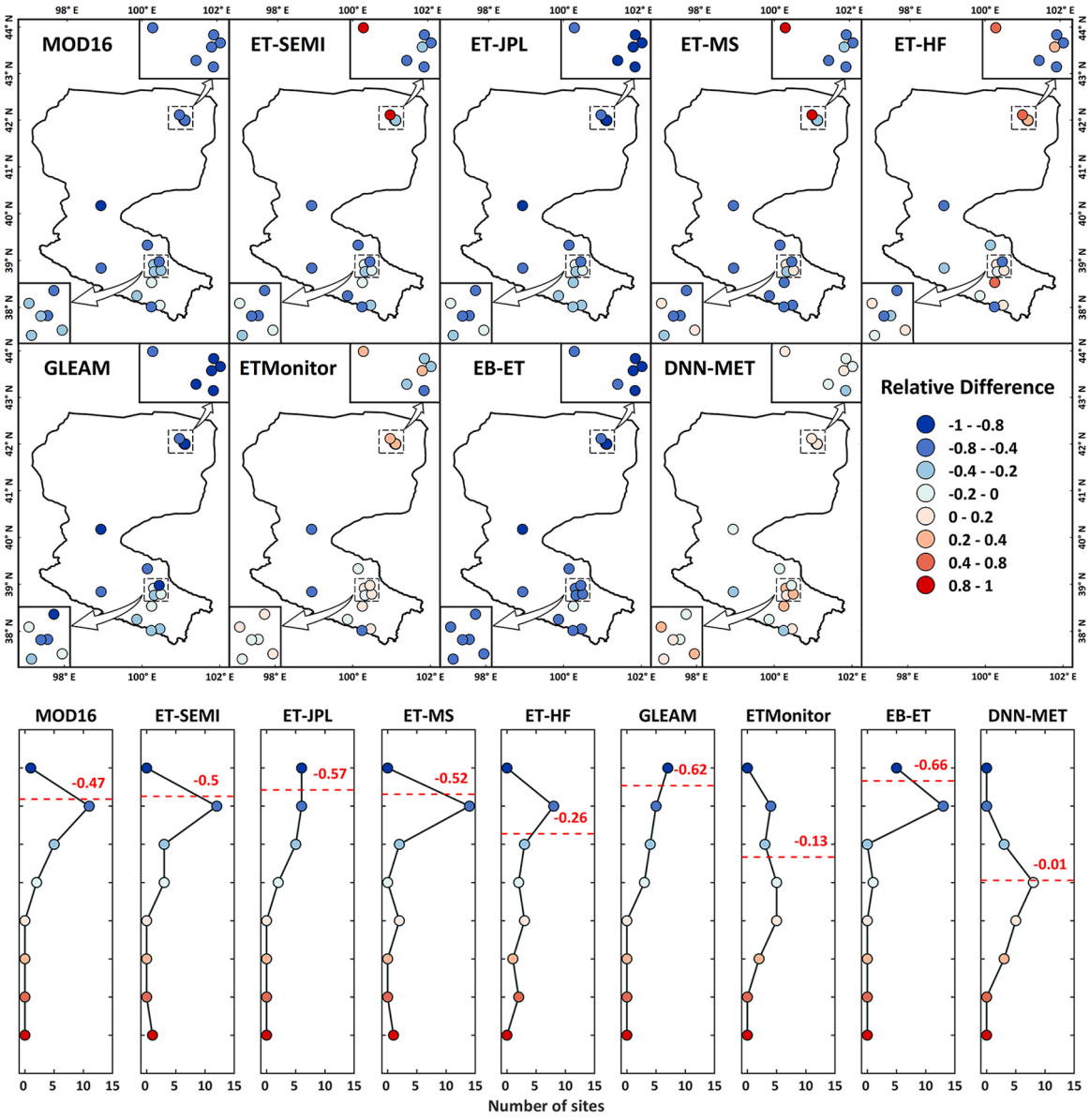


**Fig. 11.** Spatial distribution of the mean annual (2008-2015) relative differences at each EC site for all ET products. The points with negative values (blue points) represent underestimation, while the points with positive values (red points) represent overestimation. The dot plots in the bottom present the number of sites within each relative error range. The red dotted lines indicate the overall mean annual relative errors for each ET product at all EC sites.

conduct this investigation. We trained the DNN model using 80% of the EC sites (Table 2) to merge the five ET products. Then, with the same training samples, we used a DNN model to implement the upscaling of ET using all the forcing data of these five products. Table 3 presents a comparison of the performances of the estimates resulting from merging and upscaling ET at 4 validation sites. With the same satellite-derived data sources and ground observations, upscaling ET estimates showed comparable performance with merging ET estimates (the KGEs are 0.82 and 0.86 for upscaling and merging ET, respectively). The ET estimates showed highly consistent spatial characteristics over the whole basin, and the discrepancies between the two ET estimates were small in most portions of the HRB (Fig. 12 a).

However, these two approaches showed substantial discrepancies over barren regions where vegetation coverage was sparse. For instance,

**Table 3**
The comparison of evaluation parameters ($R^2$, RMSE (mm/day), Bias (mm/day) and KGE) between merging and upscaling ET estimates at 4 validation sites. All correlation coefficients are significant with 99% confidence.

| Validation Results | Hulugou | | Linze | | Huazhaizi | | Populus euphratica | | All | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Merging ET | Upscaling ET | Merging ET | Upscaling ET | Merging ET | Upscaling ET | Merging ET | Upscaling ET | Merging ET | Upscaling ET |
| **$R^2$** | 0.86 | 0.79 | 0.80 | 0.76 | 0.45 | 0.33 | 0.86 | 0.88 | 0.78 | 0.72 |
| **RMSE** | 0.50 | 0.61 | 0.99 | 1.02 | 0.78 | 0.91 | 0.56 | 0.53 | 0.69 | 0.76 |
| **Bias** | -0.01 | -0.12 | 0.31 | 0.20 | 0.21 | 0.28 | -0.10 | -0.08 | 0.08 | 0.05 |
| **KGE** | **0.85** | **0.82** | **0.80** | **0.80** | **0.59** | **0.47** | **0.89** | **0.85** | **0.86** | **0.82** |

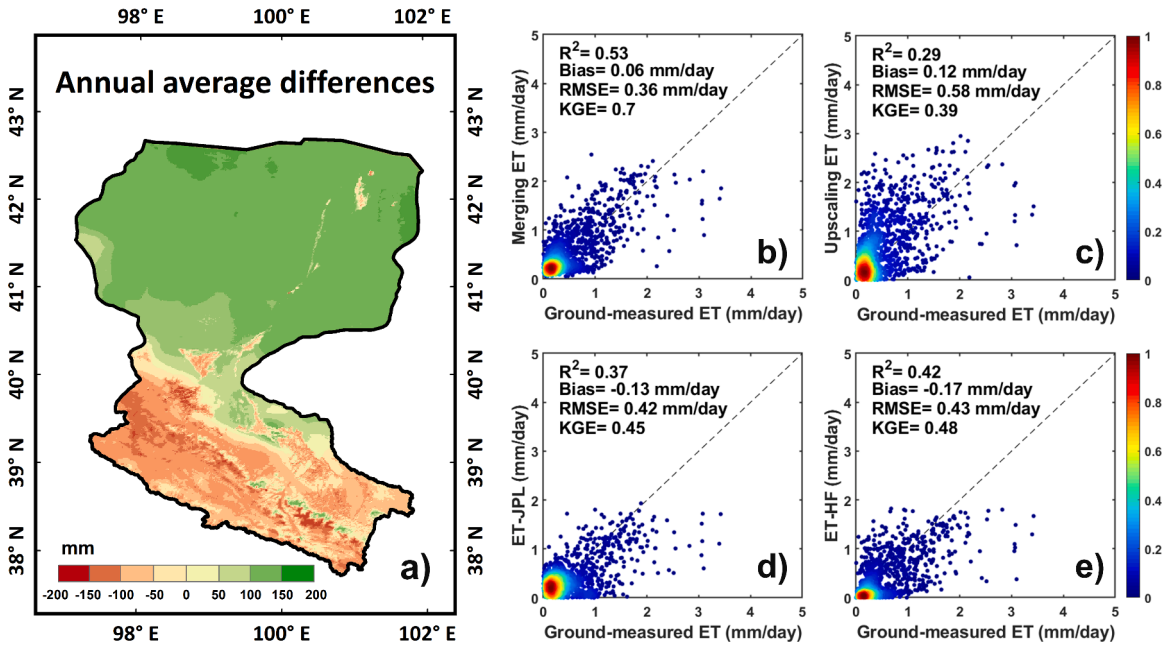


**Fig. 12.** a) Maps of annual average spatial differences between upscaling ET from flux towers to the regional scale and merging five ET products over the Heihe River Basin from 2012 to 2015. Scatterplots of daily ET observations and ET estimates in the extreme case of NDVI<0.15 are shown for b) merging ET, c) upscaling ET, d) ET-JPL and e) ET-HF.

Table 3 shows that the accuracy of the upscaling ET estimates at Huazhaizi (BAR) (which resulted in a KGE of 0.47) is significantly lower than the accuracy of the merging ET estimates (KGE of 0.59). For further exploration, we used ET estimates at 4 validation sites in the extreme case of NDVI<0.15 to assess the response of each method. A key result is that the merging of multiple ET products is systematically better than

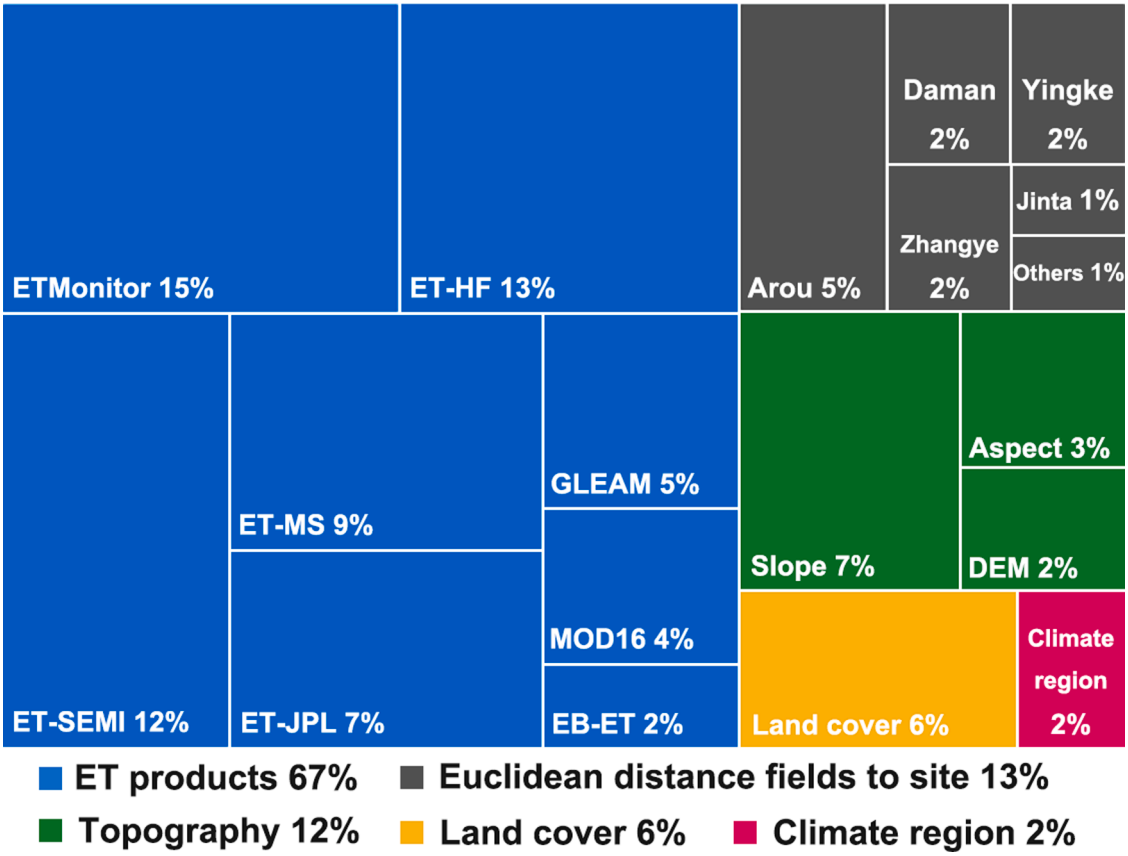


**Fig. 13.** The contributions of inputs to DNN-MET final output is dominated by the remotely sensed ET products, followed by Euclidean distance fields to eddy flux site and topography, with land cover and climate region contributing minor roles. The contributions are calculated based on SHAP values for the ET products and ancillary variables used in the merging procedure.

upscaling ET from the site scale to the regional scale for predicting extreme data sets (Fig. 12 b and c). For the extreme case of NDVI<0.15, merging ET performed much better than upscaling ET (merging ET resulted in a KGE of 0.70, compared to 0.39 for upscaling ET), highlighting its capacity to better predict the impact of the ecosystem on the ET process over barren regions. This may be attributable in part to the fact that ET products used in the merging procedure were constrained by process-based models; these physics-constrained estimates can better generalize under extreme circumstances (Zhao et al. 2019). For instance, the ET-JPL and ET-HF products used in the DNN-MET can obtain good performances when NDVI<0.15 without using calibration from ground measurements. These two products outperformed the nonphysics-constrained upscaling ET estimates with KGE values of 0.45 and 0.48 (Fig. 12 d and e). The upscaling methods are confined to the training samples, and the out-of-sample predictions may not be robust and may lead to large uncertainties (Chen et al. 2014). Therefore, our approach to merging ET products can combine deep learning model with physical knowledge to preserve the extrapolation capacity inherited from the ET algorithms and enhance the out-of-sample generalization performance for extreme cases.

### 5.2. Impact of ancillary variables, network density, and spatial resolution

To understand the impacts of ancillary variables on DNN-MET, we used SHAP (SHapley Additive exPlanations) (Lipovetsky and Conklin 2001) values to attribute prediction contribution for each input variable. SHAP values are based on the Shapley interaction index from game theory (Fujimoto et al. 2006), which explains model predictions as sums of actual contribution values of individual input features (Lundberg et al. 2018). The SHAP values for the surface properties account for 33% of all the input variables, which indicates the ancillary variables played an important role in the merging procedure (Fig. 13).

ET variations are affected by potential ET, radiation, temperature, wind speed, precipitation, soil moisture, vegetation distribution, etc. (Bertoldi et al. 2007; Jiménez et al. 2011; Kalma et al. 2008; Mueller et al. 2011; Wang and Dickinson 2012). All of these factors are associated with surface properties such as topography, land cover, and climate. For example, Yang et al. found that as one of the dominant factors affecting ET, the potential ET in the Qilian Mountains was not only sensitive to meteorological factors ($R_n$, RH, $T_a$, and WS) but was also sensitive to elevation gradients (Yang et al. 2019). An increase in elevation of 1500 meters will result in a 30 mm/month reduction in potential ET, which causes ET to decrease substantially as the elevation gradient increases (Yang et al. 2019). This is mainly due to the topography of mountainous regions directly affecting the distribution of solar radiation; all of the meteorological factors related to the ET process change markedly with elevation, slope and aspect. Chen et al. (2019b) also indicated that changes in land cover types lead to large ET variations over the HRB. Crop expansion from barren lands caused most of the increases in ET, while the conversion of alpine grass to desert grass caused most of the decreases in ET (Chen et al. 2019b).

EDFs to training sites, used as spatial autocorrelation indicators, incorporated geographical proximity effects into the merging process. Gridded layers of EDFs from each EC training site, they reflect the distance from each pixel to the EC sites directly. They account for spatial autocorrelations between observations in the geographical measures of proximity and connectivity. Fig. 13 shows that the SHAP values of the EDFs to all the sites account for 13% of the input variables, which indicates that the EDFs were attributed as most important prediction contributors among the ancillary information. Among the 19 EC sites, Arou had the highest SHAP value. This was mainly because Arou was the EC site with the longest observation duration, and it provided continuous and stable routine in situ measurements. So the location of Arou site was indicative and meaningful.

Fig. 7 illustrates that a high number of EC sites used in the training of DNN-MET leads to better performance of ET spatiotemporal representation. When we decreased the training samples to 20% (4 sites), DNN-MET still obtained improved performance, which demonstrates the effective spatial merging performance of our approach. Although a high density of ground observations is essential for ET spatial predictions, the incorporation of geographic proximity effects through EDFs can allow full use of the spatial information of sparse sites, which enhances the ET spatial prediction performance in data-scarce regions. The emphasis here is that the 20% training sites we choose here are the least site collocations to reflect the main plant function types and entire watershed variability. If we continue to reduce the number of training sites, the training samples with insufficient representativeness would lead to inferior model performance and large uncertainties.

DNN-MET also yielded improved performance based on coarse-resolution ET products (Fig. 8). This is mainly because the ancillary variables used in the merging process enhanced the spatial association between large-scale mixed pixels from satellite observations, which was ignored in traditional nonspatial ET merging methods. By blending ET products, ground observations and ancillary information, our approach made it possible to generate optimal ET spatial predictions at different spatial resolutions using sparse observation sites.

### 5.3. Uncertainty in the DNN-MET estimate

Although DNN-MET improved the representation of ET spatiotemporal patterns, there are still substantial uncertainties in the satellite-based ET estimates. These uncertainties mainly come from the data sources, DNN model structures and scaling effects (Baldocchi 2008; Mu et al. 2007; Polhamus et al. 2013).

The uncertainties from data sources comprise three parts: in situ measurements, satellite-based observations and meteorological reanalysis data. For EC in situ measurements, the energy imbalance always leads to inaccuracies in the interpretations (Mahrt 2010), and errors of approximately 5-20% are generally given (Foken 2008). Although we corrected energy imbalance using the method developed by Twine et al. (Twine et al. 2000), these corrections also cause certain errors (Finnigan et al. 2003; Twine et al. 2000). Previous studies revealed approximately 15-30% relative errors for vegetation parameters derived from satellite observations (Kalma et al. 2008; Wang and Dickinson 2012). Similarly, the classification accuracies of previous land-cover datasets are less than 75% (Hansen et al. 2000), and the agreement and consistency values are even lower in China (Liu et al. 2005). It was shown that a 3 K error in the LST product generated deviations of 75% in H from a two-source model (TSM), and the difference between LST and $T_a$ used in a SEBS model yielded deviations in H up to 45% (Kalma et al. 2008). Meteorological reanalysis data also contain a large Bias when compared to in situ measurements (Rienecker et al. 2011; Zhao et al. 2006), and no individual reanalysis dataset can provide the most accurate land surface energy budget estimation (Badgley et al. 2015; Shi and Liang 2014; Wang and Zeng 2012). Therefore, the uncertainties could be inherited through errors from these input data sources.

Although the DNN model is powerful, it has apparent deficiencies (Krizhevsky et al. 2012). It is widely recognized that DNN requires a great amount of training data, which limits its performances on small-scale data tasks (Hinton et al. 2012). The lack of a sufficient amount of labeled data would lead to the inferior performance of a DNN (Zhou and Feng 2017). Furthermore, DNN is a complicated model that always requires many hyperparameters, and the learning performance of DNN depends seriously on their careful tuning (Lecun et al. 1998; Simonyan and Zisserman 2014). The inclusion of too many interfering factors and almost infinite configurational combinations cause great uncertainty in the DNN model construction process (Zhou and Feng 2017).

The errors from scaling effects are also an important problem. The mismatched spatial scales between in situ measurements and satellite observations may cause large uncertainties (Fisher et al. 2020; Yao et al. 2017b). The limited spatial representation of EC measurements makes it

difficult to evaluate ET estimates at the satellite pixel scale, especially over heterogeneous underlying surfaces (Kalma et al. 2008). When using the EC measurements as the "true" values to evaluate and calibrate the merging models, the uncertainties from scaling effects vary from 5% to 25% (Li et al. 2009b).

## 6. Conclusions

We proposed a DNN based spatial integration framework by merging satellite-derived ET products, ground-based measurements and ancillary information to enhance ET estimates over heterogeneous surface. The key results of our approach, applied over the HRB, can be summarized as follows:

(a) The validation results showed that DNN-MET was superior to all other methods we evaluated.
(b) The eight individual ET products performed better in mountainous regions than in arid desert regions, while DNN-MET performed well for all terrain and climate conditions.
(c) The more EC sites used in training, the better DNN-MET performed. DNN-MET can also improve the ET spatiotemporal estimation in data-scarce regions.
(d) Despite different spatial resolutions between multiple ET products, DNN-MET demonstrated improved spatiotemporal ET predictions.
(e) DNN-MET outperformed other merging models, including a machine learning method (RF), a predictive probability density function (BMA), a traditional averaging method (SA), and a state-of-the-art global merging ET product (GLASS ET).
(f) Compared with the upscaling of ET from flux towers to the regional scale, our approach combined the deep learning method and ET process-based physical constraints, which enhanced the out-of-sample generalization performance for the extreme cases.
(g) Ancillary variables played an important role in the merging procedure.

## 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

Authors would like to thank Dr. Zhongli Zhu, Dr. Linna Chai and Dr. Tongren Xu from Faculty of Geographical Science, Beijing Normal University, China for providing the flux observation data. We gratefully acknowledge Kun Yang from Institute of Tibetan Plateau Research, Chinese Academy of Sciences for providing gridded China Meteorological Forcing Dataset (CMFD), and thank Li Jia from Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences for providing ETMonitor dataset. Ground-observed flux observation data were downloaded from WATER and HiWATER experiments (http://westdc.westgis.ac.cn/) under the fair-use policy. This work was partially supported by the National Key Research and Development Program of China (No. 2016YFA0600103), the Strategic Priority Research Program of the Chinese Academy of Sciences (No. XDA20100101) and the Natural Science Fund of China (No. 41671331).

## Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.agrformet.2021.108582.

## References

Aires, F., 2014. Combining Datasets of Satellite-Retrieved Products. Part I: Methodology and Water Budget Closure. Journal of Hydrometeorology 15, 1677–1691.
Allen, R.G., Pereira, L.S., Howell, T.A., Jensen, M.E., 2011. Evapotranspiration information reporting: I. Factors governing measurement accuracy. Agric. Water Manage. 98, 899–920.
Badgley, G., Fisher, J.B., Jiménez, C., Tu, K.P., Vinukollu, R., 2015. On Uncertainty in Global Terrestrial Evapotranspiration Estimates from Choice of Input Forcing Datasets*. Journal of Hydrometeorology 16, 1449–1455.
Baez-Villanueva, O.M., Zambrano-Bigiarini, M., Beck, H.E., McNamara, I., Ribbe, L., Nauditt, A., Birkel, C., Verbist, K., Giraldo-Osorio, J.D., Xuan Thinh, N., 2020. RF-MEP: A novel Random Forest method for merging gridded precipitation products and ground-based measurements. Remote Sens. Environ. 239, 111606.
Baldocchi, D., 2008. Breathing of the terrestrial biosphere: lessons learned from a global network of carbon dioxide flux measurement systems. Aust. J. Bot. 56, 1–26.
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 U, 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. Meteorol. Soc. 82, 2415–2434.
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, 042609.
Benvenuto, N., Piazza, F., 1992. On the complex backpropagation algorithm. Trans. Sig. Proc. 40, 967–969.
Bertoldi, G., Albertson, J.D., Kustas, W.P., Li, F., Anderson, M.C, 2007. On the opposing roles of air temperature and wind speed variability in flux estimation from remotely sensed land surface states. Water Resour. Res. 43 p.W10433.10431-W10433.10413.
Bonan, G., 2015. Ecological Climatology: Concepts and Applications. Cambridge University Press, Cambridge, 3 ed.
Breiman, L., 2001. Random Forests. Mach. Learn. 45, 5–32.
Chen, J., Saunders, S.C., Crow, T.R., Naiman, R.J., Brosofske, K.D., Mroz, G.D., Brookshire, B.L., Franklin, J.F., 1999. Microclimate in Forest Ecosystem and Landscape Ecology: Variations in local climate can be used to monitor and compare the effects of different management regimes. Bioscience 49, 288–297.
Chen, X., Massman, W.J., Su, Z., 2019a. A Column Canopy-Air Turbulent Diffusion Method for Different Canopy Structures. Journal of Geophysical Research: Atmospheres 124, 488–506.
Chen, X., Su, Z., Ma, Y., Yang, K., Wen, J., Zhang, Y., 2013. An Improvement of Roughness Height Parameterization of the Surface Energy Balance System (SEBS) over the Tibetan Plateau. Journal of Applied Meteorology & Climatology 52, 607–622.
Chen, Y., Wang, S., Ren, Z., Huang, J., Wang, X., Liu, S., Deng, H., Lin, W., 2019b. Increased evapotranspiration from land cover changes intensified water crisis in an arid river basin in northwest China. J. Hydrol. 574, 383–397.
Chen, Y., Xia, J., Liang, S., Feng, J., Fisher, J.B., 2014. Comparison of satellite-based evapotranspiration models over terrestrial ecosystems in China. Remote Sens. Environ. 140, 279–293.
Chen, Y., Yuan, W., Xia, J., Fisher, J.B., Dong, W., Zhang, X., Liang, S., Ye, A., Cai, W., Feng, J., 2015. Using Bayesian model averaging to estimate terrestrial evapotranspiration in China. J. Hydrol. 528, 537–549.
Cheng, G., Li, X., Zhao, W., Xu, Z., Feng, Q., Xiao, S., Xiao, H., 2014. Integrated study of the water–ecosystem–economy in the Heihe River Basin. Natl. Sci. Rev. 1, 413–428.
Chollet, F.o., 2017. Deep Learning with Python. Manning Publications.
Duan, Q., Phillips, T.J., 2010. Bayesian estimation of local signal and noise in multimodel simulations of climate change. J. Geophys. Res. 115.
Ershadi, A., McCabe, M.F., Evans, J.P., Chaney, N.W., Wood, E.F., 2014. Multi-site evaluation of terrestrial evaporation models using FLUXNET data. Agricultural & Forest Meteorology 187, 46–61.
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.
Finnigan, J.J., Clement, R., Malhi, Y., Leuningh, R., Cleugh, H.A, 2003. A Re-Evaluation of Long-Term Flux Measurement Techniques Part I: Averaging and Coordinate Rotation. Boundary Layer Meteorology 1–48.
Fisher, J.B., Lee, B., Purdy, A.J., Halverson, G.H., Dohlen, M.B., Cawse-Nicholson, K., Wang, A., Anderson, R.G., Aragon, B., Arain, M.A., Baldocchi, D.D., Baker, J.M., Barral, H., Bernacchi, C.J., Bernhofer, C., Biraud, S.C., Bohrer, G., Brunsell, N., Cappelaere, B., Castro-Contreras, S., Chun, J., Conrad, B.J., Cremonese, E., Demarty, J., Desai, A.R., De Ligne, A., Foltýnová, L., Goulden, M.L., Griffis, T.J., Grünwald, T., Johnson, M.S., Kang, M., Kelbe, D., Kowalska, N., Lim, J.-H., Maïnassara, I., McCabe, M.F., Missik, J.E.C., Mohanty, B.P., Moore, C.E., Morillas, L., Morrison, R., Munger, J.W., Posse, G., Richardson, A.D., Russell, E.S., Ryu, Y., Sanchez-Azofeifa, A., Schmidt, M., Schwartz, E., Sharp, I., Šigut, L., Tang, Y., Hulley, G., Anderson, M., Hain, C., French, A., Wood, E., Hook, S., 2020. ECOSTRESS: NASA's Next Generation Mission to Measure Evapotranspiration From the International Space Station. Water Resour. Res. 56, e2019WR026058.
Fisher, J.B., Melton, F., Middleton, E., Hain, C., Anderson, M., Allen, R., McCabe, M.F., Hook, S., Baldocchi, D., Townsend, P.A., Kilic, A., Tu, K., Miralles, D.D., Perret, J., Lagouarde, J.-P., Waliser, D., Purdy, A.J., French, A., Schimel, D., Famiglietti, J.S., Stephens, G., Wood, E.F., 2017. The future of evapotranspiration: Global

requirements for ecosystem functioning, carbon and climate feedbacks, agricultural management, and water resources. Water Resour. Res. 53, 2618–2626.
Fisher, J.B., Tu, K.P., Baldocchi, D.D., 2008. Global estimates of the land–atmosphere water flux based on monthly AVHRR and ISLSCP-II data, validated at 16 FLUXNET sites. Remote Sens. Environ. 112, 901–919.
Foken, T., 2008. THE ENERGY BALANCE CLOSURE PROBLEM: AN OVERVIEW. Ecol. Appl. 18, 1351–1367.
Fujimoto, K., Kojadinovic, I., Marichal, J., 2006. Axiomatic characterizations of probabilistic and cardinal-probabilistic interaction indices. Games and Economic Behavior 55, 72–99.
Geurts, P., Ernst, D., Wehenkel, L., 2006. Extremely randomized trees. Machine Learning 63, 3–42.
Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. J. Hydrol. 377, 80–91.
Hansen, M.C., Sohlberg, R., Defries, R.S., Townshend, J.R.G., 2000. Global land cover classification at 1 km spatial resolution using a classification tree approach. Int. J. Remote Sens. 21, 1331–1364.
He, X., Xu, T., Xia, Y., Bateni, S.M., Zhao, J., 2020. A Bayesian Three-Cornered Hat (BTCH) Method: Improving the Terrestrial Evapotranspiration Estimation. Remote Sensing 12, 878.
Hengl, T., Nussbaum, M., Wright, M.N., Heuvelink, G.B.M., Gräler, B., 2018. Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. PeerJ 6, e5518.
Hinton, G., Deng, L., Yu, D., Dahl, G.E., Mohamed, A., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T.N., Kingsbury, B., 2012. Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups. IEEE Signal Process. Mag. 29, 82–97.
Hu, G., Jia, L., 2015. Monitoring of Evapotranspiration in a Semi-Arid Inland River Basin by Combining Microwave and Optical Remote Sensing Observations. Remote Sensing 7, 3056–3087.
Ienco, D., Interdonato, R., Gaetano, R., Ho Tong Minh, D., 2019. Combining Sentinel-1 and Sentinel-2 Satellite Image Time Series for land cover mapping via a multi-source deep learning architecture. ISPRS J. Photogramm. Remote Sens. 158, 11–22.
Ivakhnenko, A.G., Lapa, V.G., 1966. CYBERNETIC PREDICTING DEVICES.
Jiménez, C., Martens, B., Miralles, D.M., Fisher, J.B., Beck, H.E., Fernández-Prieto, D., 2018. Exploring the merging of the global land evaporation WACMOS-ET products based on local tower measurements. Hydrol. Earth Syst. Sci. 22, 4513–4533.
Jiménez, C., Prigent, C., Mueller, B., Seneviratne, S.I., McCabe, M.F., Wood, E.F., Rossow, W.B., Balsamo, G., Betts, A.K., Dirmeyer, P.A., Fisher, J.B., Jung, M., Kanamitsu, M., Reichle, R.H., Reichstein, M., Rodell, M., Sheffield, J., Tu, K., Wang, K., 2011. Global intercomparison of 12 land surface heat flux estimates. Journal of Geophysical Research: Atmospheres 116.
Jung, M., Reichstein, M., Ciais, P., Seneviratne, S.I., Sheffield, J., Goulden, M.L., Bonan, G., Cescatti, A., Chen, J., de Jeu, R., Dolman, A.J., Eugster, W., Gerten, D., Gianelle, D., Gobron, N., Heinke, J., Kimball, J., Law, B.E., Montagnani, L., Mu, Q., Mueller, B., Oleson, K., Papale, D., Richardson, A.D., Roupsard, O., Running, S., Tomelleri, E., Viovy, N., Weber, U., Williams, C., Wood, E., Zaehle, S., Zhang, K., 2010. Recent decline in the global land evapotranspiration trend due to limited moisture supply. Nature 467, 951–954.
Kalma, J.D., Mcvicar, T.R., Mccabe, M.F., 2008. Estimating Land Surface Evaporation: A Review of Methods Using Remotely Sensed Surface Temperature Data. Surv. Geophys. 29, 421–469.
Kool, D., Agam, N., Lazarovitch, N., Heitman, J.L., Sauer, T.J., Bengal, A., 2014. A review of approaches for evapotranspiration partitioning. Agric. For. Meteorol. 184, 56–70.
Kottek, Markus, Grieser, Jürgen, Beck, Christoph, Rudolf, Bruno, Rubel, & Franz, 2006. World Map of the Köppen-Geiger climate classification updated. Meteorologische Zeitschrift.
Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet classification with deep convolutional neural networks. In: Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1. Curran Associates Inc, Lake Tahoe, Nevada, pp. 1097–1105.
Lecun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-based learning applied to document recognition. Proc. IEEE 86, 2278–2324.
Li, X., Cheng, G., Liu, S., Xiao, Q., Ma, M., Jin, R., Che, T., Liu, Q., Wang, W., Qi, Y., 2013a. Heihe Watershed Allied Telemetry Experimental Research (HiWATER): Scientific Objectives and Experimental Design. Bull. Am. Meteorol. Soc. 94, 1145–1160.
Li, X., Li, X., Li, Z., Ma, M., Wang, J., Xiao, Q., Liu, Q., Che, T., Chen, E., Yan, G., Hu, Z., Zhang, L., Chu, R., Su, P., Liu, Q., Liu, S., Wang, J., Niu, Z., Chen, Y., Jin, R., Wang, W., Ran, Y., Xin, X., Ren, H., 2009a. Watershed Allied Telemetry Experimental Research. Journal of Geophysical Research: Atmospheres 114.
Li, Z., Tang, B., Wu, H., Ren, H., Yan, G., Wan, Z., Trigo, I.F., Sobrino, J.A., 2013b. Satellite-derived land surface temperature: Current status and perspectives. Remote Sens. Environ. 131, 14–37.
Li, Z.L., Tang, R., Wan, Z., Bi, Y., Zhou, C., Tang, Bohui, Yan, G., Zhang, X, 2009b. A Review of Current Methodologies for Regional Evapotranspiration Estimation from Remotely Sensed Data. Sensors 9, 3801–3853.
Liang, S., Cheng, J., Jia, K., Jiang, B., Liu, Q., Xiao, Z., Yao, Y., Yuan, W., Zhang, X., Zhao, X., Zhou, J., 2020. The Global LAnd Surface Satellite (GLASS) product suite. Bull. Am. Meteorol. Soc. 1–37.
Liang, S., Kustas, W., Schaepman-Strub, G., Li, X., 2010a. Impacts of Climate Change and Land Use Changes on Land Surface Radiation and Energy Budgets. *Selected Topics in Applied Earth Observations and Remote Sensing*. IEEE Journal of, 3 219–224.
Liang, S., Wang, D., He, T., Yu, Y., 2019. Remote sensing of earth's energy budget: synthesis and review. Int. J. Digital Earth 12, 737–780.
Liang, S., Wang, D., Tao, X., Cheng, J., Yao, Y., Zhang, X., He, T, 2018. Methodologies for Integrating Multiple High-Level Remotely Sensed Land Products. In: Liang, S. (Ed.), Comprehensive Remote Sensing. Elsevier, Oxford, pp. 278–317.
Liang, S., Wang, J., 2019. Advanced remote sensing: Terrestrial information extraction and applications. Academic Press/Elsevier, Cambridge, UK, pp. 1–986.
Liang, S., Wang, K., Zhang, X., Wild, M., 2010b. Review on Estimation of Land Surface Radiation and Energy Budgets From Ground Measurement, Remote Sensing and Model Simulations. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 3, 225–240.
Lipovetsky, S., Conklin, M., 2001. Analysis of regression in game theory approach. Applied Stochastic Models in Business and Industry 17, 319–330.
Liu, J., Liu, M., Tian, H., Zhuang, D., Zhang, Z., Zhang, W., Tang, X., Deng, X., 2005. Spatial and temporal patterns of China's cropland during 1990–2000: An analysis based on Landsat TM data. Remote Sens. Environ. 98, 442–456.
Liu, S., Li, X., Xu, Z., Che, T., Xiao, Q., Ma, M., Liu, Q., Jin, R., Guo, J., Wang, L., 2018. The Heihe Integrated Observatory Network: A Basin-Scale Land Surface Processes Observatory in China. Vadose Zone J. 17.
Liu, S., Xu, Z., Zhu, Z., Jia, Z., Zhu, M., 2013. Measurements of evapotranspiration from eddy-covariance systems and large aperture scintillometers in the Hai River Basin. China. *Journal of Hydrology* 487, 24–38.
Lundberg, S., Erion, G.G., Lee, S., 2018. Consistent Individualized Feature Attribution for Tree Ensembles. arXiv: Learning.
Maes, W.H., Pagán, B.R., Martens, B., Gentine, P., Guanter, L., Steppe, K., Verhoest, N.E. C., Dorigo, W., Li, X., Xiao, J., Miralles, D.G., 2020. Sun-induced fluorescence closely linked to ecosystem transpiration as evidenced by satellite data and radiative transfer models. Remote Sens. Environ. 249, 112030.
Mahrt, L., 2010. Computing turbulent fluxes near the surface: Needed improvements. Agric. For. Meteorol. 150, 501–509.
Mao, J., Fu, W., Shi, X., Ricciuto, D.M., Fisher, J.B., Dickinson, R.E., Wei, Y., Shem, W., Piao, S., Wang, K., Schwalm, C.R., Tian, H., Mu, M., Arain, A., Ciais, P., Cook, R., Dai, Y., Hayes, D., Hoffman, F.M., Huang, M., Huang, S., Huntzinger, D.N., Ito, A., Jain, A., King, A.W., Lei, H., Lu, C., Michalak, A.M., Parazoo, N., Peng, C., Peng, S., Poulter, B., Schaefer, K., Jafarov, E., Thornton, P.E., Wang, W., Zeng, N., Zeng, Z., Zhao, F., Zhu, Q., Zhu, Z., 2015. Disentangling climatic and anthropogenic controls on global terrestrial evapotranspiration trends. Environ. Res. Lett. 10, 094008.
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.
Miralles, D.G., Holmes, T.R.H., De Jeu, R.A.M., Gash, J.H., Meesters, A.G.C.A., Dolman, A.J., 2011. Global land-surface evaporation estimated from satellite-based observations. Hydrol. Earth Syst. Sci. 15, 453–469.
Monteith, J.I.L., 1965. Evaporation and environment. Symp Soc Exp Biol 19, 205–234.
Mu, Q., Heinsch, F.A., Zhao, M., Running, S.W., 2007. Development of a global evapotranspiration algorithm based on MODIS and global meteorology data. Remote Sens. Environ. 111, 519–536.
Mu, Q., Zhao, M., Running, S.W., 2011. Improvements to a MODIS global terrestrial evapotranspiration algorithm. Remote Sens. Environ. 115, 1781–1800.
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.
Mueller, B., Seneviratne, S.I., Jimenez, C., Corti, T., Hirschi, M., Balsamo, G., Ciais, P., Dirmeyer, P., Fisher, J.B., Guo, Z., Jung, M., Maignan, F., McCabe, M.F., Reichle, R., 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.
Pan, M., Wood, E.F., 2006. Data Assimilation for Estimating the Terrestrial Water Budget Using a Constrained Ensemble Kalman Filter. Journal of Hydrometeorology 7, 534–547.
Polhamus, A., Fisher, J.B., Tu, K.P., 2013. What controls the error structure in evapotranspiration models? Agric. For. Meteorol. 169, 12–24.
Priestley, C.H.B., Taylor, R.J., 1972. On the assessment of surface heat flux and evaporation using large-scale parameters. Weather Review 100, 81–92.
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.
Rienecker, M.M., Suarez, M.J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M.G., Schubert, S.D., Takacs, L., Kim, G.-K., Bloom, S., Chen, J., Collins, D., Conaty, A., da Silva, A., Gu, W., Joiner, J., Koster, R.D., Lucchesi, R., Molod, A., Owens, T., Pawson, S., Pegion, P., Redder, C.R., Reichle, R., Robertson, F. R., Ruddick, A.G., Sienkiewicz, M., Woollen, J., 2011. MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications. Journal of Climate 24, 3624–3648.
Rosenfeld, A., Pfaltz, J.L., 1968. Distance functions on digital pictures. Pattern Recognit. 1, 33–61.
Saggi, M.K., Jain, S., 2019. Reference evapotranspiration estimation and modeling of the Punjab Northern India using deep learning. Comput. Electron. Agric. 156, 387–398.
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 Sensing 12, 687.
Shi, Q., Liang, S., 2014. Surface-sensible and latent heat fluxes over the Tibetan Plateau from ground measurements, reanalysis, and satellite data. Atmos. Chem. Phys. 14, 5659–5677.

Shuttleworth, W.J., & Wallace, J.S. (1985). Evaporation from sparse crops-an energy combination theory, 111, 839-855.
Simonyan, K., Zisserman, A., 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition. Computer ence.
Su, Z., 2002. The Surface Energy Balance System (SEBS) for estimation of turbulent heat fluxes. Hydrol. Earth Syst. Sci. 6, 85–100.
Talsma, C.J., Good, S.P., Carlos, J., Brecht, M., Fisher, J.B., Miralles, D.G., Mccabe, M.F., Purdy, A.J., 2018. Partitioning of evapotranspiration in remote sensing-based models. Agricultural & Forest Meteorology 131–143. *s 260–261*.
Taylor, K.E., 2001. Summarizing multiple aspects of model performance in a single diagram. Journal of Geophysical Research: Atmospheres 106, 7183–7192.
Trenberth, K.E., Fasullo, J.T., Kiehl, J.T., 2009. Earth's Global Energy Budget, 90. Bulletin of the American Meteorological Society, pp. 311–323.
Twine, T.E., Kustas, W.P., Norman, J.M., Cook, D.R., Houser, P.R., Meyers, T.P., Prueger, J.H., Starks, P.J., Wesely, M.L., 2000. Correcting eddy-covariance flux underestimates over a grassland. Agric. For. Meteorol. 103, 279–300.
Velpuri, N.M., Senay, G.B., Singh, R.K., Bohms, S., Verdin, J.P., 2013. A comprehensive evaluation of two MODIS evapotranspiration products over the conterminous United States: Using point and gridded FLUXNET and water balance ET. Remote Sens. Environ. 139, 35–49.
Vinukollu, R.K., Meynadier, R., Sheffield, J., Wood, E.F., 2011. Multi-model, multi-sensor estimates of global evapotranspiration: climatology, uncertainties and trends. Hydrol. Processes 25, 3993–4010.
Wang, A., Zeng, X., 2012. Evaluation of multireanalysis products with in situ observations over the Tibetan Plateau. Journal of Geophysical Research: Atmospheres 117.
Wang, K., Dickinson, R.E., 2012. A review of global terrestrial evapotranspiration: Observation, modeling, climatology, and climatic variability. Rev. Geophys. 50.
Wang, K., Dickinson, R.E., Wild, M., Liang, S., 2010. Evidence for decadal variation in global terrestrial evapotranspiration between 1982 and 2002: 1. Model development. Journal of Geophysical Research: Atmospheres 115.
Xiong, Y.J., Zhao, S.H., Tian, F., Qiu, G.Y., 2015. An evapotranspiration product for arid regions based on the three-temperature model and thermal remote sensing. J. Hydrol. 530, 392–404.
Xu, Z., Liu, S., Zhu, Z., Zhou, J., Shi, W., Xu, T., Yang, X., Zhang, Y., He, X., 2020. Exploring evapotranspiration changes in a typical endorheic basin through the integrated observatory network. Agric. For. Meteorol. 290, 1–14.
Yang, 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.
Yao, Y., Liang, S., Cheng, J., Liu, S., Fisher, J.B., 2013. MODIS-driven estimation of terrestrial latent heat flux in China based on a modified Priestley-Taylor algorithm. Agricultural & Forest Meteorology 187–202. *171-172*.
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.
Yao, Y., Liang, S., Li, X., Chen, J., Wang, K., Jia, K., Cheng, J., Jiang, B., Fisher, J.B., Mu, Q., Grünwald, T., Bernhofer, C., Roupsard, O., 2015. A satellite-based hybrid algorithm to determine the Priestley–Taylor parameter for global terrestrial latent heat flux estimation across multiple biomes. Remote Sens. Environ. 165, 216–233.
Yao, Y., Liang, S., Li, X., Hong, Y., Fisher, J.B., Zhang, N., Chen, J., Cheng, J., Zhao, S., Zhang, X., 2014. Bayesian multimodel estimation of global terrestrial latent heat flux from eddy covariance, meteorological, and satellite observations. J. Geophys. Res. 119, 4521–4545.
Yao, Y., 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.
Yao, Y., Liang, S., Yu, J., Chen, J., Liu, S., Lin, Y., Fisher, J.B., McVicar, T.R., Cheng, J., Jia, K., Zhang, X., Xie, X., Jiang, B., Sun, L., 2017c. A simple temperature domain two-source model for estimating agricultural field surface energy fluxes from Landsat images. Journal of Geophysical Research: Atmospheres 122, 5211–5236.
Yao, Y., Zhang, Y., Liu, Q., Liu, S., Jia, K., Zhang, X., Xu, Z., Xu, T., Chen, J., Fisher, J.B., 2019. Evaluation of a satellite-derived model parameterized by three soil moisture constraints to estimate terrestrial latent heat flux in the Heihe River basin of Northwest China. Sci. Total Environ. 695, 133787.
Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., Xu, H., Tan, W., Yang, Q., Wang, J., Gao, J., Zhang, L., 2020. Deep learning in environmental remote sensing: Achievements and challenges. Remote Sens. Environ. 241, 111716.
Zhao, M., Running, S.W., Nemani, R.R., 2006. Sensitivity of Moderate Resolution Imaging Spectroradiometer (MODIS) terrestrial primary production to the accuracy of meteorological reanalyses. Journal of Geophysical Research: Biogeosciences 111.
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.
Zhong, B., Yang, A., Nie, A., Yao, Y., Zhang, H., Wu, S., Liu, Q., 2015. Finer Resolution Land-Cover Mapping Using Multiple Classifiers and Multisource Remotely Sensed Data in the Heihe River Basin. IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing 8, 4973–4992.
Zhou, Z., Feng, J., 2017. Deep forest: towards an alternative to deep neural networks. In: international joint conference on artificial intelligence, pp. 3553–3559.

Supporting Online Material for

# DNN-MET: A Deep Neural Networks method to integrate satellite-derived evapotranspiration products, eddy covariance observations and ancillary information

Ke Shang, Yunjun Yao *, Shunlin Liang, Yuhu Zhang, Joshua B. Fisher, Jiquan Chen, Shaomin Liu, Ziwei Xu, Yuan Zhang, Kun Jia, Xiaotong Zhang, Junming Yang, Xiangyi Bei, Xiaozheng Guo, Ruiyang Yu, Zijing Xie, Lilin Zhang

*To whom correspondence should be addressed. E-mail: boyyunjun@163.com

**This PDF file includes:**

# Supporting Online Material

**Figure S1**

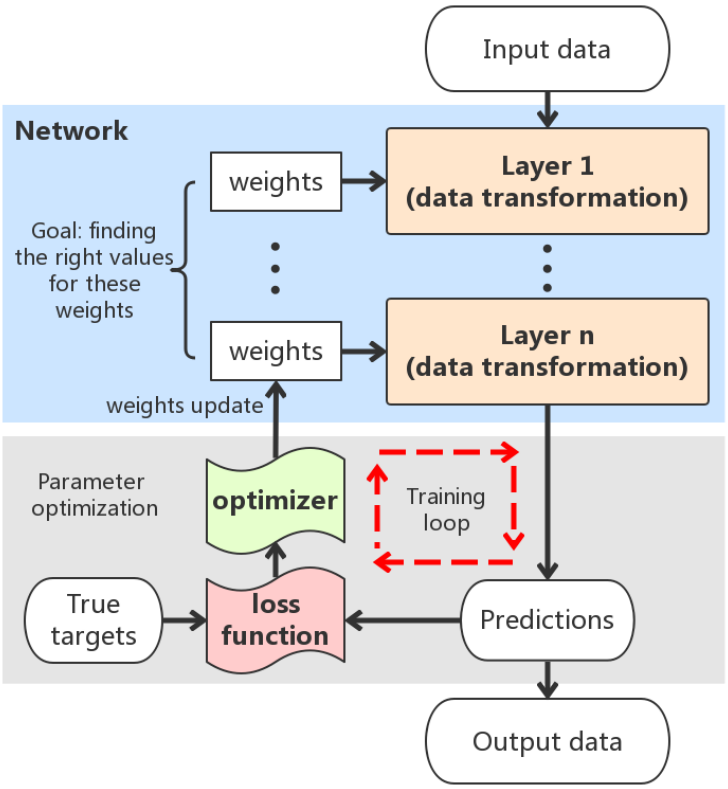


Fig. S1. Relationships between the layers, network, loss function and optimizer in a DNN model

**Figure S2**

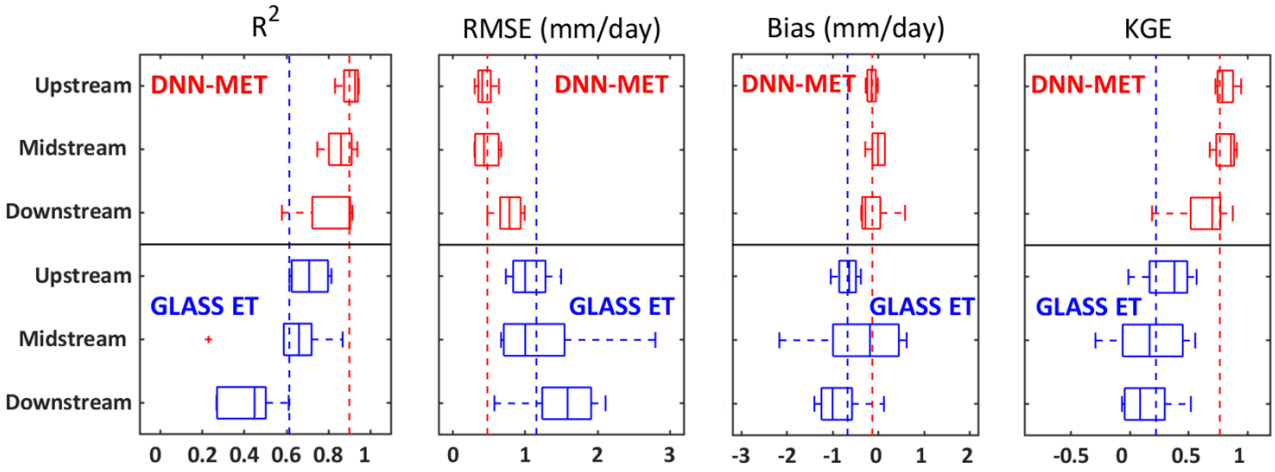


Fig. S2. Comparison of performances between DNN-MET and GLASS ET for the three reaches using $R^2$, RMSE, Bias and KGE at 8-day spatial resolution. The solid lines in the boxes indicate the median values, the box edges indicate the first and third quartile values, the end of the whiskers indicate values no greater than 1.5 times the interquartile value, and the most extreme data points beyond the whiskers indicate the extreme outliers. The red and blue dotted lines indicate the median values of the assessment metrics across the whole basin for DNN-MET and GLASS ET, respectively.

## Table S1

Information for the 19 EC flux tower sites in the Heihe River Basin, including the ID, site name, latitude N (Lat N), longitude E (Lon E), location, International Geosphere-Biosphere Programme land cover types (IGBP), elevation (m), and the duration of data collections.

| Site ID | Site name | Lat (N), Long (E) | Location | IGBP | Elevation (m) | Duration used in study |
|---|---|---|---|---|---|---|
| 1 | Arou | 38.0473, 100.4643 | Upstream | GRA | 3033 | Jun 2008 – Dec 2015 |
| 2 | Dashalong | 38.84, 98.9406 | Upstream | GRA | 3739 | Aug 2013 – Dec 2015 |
| 3 | Hulugou | 38.25, 99.8667 | Upstream | GRA | 3232 | Sep 2011 – Dec 2015 |
| 4 | Yakou | 38.0142, 100.2421 | Upstream | GRA | 4147 | Jan 2015 – Dec 2015 |
| 5 | Guantan | 38.5333, 100.2503 | Upstream | ENF | 2835 | Jan 2008 – Dec 2011 |
| 6 | Yingke | 38.8571, 100.4103 | Midstream | CRO | 1519 | Jan 2008 – Dec 2011 |
| 7 | Linze | 39.3281, 100.1408 | Midstream | CRO | 1399 | Apr 2013 – Dec 2014 |
| 8 | Jinta | 40.1722, 98.9287 | Midstream | CRO | 1252 | Jun 2008 – Aug 2008 |
| 9 | Daman | 38.8555, 100.3722 | Midstream | CRO | 1556 | Jun 2012 – Dec 2015 |
| 10 | Bajitan | 38.915, 100.3042 | Midstream | BAR | 1562 | Jun 2012 – Dec 2014 |
| 11 | Huazhaizi | 38.7652, 100.3186 | Midstream | BAR | 1731 | Jun 2012 – Dec 2015 |
| 12 | Shenshawo | 38.7892, 100.4933 | Midstream | BAR | 1694 | Jun 2012 – Dec 2014 |
| 13 | Zhangye wetland | 38.9751, 100.4464 | Midstream | WET | 1460 | Jun 2012 – Sep 2015 |
| 14 | Populus euphratica | 41.9932, 101.1239 | Downstream | DBF | 876 | Jul 2013 – Dec 2015 |
| 15 | Mixed forest | 41.9903, 101.1335 | Downstream | DBF | 874 | Jul 2013 – Dec 2015 |
| 16 | Sidaoqiao | 42.0012, 101.1374 | Downstream | SHR | 873 | Jul 2013 – Dec 2015 |
| 17 | Cropland | 42.0048, 101.1338 | Downstream | CRO | 875 | Jul 2013 – Oct 2015 |
| 18 | Barren land | 41.9993, 101.1326 | Downstream | BAR | 878 | Jul 2013 – Dec 2015 |
| 19 | Desert | 42.1137, 100.9872 | Downstream | BAR | 1054 | May 2015 – Dec 2015 |

## Table S2

Summary of the satellite-derived ET products used in this study. SWE refers to snow water equivalent, VF refers to vegetation fraction, and Pres refers to pressure.

| ET products | Method | Spatial resolution | Temporal resolution | Main forcing data | References |
|---|---|---|---|---|---|
| MOD16 | Penman–Monteith | 0.01 ° | daily | $R_n$, $T_a$, RH, LAI, FPAR | (Mu et al. 2011) |
| ET-SEMI | Penman–Monteith | 0.01 ° | daily | $R_s$, $T_a$, RH, WS, NDVI | (Wang et al. 2010) |
| ET-JPL | Priestley–Taylor | 0.01 ° | daily | $R_n$, $T_a$, RH, NDVI, LAI | (Fisher et al. 2008) |
| ET-MS | Priestley–Taylor | 0.01 ° | daily | $R_n$, $T_a$, DT, NDVI | (Yao et al. 2013) |
| ET-HF | Empirical Hybrid Formulation | 0.01 ° | daily | $R_n$, $T_a$, NDVI, SM | (Yao et al. 2019) |
| GLEAM | Priestley–Taylor | 0.25 ° | daily | Rad, $T_a$, P, SWE, VOD, SM, VF | (Martens et al. 2017) |
| ETMonitor | Shuttleworth–Wallace Dual-Source model | 0.01 ° | daily | Rad, $T_a$, Pres, RH, WS, P, NDVI, SM | (Hu and Jia 2015) |
| EB-ET | Surface Energy Balance System | 0.05 ° | daily | $T_a$, RH, WS, Pres, Rad, LST | (Chen et al. 2019) |
| GLASS ET | Bayesian model averaging method | 0.01 ° | 8-day | $R_s$, $R_n$, $T_a$, RH, WS, Albedo, LAI, NDVI | (Liang et al. 2020) |

## Table S3

Statistics of estimated daily ET against the eddy-flux tower observations. All correlation coefficients are significant with 99% confidence.

| ET products | $R^2$ | RMSE (mm/day) | Bias (mm/day) | KGE |
|---|---|---|---|---|
| MOD16 | 0.38 | 1.50 | -0.72 | 0.22 |
| ET-SEMI | 0.62 | 1.37 | -0.76 | 0.25 |
| ET-JPL | 0.41 | 1.58 | -0.87 | 0.15 |
| ET-MS | 0.59 | 1.42 | -0.80 | 0.21 |
| ET-HF | 0.60 | 1.12 | -0.39 | 0.57 |
| GLEAM | 0.24 | 1.72 | -0.93 | 0.11 |
| ETMonitor | 0.70 | 0.95 | -0.20 | 0.77 |
| EB-ET | 0.15 | 1.84 | -1.01 | 0.06 |
| SA | 0.63 | 1.33 | -0.71 | 0.28 |
| BMA | 0.68 | 1.19 | -0.59 | 0.39 |
| RF | 0.71 | 0.89 | -0.02 | 0.80 |
| DNN-MET | 0.76 | 0.82 | -0.02 | 0.83 |

## Algorithms for ET products

### Text S1. MOD16 model

The MOD16 product used in this study was generated based on the MODIS global ET algorithm framework proposed by Mu et al. (Mu et al. 2011). This algorithm is based on the Penman-Monteith equation, which can be written as:

$$ET = \frac{\Delta(R_n - G) + \rho C_p VPD / g_a}{\Delta + \gamma(1 + g_s / g_a)} \quad \text{(S1)}$$

where $\Delta$ is the slope of the saturation water vapor pressure curve, $\gamma$ is the psychrometric constant, and $\rho$ is the air density. The term $g_a$ is the aerodynamic resistance, and $g_s$ is the surface resistance. Mu et al. (Mu et al. 2011) set the different stomatal conductance constraints across all the biomes. Studies have shown that stable coefficients for different biomes can reduce the impact of incorrect classification of land cover types (Yao et al. 2017). Therefore, for all biomes, our study set the optimum air temperature ($T_o$) as 25 ℃ to express the temperature constraint ($m_T$) (Fisher et al. 2008; Yao et al. 2013) and set the $VPD$ values for nearly complete inhibition ($VPD_c$) and no inhibition ($VPD_o$) to transpiration as 650 Pa and 2900 Pa, respectively, to express the moisture constraint ($m_C$).

$$m_T = \exp\left[-\left(\frac{T_a - T_o}{T_o}\right)^2\right] \quad \text{(S2)}$$

$$m_C = \begin{cases} 1.0 & VPD \le VPD_o \\ \frac{VPD_c - VPD}{VPD_c - VPD_o} & VPD_o < VPD < VPD_c \\ 0.1 & VPD \ge VPD_c \end{cases} \quad \text{(S3)}$$

A detailed algorithm description can be obtained from Mu et al. (Mu et al. 2011).

### Text S2. ET-SEMI model

The ET-SEMI product was produced by a semiempirical Penman-based method developed by Wang et al. (Wang et al. 2010). The ET-SEMI calculates the total ET by combining the energy

control part ($ET_{ec}$) and the aerodynamic control part ($ET_{ac}$).

$$ET = k_1(ET_{ec} + ET_{ac}) + k_2(ET_{ec} + ET_{ac})^2 \quad \text{(S4)}$$

$$ET_{ec} = \frac{\Delta}{\Delta+\gamma} R_s[k_3 + k_4 NDVI + \left(1 - \frac{RH}{100}\right)(k_5 + k_6 NDVI)] \quad \text{(S5)}$$

$$ET_{ac} = \frac{\gamma}{\Delta+\gamma} WS[k_7 + \left(1 - \frac{RH}{100}\right)(k_8 + k_9 NDVI)]VPD \quad \text{(S6)}$$

The coefficient ($k_i, i = 1, \ldots 9$) is determined and validated using continuous measurements collected at 64 global sites. A major advantage of this model is that it avoids using the near-surface air temperature or humidity differences that are widely used in many other remote sensing ET algorithms (Wang et al. 2007). Consequently, the sensitivity to errors in the input data is substantially lessened (Wang et al. 2008).

**Text S3. ET-JPL model**

The ET-JPL model was produced by a novel Priestley–Taylor equation proposed by Fisher et al. (Fisher et al. 2008). ET-JPL divides the total ET into three parts: soil evaporation ($ET_{soil}$), canopy transpiration ($ET_{canopy}$) and interception evaporation ($ET_{interception}$).

$$ET = ET_{soil} + ET_{canopy} + ET_{interception} \quad \text{(S7)}$$

$$ET_{soil} = \alpha \frac{\Delta}{\Delta+\gamma}[f_w + f_s(1 - f_w)](R_{nSoil} - G) \quad \text{(S8)}$$

$$ET_{canopy} = \alpha \frac{\Delta}{\Delta+\gamma} f_c f_{pt} f_{pm}(1 - f_w) R_{nCanopy} \quad \text{(S9)}$$

$$ET_{interception} = \alpha \frac{\Delta}{\Delta+\gamma} f_w R_{nCanopy} \quad \text{(S10)}$$

$$f_w = RH^4 \quad \text{(S11)}$$

$$f_s = RH^{VPD} \quad \text{(S12)}$$

$$f_c = F_{APAR}/F_{IPAR} \quad \text{(S13)}$$

where $\alpha$ is the Priestley-Taylor coefficient (1.26), $f_w$ is the wet surface fraction, $f_s$ is the SM constraint, $f_c$ is the canopy fraction, $f_{pt}$ is the plant temperature constraint, and $f_{pm}$ is the plant

moisture constraint. A more detailed description of ET-JPL can be found in Fisher et al. (Fisher et al. 2008). The PT-JPL ET is produced operationally in the NASA ECOsystem Spaceborne Thermal Radiometer Experiment on Space Station (ECOSTRESS) mission (Fisher et al. 2020).

**Text S4. ET-MS model**

The ET-MS product was generated by a modified satellite-based Priestley–Taylor algorithm proposed by Yao et al. (Yao et al. 2013). ET-MS separates ET into four components: $ET_{canopy}$, $ET_{interception}$, unsaturated soil evaporation ($ET_{dSoil}$), and saturated wet soil evaporation ($ET_{wSoil}$). The ET-MS can be written as:

$$ET = ET_{dSoil} + ET_{wSoil} + ET_{canopy} + ET_{interception} \quad \text{(S14)}$$

$$ET_{dSoil} = \alpha \frac{\Delta}{\Delta+\gamma}(1 - f_w) f_s (R_{nSoil} - G) \quad \text{(S15)}$$

$$ET_{wSoil} = \alpha \frac{\Delta}{\Delta+\gamma} f_w (R_{nSoil} - G) \quad \text{(S16)}$$

$$ET_{canopy} = \alpha \frac{\Delta}{\Delta+\gamma} f_v f_{pt} (1 - f_w) R_{nCanopy} \quad \text{(S17)}$$

$$ET_{interception} = \alpha \frac{\Delta}{\Delta+\gamma} f_w R_{nCanopy} \quad \text{(S18)}$$

$$f_w = {f_s}^4 \quad \text{(S19)}$$

$$f_s = \left(\frac{1}{DT}\right)^{DT/DT_{max}} \quad \text{(S20)}$$

$$f_v = \frac{NDVI - NDVI_{min}}{NDVI_{max} - NDVI_{min}} \quad \text{(S21)}$$

where $f_v$ is the vegetation cover fraction. This method is constrained by NDVI and apparent thermal inertia (ATI), the latter of which is derived from the diurnal temperature range (DT) to avoid the complicated computations of aerodynamic resistance parameters (Yao et al. 2013).

**Text S5. ET-HF model**

The ET-HF product was generated using an empirical hybrid formulation parameterized by SM (Yao et al. 2019). This model characterizes ET using $LE/R_n$, which is a classical soil moisture and

energy-limited LE regime (Seneviratne et al. 2010). The ET-HF can be expressed as:

$$ET = R_n(a_1 + a_2 T_a + a_3 NDVI + a_4 f_s) \quad \text{(S22)}$$

where $a_i$ $(i = 1, \ldots 4)$ is the empirical coefficient, and $f_s$ refers to SM, RH or DT. The empirical coefficients for different $f_s$ terms were determined by surface and satellite observations. In this study, we used SM as the soil moisture constraint. The empirical coefficients for $a_i$ $(i = 1, \ldots 4)$ are 0.036, 0.006, 0.710 and 0.708, respectively (Yao et al. 2019).

**Text S6. GLEAM model**

The GLEAM model estimates the total ET based on four components: a Priestley-Taylor based potential ET module, a semiempirical stress module, a precipitation- and SM-driven soil module and a Gash model-based interception module (Martens et al. 2017). The Priestley-Taylor based actual ET module is expresses as:

$$ET = \alpha \frac{\Delta}{\Delta+\gamma} f(R_n - G) \quad \text{(S23)}$$

Where a is a Priestley-Taylor coefficient (1.26) under wet conditions, and $f$ is the environmental stress factor. The potential ET for the bare soil and vegetation canopy are converted into actual ET values by a stress factor based on VOD and root-zone SM. The in situ surface SM is assimilated into the soil profile to remove random forcing errors (Martens et al. 2017; Miralles et al. 2011). The canopy-intercepted ET is simulated by a Gash analytical model (Gash 1979). The latest global daily/monthly/yearly GLEAM ET products are available at www.gleam.eu.

**Text S7. ETMonitor model**

The ETMonitor monitors daily ET over the HRB by combining microwave and optical remote sensing observations for all sky conditions (Hu and Jia 2015). The rainfall interception is calculated based on the Gash model (van Dijk and Bruijnzeel 2001). The soil evaporation and vegetation

transpiration were calculated by the Shuttleworth–Wallace (Shuttleworth and Wallace 1985) model with a range of parameterized aerodynamic and bulk surface resistances:

$$\lambda E = C_c PM_{canopy} + C_s PM_{soil} \quad \text{(S24)}$$

$$PM_{canopy} = \frac{\Delta(R_n - G) + [\rho C_p VPD - \Delta r_{ac}(R_{nSoil} - G)]/(r_{aa} + r_{ac})}{\Delta + \gamma[1 + r_{sc}/(r_{aa} + r_{ac})]} \quad \text{(S25)}$$

$$PM_{soil} = \frac{\Delta(R_n - G) + (\rho C_p VPD - \Delta r_{as} R_{nCanopy})/(r_{aa} + r_{as})}{\Delta + \gamma[1 + r_{ss}/(r_{aa} + r_{as})]} \quad \text{(S26)}$$

where $PM_{canopy}$ and $PM_{soil}$ are components related to canopy transpiration and soil evaporation; $C_c$ is the canopy resistance coefficient; $C_s$ is the soil resistance coefficient; $C_p$ refers to the specific heat of air; $r_{ss}$ is the soil surface resistance; $r_{ac}$, $r_{aa}$ and $r_{as}$ are the aerodynamic resistances between the canopy and the canopy source height, the canopy source height and the reference height above the canopy, and the soil surface and the canopy source height, respectively. More detailed information about the algorithms and parameters used in ETMonitor can be obtained in Hu and Jia 2015 (Hu and Jia 2015). The 0.01 ° ETMonitor can be obtained from the Global Change Research Data Publishing & Repository (http://www.geodoi.ac.cn/WebCn/Default.aspx).

**Text S8. EB-ET model**

The EB-ET is produced based on the Monin–Obukhov similarity theory (MOST) and the SEBS model. To accurately simulate the sensible heat flux ($H$), an a priori estimate of roughness height and the excess resistance parameter $kB^{-1}$ (Chen et al. 2019; Chen et al. 2013) are required. The $H$ term can be written as:

$$H = ku\rho C_p(\theta_0 - \theta_a)\left[\ln\left(\frac{z-d}{z_{0h}}\right) - \Psi_h\left(\frac{z-d}{L}\right) + \Psi_h\left(\frac{z_{0h}}{L}\right)\right]^{-1} \quad \text{(S27)}$$

where $u$ is the friction velocity, $\theta_0$ is the potential temperature at the surface, $\theta_a$ is the potential air temperature at height $z$, $d$ is the zero plane displacement height, and $L$ is the Obukhov length. The excess resistance parameter can be expressed as:

$$z_{0h} = z_{0m}/exp(kB^{-1}) \quad \text{(S28)}$$

$$kB^{-1} = f_c^2 kB_c^{-1} + f_s^2 kB_s^{-1} + 2f_c f_s kB_m^{-1} \quad \text{(S29)}$$

where $z_{0h}$ and $z_{0m}$ are the roughness heights for heat transfer and momentum transfer, respectively; $kB_c^{-1}$ is the excess resistance parameter of the canopy; $kB_s^{-1}$ is the excess resistance parameter of bare soil; and $kB_m^{-1}$ is the excess resistance parameter for the mixed bare soil and canopy (Chen et al. 2013). The daily EB-ET product is available at the National Tibetan Plateau Data Center (https://data.tpdc.ac.cn/zh-hans/).

**Text S9. GLASS ET product**

The GLASS ET product uses BMA method to enhance global terrestrial ET estimation by merging five process-based ET products including MOD16 (Mu et al. 2011), RRS-PM (Yuan et al. 2010), PT-JPL (Fisher et al. 2008), MS-PT (Yao et al. 2013), and UMD-SEMI (Wang et al. 2010). 240 EC tower sites from FLUXNET across the world are used to calibrate this merging method (Liang et al. 2020). BMA adjusts the predictive probability density function (PDF) using ground-based observations to improve ET accuracy. The PDF can be written as:

$$PDF = \sum_{i=1}^{n} u_i \cdot p(y|M_i) \quad \text{(S30)}$$

where $u_i$ is a statistical weight that denotes the posterior probability for model $M_i$; $p(y|M_i)$ denotes the PDF using model $M_i$ alone; and $n$ is the total number of models used in BMA (Yao et al. 2014). The 0.01 °GLASS ET product with 8-day temporal resolution can be download from http://www.glass.umd.edu.

## References


Chen, X., Massman, W.J., & Su, Z. (2019). A Column Canopy-Air Turbulent Diffusion Method for Different Canopy Structures. *Journal of Geophysical Research: Atmospheres, 124*, 488-506

Chen, X., Su, Z., Ma, Y., Yang, K., Wen, J., & Zhang, Y. (2013). An Improvement of Roughness Height Parameterization of the Surface Energy Balance System (SEBS) over the Tibetan Plateau. *Journal of Applied Meteorology & Climatology, 52*, 607-622

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

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

Gash, J.H.C. (1979). An analytical model of rainfall interception by forests. *Quarterly Journal of the Royal Meteorological Society, 105*, 43-55

Hu, G., & Jia, L. (2015). Monitoring of Evapotranspiration in a Semi-Arid Inland River Basin by Combining Microwave and Optical Remote Sensing Observations. *Remote Sensing, 7*, 3056-3087

Liang, S., Cheng, J., Jia, K., Jiang, B., Liu, Q., Xiao, Z., Yao, Y., Yuan, W., Zhang, X., Zhao, X., & Zhou, J. (2020). The Global LAnd Surface Satellite (GLASS) product suite. *Bulletin of the American Meteorological Society*, 1-37

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

Miralles, D.G., Holmes, T.R.H., De Jeu, R.A.M., Gash, J.H., Meesters, A.G.C.A., & Dolman, A.J. (2011). Global land-surface evaporation estimated from satellite-based observations. *Hydrol. Earth Syst. Sci., 15*, 453-469

Mu, Q., Zhao, M., & Running, S.W. (2011). Improvements to a MODIS global terrestrial evapotranspiration algorithm. *Remote Sensing of Environment, 115*, 1781-1800

Seneviratne, S.I., Corti, T., Davin, E.L., Hirschi, M., Jaeger, E.B., Lehner, I., Orlowsky, B., & Teuling, A.J. (2010). Investigating soil moisture–climate interactions in a changing climate: A review. *Earth-Science Reviews, 99*, 125-161

Shuttleworth, W.J., & Wallace, J.S. (1985). Evaporation from sparse crops-an energy combination theory*, 111*, 839-855

van Dijk, A.I.J.M., & Bruijnzeel, L.A. (2001). Modelling rainfall interception by vegetation of variable density using an adapted analytical model. Part 1. Model description. *Journal of Hydrology, 247*, 230-238

Wang, Kaicun, Liang, & Shunlin (2008). An Improved Method for Estimating Global Evapotranspiration Based on Satellite Determination of Surface Net Radiation, Vegetation Index, Temperature, and Soil Moisture. *J Hydrometeor*

Wang, K., Dickinson, R.E., Wild, M., & Liang, S. (2010). Evidence for decadal variation in global terrestrial evapotranspiration between 1982 and 2002: 1. Model development. *Journal of Geophysical Research: Atmospheres, 115*

Wang, K., Wang, P., Li, Z., Cribb, M., & Sparrow, M. (2007). A simple method to estimate actual evapotranspiration from a combination of net radiation, vegetation index, and temperature. *Journal of Geophysical Research Atmospheres, 112*

Yao, Y., Liang, S., Cheng, J., Liu, S., & Fisher, J.B. (2013). MODIS-driven estimation of terrestrial latent heat flux in China based on a modified Priestley-Taylor algorithm. *Agricultural & Forest Meteorology, 171-172*, 187-202

Yao, Y., Liang, S., Li, X., Hong, Y., Fisher, J.B., Zhang, N., Chen, J., Cheng, J., Zhao, S., & Zhang, X. (2014). Bayesian multimodel estimation of global terrestrial latent heat flux from eddy covariance, meteorological, and satellite observations. *Journal of Geophysical Research, 119*, 4521-4545

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. (2017). Estimation of high-resolution terrestrial evapotranspiration from Landsat data using a simple Taylor skill fusion method. *Journal of Hydrology, 553*, 508-526

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

Yuan, W., Liu, S., Yu, G., Bonnefond, J.-M., Chen, J., Davis, K., Desai, A.R., Goldstein, A.H., Gianelle, D., Rossi, F., Suyker, A.E., & Verma, S.B. (2010). Global estimates of evapotranspiration and gross primary production based on MODIS and global meteorology data. *Remote Sensing of Environment, 114*, 1416-1431