https://doi.org/10.5194/essd-2021-456
Preprint. Discussion started: 1 March 2022

CAMELE: Collocation-Analyzed Multi-source Ensembled Land Evapotranspiration Data

Changming Li[1], Hanbo Yang[1*], Wencong Yang[1], Ziwei Liu[1], Yao Jia[1], Sien Li[2], Dawen Yang[1]

[1]State Key Laboratory of Hydroscience and Engineering, Department of Hydraulic Engineering, Tsinghua University. Beijing 100084, China

[2]Center for Agricultural Water Research in China, China Agricultural University, Beijing 100083, China

[*]Correspondence: Hanbo Yang (yanghanbo@tsinghua.edu.cn)

https://doi.org/10.5194/essd-2021-456
Preprint. Discussion started: 1 March 2022

Abstract

Land evapotranspiration (ET) is a key element of Earth's water-carbon system. Accurate estimation of global land ET is essential for better understanding of land-atmosphere interaction. Past decades have witnessed the generation of various ET products. However, the widely used products still contain inherent uncertainty induced by forcing inputs and imperfect model parameterizations. In addition, direct evaluation of ET products is not feasible due to the lack of sufficient global in-situ observations, which hinders our usage and assimilation. Hence, merging a global dataset as reliable benchmark and exploring evaluation method for ET products are of great importance. The aims of our study were as followed: (1) to design and validate a collocation-based method for ET merging; (2) to generate a long-term (1981-2020) ET product employing ERA5, FLUXCOM, PMLV2, GLDAS and GLEAM at 0.1°-8Daily and 0.25°-Daily resolutions. The produced Collocation-Analyzed Multi-source Ensembled Land Evapotranspiration Data (CAMELE) was then compared with others at point and regional scales. At the point scale, the results showed that the CAMELE performed well over different vegetation coverage. The accuracy of CAMELE was validated against in-situ observations with Pearson Correlation of 0.68, 0.62 and root mean square error of 0.84 and 1.03 mm/d on average over 0.1° and 0.25°, respectively. In terms of Kling-Gupta Efficiency, CAMELE ET obtained results superior (mean 0.52) to the second best ERA5 (mean 0.44) at 0.1° basis. For global comparison, the spatial distribution of multi-year average and annual variation were in consistent with others. Our merged product revealed increased ET in South Asia, Northwest Australia, and decreases in Amazon Plain and Congo Basin. The CAMELE products is freely available at https://doi.org/10.5281/zenodo.6283239 (Li et al., 2021).

https://doi.org/10.5194/essd-2021-456
Preprint. Discussion started: 1 March 2022

1. Introduction

Land Evapotranspiration, including transpiration, soil evaporation, and evaporation from canopy interception, is the key component of global terrestrial water and energy cycle (Jung et al., 2010; Lian et al., 2018), which accounts for about 60% of water cycle. As the intermediate variable of soil moisture affecting air temperature, accurate estimation of global land evapotranspiration is critical for understanding the hydrological cycle and land-atmosphere interaction (Miralles et al., 2019; Gentine et al., 2019). Thus, providing a reliable ET dataset as benchmark for further studies is of great importance.

During past decades, a great number of studies have investigated and developed different methods for the estimation of global land evapotranspiration, which leads to lots of datasets. Due to the difference in employed algorithms and principles, discrepancies are quite common among different simulations (Restrepo-Coupe et al., 2021; Han and Tian, 2020; Zhang et al., 2021b). In addition, evaluation of ET products is always challenging due to the lack of sufficient observations at global scale, which limits the direct uses of these data (Pan et al., 2020; Baker et al., 2021).

Products are often merged to mitigate their uncertainties. Recent studies have explored several approaches to integrate multiple ET products, including Simple Average (SA) (Ershadi et al., 2014), Bayesian Model Average (BMA) (Ma et al., 2020; Zhu et al., 2016), Reliability Ensemble Average method (REA) (Yoo et al., 2020), Empirical Orthogonal Functions (EOF) (Feng et al., 2016) and machine-learning based methods (Chen et al., 2020; Yin et al., 2021). The SA method assigns the same weight to each product, which is practically unreasonable; The BMA method requires a certain input of observations with high quality and relative dense distribution (Li et al., 2021); The EOF method requires high computational cost and may introduce uncertainty by unclear refactoring scheme (Le et al., 2017). Behind these methods, the main challenge is to calculate reliable weights for inputs based on a chosen "truth" (Gruber et al., 2020; Koster et al., 2021), either by averaging or

https://doi.org/10.5194/essd-2021-456
Preprint. Discussion started: 1 March 2022

introducing other relative geographical information as benchmark (Zhang et al., 2021a). In addition, previous research mostly focused on the estimation of ET at regional scale. For a global simulation, a more simple and reliable method is required.

Without the requirement of given reference, collocation analysis methods are widely used to estimate error variances and data-truth correlations by comparing across several independent data sources (Stoffelen, 1998; Gruber et al., 2020) . Lately, these methods have been widely applied in evaluation of various geographical variables estimates, including soil moisture (Chen et al., 2018; Dong et al., 2020c), precipitation (Li et al., 2018; Dong et al., 2020a), ocean wind speed (Ribal and Young, 2020), leaf area index (Jiang et al., 2017), total water storage (Baik et al., 2021), sea ice thickness and surface salinity (Hoareau et al., 2018), and near-surface air temperature (Sun et al., 2021). The original triple-collocation framework has been expanded in recent years. (Su et al., 2014) proposed an instrumental-variable based approach by using a temporally lag-1 time series of one product as the other independent product, which only requires double collocation and is referred as single instrumental variable algorithm, or IVS. Based on that, (Dong et al., 2019) achieved a more robust solution, denoted as double instrumental variable algorithm, or IVD. (Gruber et al., 2016) extends the original algorithm to include a fourth dataset (i.e., quadruple collocation or QC) and partially address the independent assumption to calculate a part of error cross-correlation (Vogelzang and Stoffelen, 2021). To combine the benefits of both double and quadruple collocation, (Dong et al., 2020b) recently proposed the extended double instrumental or EIVD.

In this study, we proposed a collocation-based data ensembled method to merge multiple ET products and produced the collocation-analyzed multi-source ensembled land evapotranspiration data, abbreviated as CAMELE. Merging framework was validated through synthetic experiments and validation against flux tower observations. By minimizing mean square error, the optimal weights for inputs were given using collocation-based evaluation results. Finally, our merged product was

https://doi.org/10.5194/essd-2021-456
Preprint. Discussion started: 1 March 2022

compared at point and global scale with others.

2. Data

Five widely used land ET products were selected, covering the period from 1980 to 2020. In addition, in-situ observations were employed for evaluation of the framework and further comparison of our merged product. The spatial and temporal resolutions of input datasets are shown in the following Table:

TABLE.1 Summary of products involved

Name	Schemes	Spatial resolution	Temporal resolution	Time Span	Reference
ERA5	IFS	0.1°&0.25°	Hourly	1980-present	(Hersbach et al., 2020)
GLDASv2.1	Noah	0.25°	Daily	1980-2019	(Rodell et al., 2004)
PMLV2	Penman-Monteith-Leuning	0.083°	8-Daily	2002-2019	(Zhang et al., 2019)
FLUXCOM	Machine learning	0.083°	8-Daily	2001-2013	(Jung et al., 2019)
GLEAMv3.3a	Priestley-Taylor	0.25°	Daily	1980-2017	(Miralles et al., 2011)

(1) ERA5

The European Centre for Medium-Range Weather Forecasts (ECMWF) provides ERA5-Land global hourly reanalysis dataset at various resolutions, covering the period 1981 to nearly present. ERA5-Land has been produced by replaying the land component of ECMWF ERA5 climate reanalysis. The atmospheric forcing data served an indirect influence as the constraint of the model-based estimates (Hersbach et al., 2020). Land evaporation is just one of the many output variables, which containing evaporation from bare soil, evaporation from open water surface excluding oceans, evaporation from the top canopy, evaporation from vegetation transpiration, snow evaporation, potential evaporation, and total evaporation. The dataset is freely available at the Climate Change service of Copernicus center (http://cds.climate.copernicus.eu). The accumulated total evaporation was downloaded and aggregated from hourly to daily timestep over 0.1° and 0.25°

https://doi.org/10.5194/essd-2021-456
Preprint. Discussion started: 1 March 2022

resolutions in this study.

(2) GLDASv2.1 Noah

The Global Land Data Assimilation System (GLDAS) product is a land surface simulation forced by a combination of model and observation datasets, which incorporates advanced and sophisticated data assimilation methodologies (Rodell et al., 2004). GLDAS runs multiple land surface models (LSMs), including Noah, Mosaic, Variable infiltration capacity (VIC) and Community land model (CLM). These combined models provide global estimation of evapotranspiration at both fine and coarse spatial (0.01° and 0.25°) and temporal (3-hourly and monthly) resolution. More complicated descriptions of the GLDAS products are available at NASA's Hydrology Data and Information Services Center (http://disc.sci.gsfc.nasa.gov/hydrology). In this study, we employed the GLDAS 2.1 Noah model at 0.25° spatial resolution with 3-hourly frequency. The 3-hourly data were then aggregated to daily timestep to match the consistence with other products.

(3) PMLV2

The Penman-Monteith-Leuning model version 2 global evaporation (PMLV2) is produced based on Penman-Monteith-Leuning model (Zhang et al., 2019). The PML model was first proposed by (Leuning et al., 2008), and further improved by (Zhang et al., 2010). The PML version 1 (PMLv1) is based on a biophysical model that considers canopy physiological processes and soil evaporation for the accurate estimation of surface conductance (G_s), which is the focus of PM-based method. It was further incorporated with a canopy conductance (G_c) model that coupled vegetation transpiration with gross primary productivity, resulting in the PML version 2 (PMLv2) (Gan et al., 2018). (Zhang et al., 2019) applied the PMLv2 model at global scale. The daily inputs include: (1) leaf area index (LAI), white sky shortwave albedo, and emissivity from Moderate Resolution Imaging Spectroradiometer (MODIS); (2) temperature variables ($T_{max}, T_{min}, T_{avg}$), instantaneous variables (P_{surf}, P_a, U, q), and accumulated variables (P_{rcp}, R_{ln}, R_s) from GLDAS. The evaporation is divided into direct evaporation from bare soil (E_s), evaporation from solid water (water body,

https://doi.org/10.5194/essd-2021-456
Preprint. Discussion started: 1 March 2022

snow, and ice) (ET_{water}), and vegetation transpiration (E_c). The PMLv2-ET is well-calibrated against 8-daily eddy covariance data from 95 global flux towers for ten land cover types (Kong et al., 2020).The data is freely available at the data center of institute of Tibetan Plateau Research, Chinese Academy of Sciences via application (https://data.tpdc.ac.cn/zh-hans/data/48c16a8d-d307-4973-abab-972e9449827c/?q=).

In this study, the 8-daily PMLv2 data were used and interpolated to 0.1° using the MATLAB Gaussian process regression package. The accumulated total evaporation is calculated as:

$$ET = E_s + E_c + ET_{water}$$

Data with abnormal value were removed.

(4) FLUXCOM

FLUXCOM is a machine-learning-based merging data of global land-atmosphere energy fluxes, which is the combination of remote sensing data and meteorological data (Jung et al., 2019). FLUXCOM uses several machine-learning-based regression tools, including tree-based methods, regression splines, neural networks, and kernel methods. The outputs were designed following two complementary strategies: (1) FLUXCOM-RS: merging exclusively remote sensing data to produce flux data with high spatial resolution. (2) FLUXCOM-RS+METEO: merging meteorological observations with remote sensing data at daily temporal resolution. The exclusive ensemble of RS data allows for generating gridded flux products at 500m spatial resolution, with relatively low frequency of 8-daily. Additionally, the FLUXCOM-RS data only cover the period after 2000 due to data availability. While the merging of meteorological data and remote sensing data extended the coverage (since 1980) with the cost of relatively coarser spatial resolution (0.5°). More detailed descriptions of the FLUXCOM dataset are available on the FLUXCOM website (http://fluxcom.org/). Data is freely available via contact.

In this study, we employed the FLUXCOM-RS 8-daily 0.0833° energy flux data and convert the latent heat to evaporation using ERA5-Land aggregated daily air temperature. The conversion follows the equation:

https://doi.org/10.5194/essd-2021-456
Preprint. Discussion started: 1 March 2022

$$ET = \frac{LE \times 30 \times 60}{(2.501 - 0.002361 \times T) \times 10^6}$$

Where ET is the evapotranspiration ($\mathrm{kg} \cdot \mathrm{m}^{-2} \cdot \mathrm{s}^{-1}$), LE is the latent heat flux ($\mathrm{W} \cdot \mathrm{m}^{-2}$), T is the air temperature (K). Furthermore, the original evaporation data were interpolated to 0.1° using the MATLAB Gaussian process regression package.

(5) GLEAM v3.3a

In this study, the Global Land Evaporation Amsterdam Model version 3.3a (GLEAM v3.3a) dataset (Miralles et al., 2011; Martens et al., 2017) at 0.25° are used. This version of GEAM provides daily estimation of actual evaporation (E), bare soil evaporation (E_b), canopy interception (E_i), transpiration from vegetation (E_t), potential evaporation (E_p), and snow sublimation (E_s) for the period 2003-2018. The data is freely available on VU university Amsterdam Geoservices website (http://geoservices.falw.vu.nl).

GLEAM is based on the Priestley-Taylor framework (Priestley and Taylor, 1972), which employs reanalysis temperature and radiation to estimate potential ET (PET). Furthermore, the PET is reduced to actual ET using remotely sensed soil moisture and vegetation optical-depth measurements. The GLEAM AET data was validated at 43 FLUXNET flux sites and had been proven to provide solid AET estimation (Majozi et al., 2017). Since ERA5-Land and GLEAMv3.3a both employ the ECMWF atmospheric reanalysis data, they may suffer the uncertainty from the same origin. However, due to the indirect influence of atmospheric data used in ERA5-Land, we can still assume that these two products are independent.

(6) In-situ observations

The FLUXNET2015 Tier 1 (http://fluxnet.fluxdata.org/) half-hourly eddy-covariance data are used in our study (Pastorello et al., 2020). After data filtering and processing, 82 sites are selected, and the observations are aggregated to daily timestep as reference data for evaluation of other products.

Following a filtering process by (Lin et al., 2019; Li et al., 2019), original hourly data is selected. Firstly, only the measured and good-quality gap-filled data are used for quality control. Secondly, to reduce the impact of canopy interception (Medlyn et al.,

https://doi.org/10.5194/essd-2021-456
Preprint. Discussion started: 1 March 2022

2017; Knauer et al., 2018), we excluded days with rainfall, as well as one extra subsequent day after rainy events. Thirdly, data records with negative GPP, ET and VPD were removed. When the number of valid half-hourly observations is higher than 38 (about 80%) per day, the daily total ET is calculated as:

$$ET = \frac{\sum_{i=1}^{N}(E_i \times 48)}{N}$$

$$ET_i = \frac{LE_i \times 30 \times 60}{(2.501 - 0.002361 \times T_i) \times 10^6}$$

Where N is the number of valid half-hourly observations; LE_i is the half-hourly observed latent heat flux ($\mathrm{W \cdot m^{-2}}$); T_i is the air temperature (K).

If the number of valid data is below 38, the daily value is set as fill value. Additionally, previous studies illustrated that FLUXNET2015 data suffered from an energy imbalance problem. Thus, following the method proposed by (Twine et al., 2000), the measured ET data are corrected. The sites are distributed globally, mostly located in North America and Europe. The International Geosphere–Biosphere Program (IGBP) land cover classification system (Loveland et al., 1999) is employed to distinguish the nine PFTs across sites, including evergreen needleleaf forests (ENF), evergreen broadleaf forests (EBF), deciduous broadleaf forests (DBF), croplands (CRO), grasslands (GRA), savannas (SAV), woody savannas (WSA), and mixed forests (MF). The selected sites cover the period from 2003 to 2017 and each has at least 3 years of reliable data. Detailed information are included in the Appendix.

3. Method

In our study, the merging process contained three steps: (1) uncertainty characterization of inputs using collocation analysis methods; (2) calculation of optimal weights for each product by minimizing the mean square error; (3) linear combination of inputs and products of merged product over various resolutions. Figure 1 represents the general process for data merging.

https://doi.org/10.5194/essd-2021-456
Preprint. Discussion started: 1 March 2022

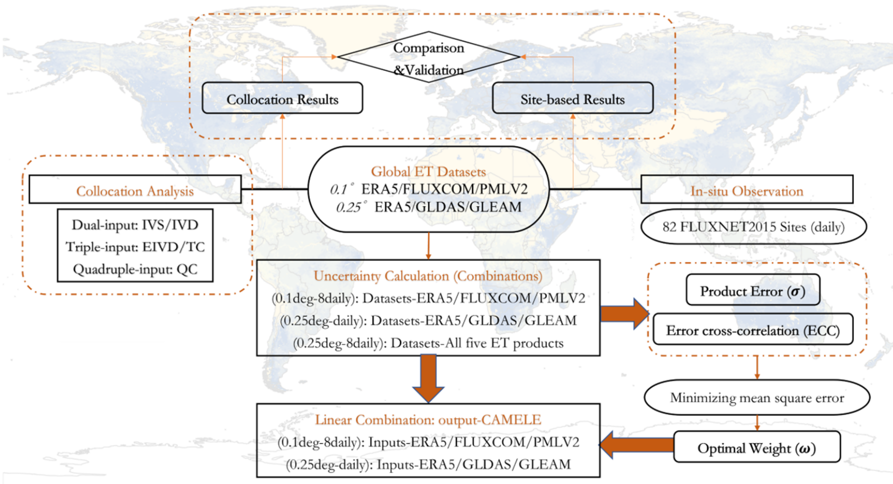


FIGURE.1 A flowchart for the data merging process

3.1 Uncertainty characterization

The challenge for the evaluation of global ET products is due to the lack of reliable benchmark. While the main advantage of collocation analysis methods is that no reference is required. In collocation analysis, independent products of a geophysical variable are typically assumed to be linearly related to the true signal (Mccoll et al., 2016). This linear model can be expressed as:

$$x = \beta_x P + B_x + \varepsilon_x$$

Where x is the product, P is the true signal; β_x and B_x are the ordinary least squares intercept and slope; and ε_x is zero-mean random error. This model is referred as the additive error structure model, while in practice, multiplicative error model in conjunction log transformation is more preferred (Li et al., 2018).

The basic assumptions adopted in collocation contain: (1) error orthogonality, assuming that the random error is independent with the true signal, which can be expressed as: $C_{P\varepsilon} = \overline{P\varepsilon} - \overline{P}\overline{\varepsilon} = 0$; (2) zero error cross-correlation, requiring the independence of each two products, which can be expressed as: $\overline{\varepsilon_x \varepsilon_y} = \overline{\varepsilon_y \varepsilon_z} = \overline{\varepsilon_x \varepsilon_z} = 0$; (3) the random error of each products is zero-mean, which means $\overline{\varepsilon} = 0$.

https://doi.org/10.5194/essd-2021-456
Preprint. Discussion started: 1 March 2022

Based on these assumptions, the covariances between the products and the Pearson correlation (R^2) of each product against the true signal can be solved.

The triple-collocation method (TC) requires a triplet of independent data sources (Stoffelen, 1998; Gruber et al., 2020) .The collocation analysis relies highly on the assumption that all datasets are mutually independent, which means error cross-correlation (ECC) is considered as zero (Gruber et al., 2020). As illustrated by (Yilmaz and Crow, 2014), the violation of zero ECC assumption usually results in underestimation of data errors. However, it is usually difficult to find three independent datasets in practice. To address the problem, (Su et al., 2014) proposed the instrumental-variable based approach by using a temporally lag-1 time series of one product as the third independent product, which only requires double collocation and is referred as single instrumental variable algorithm, or IVS. Based on that, (Dong et al., 2019) achieved a more robust solution, denoted as double instrumental variable algorithm, or IVD. (Gruber et al., 2016) extends the original algorithm to include a fourth dataset (i.e., quadruple collocation or QC) and partially address the independent assumption to calculate a part of ECCs (Vogelzang and Stoffelen, 2021). To combine the benefits of both double and quadruple collocation, (Dong et al., 2020b) recently proposed the extended double instrumental (EIVD), by which an ECC can be estimated using three datasets. Detailed deviations of each method were included in the Appendix.

To characterize the uncertainties of inputs, all five collocation analysis methods were employed at both 0.1° and 0.25°, daily and 8-daily resolution. Different methods can also be categorized by number of inputs: (1) Dual inputs (IVS/IVD); (2) Triple inputs (TC/EIVD); (3) Quadruple inputs (QC). For dual-input methods, IVS required the selection of product to derive the lag-1 series as the third input, while IVD used the lag-1 variances of both products. For triple-input methods, EIVD required the identification of two products with non-zero error-correlation-covariance, while TC assumed all three products were mutual-independent. For quadruple-input method, the

https://doi.org/10.5194/essd-2021-456
Preprint. Discussion started: 1 March 2022

requirement of QC was the same as EIVD. Therefore, taken the combinations over 0.25°/8-daily resolution for example, the number of combination scenarios for: (1) IVS: $\binom{5}{2} \times \binom{2}{1} = 20$; (2) EIVD: $\binom{5}{3} \times \binom{3}{2} = 30$; (3) QC: $\binom{5}{4} \times \binom{4}{2} = 30$. Detailed description of combinations could be found in the Appendix.

3.2 Calculation of Optimal Weights

Given specific variances of inputs, linear combination could serve as a simple and efficient solution for data assimilation. In this study, each product is assigned with the optimal weight (ω) that minimizing the mean square error (Bates and Granger, 1969; Kim et al., 2020) using error variances (σ_{ε_i}) and the ECC ($\sigma_{\varepsilon_i \varepsilon_j}$) as:

$$\omega_{ij} = \frac{\sigma_{\varepsilon_i}^2 - \sigma_{\varepsilon_i \varepsilon_j}^4 \sigma_{\varepsilon_i} \sigma_{\varepsilon_j}}{\sigma_{\varepsilon_i}^2 + \sigma_{\varepsilon_j}^2 - 2\sigma_{\varepsilon_i \varepsilon_j}^4 \sigma_{\varepsilon_i} \sigma_{\varepsilon_j}}$$

And the combined product θ_c is calculated as:

$$\theta_c = \sum_{i=1}^{N} \theta_i \omega_i$$

Where ω_i is the weighted arithmetic mean for each product, for a dual-input combination, the value of ω_i is calculated as:

$$\omega_i = \frac{\omega_{ij}}{\omega_{ij} + \omega_{ji}}$$

For a triple-input combination, the value of ω_i is given as:

$$\omega_i = \frac{\omega_{ij} + \omega_{ik}}{(\omega_{ij} + \omega_{ik}) + (\omega_{ji} + \omega_{jk}) + (\omega_{ki} + \omega_{kj})}$$

In addition, for a quadruple-input, the value is:

$$\omega_i = \frac{\omega_{ij} + \omega_{ik} + \omega_{im}}{(\omega_{ij} + \omega_{ik} + \omega_{im}) + (\omega_{ji} + \omega_{jk} + \omega_{jm}) + (\omega_{ki} + \omega_{kj} + \omega_{km}) + (\omega_{mi} + \omega_{mj} + \omega_{mk})}$$

3.3 Merging combination

The data were produced over 0.1°-8Daily and 0.25°-Daily resolutions based on evaluation results using IVD and EIVD algorithms. The selection of algorithm was

https://doi.org/10.5194/essd-2021-456
Preprint. Discussion started: 1 March 2022

based on the comparison results. The performance of IVD and EIVD generally outperformed other methods at chosen resolutions. In this study, each product is assigned with the optimal weight (ω) that minimizing the mean square error (Bates and Granger, 1969; Kim et al., 2020) using collocation-evaluated error variances (σ_{ε_i}) and the ECC ($\sigma_{\varepsilon_i\varepsilon_j}$). The table below showed the data used for merging during different period.

TABLE.2 Combination of inputs and accessible methods

Time Period	Available Products	Method
Scenario 1 (0.1°-8Daily)		
Time Period	**Available Products**	**Method**
(2001.01.01-2002.07.03)	ERA5/ FLUXCOM	IVD
(2002.07.04-2013.12.27)	ERA5/ FLUXCOM/ PMLV2	EIVD
(2013.12.28-2019.08.29)	ERA5/ PMLV2	IVD
Scenario 2 (0.25°-Daily)		
Time Period	**Available Products**	**Method**
(1981.01.01-2003.02.01)	ERA5/ GLEAM	IVD
(2003.02.02-2018.12.31)	ERA5/ FLUXCOM/ PMLV2	EIVD
(2019.01.01-2020.08.31)	ERA5/ GLDAS	IVD

4. Validation of framework

In this section, the validation of our framework was conducted as follow: (1) synthetic experiments were design to validate the merging route and provide information for the selection of proper collocation methods; (2) collocation-based evaluation results for inputs were compared against site-based evaluation using FLUXNET. Here, we used three indexes for comparison, including:

Pearson's Correlation (R^2)

$$R^2 = \frac{[\sum(S_i - \overline{S})(R_i - \overline{R})]^2}{\sum(S_i - \overline{S})^2 \sum(R_i - \overline{R})^2}$$

Root-mean-squared-error ($RMSE$)

https://doi.org/10.5194/essd-2021-456
Preprint. Discussion started: 1 March 2022

$$RMSE = \sqrt{\frac{\sum(S_i - R_i)^2}{N}}$$

Standard Deviation (SD)

$$SD = \sqrt{\frac{\sum(x_i - \overline{x})^2}{N}}$$

4.1 Synthetic validation

In our study, five synthetic experiments were designed to evaluate the merging route and compare different collocation analysis algorithms. Our aims were to (1) demonstrate that the merging data using linear combination based on minimizing mean square error outperforms any of the parent products; (2) compare the improvement of merging data over parent products using optimal weights derived by different methods; (3) evaluate how the violation of assumptions will impact the results; (4) illustrate that even though the assumptions are not perfectly valid, the merging product is still improved. Each experiment was repeated 1000 times using bootstrap to reduce the inherent uncertainty. In addition, only multiplicative error structure was employed in our experiments due to its applicability on collocation error characterization (Li et al., 2018; Gruber et al., 2020).

4.1.1 Design of synthetic experiment

A true signal (t) was generated following a passion distribution with a sample size of N: $t = 0.85P(0 \sim 50, N)$ (Kim et al., 2020). Then a collection of synthetic products, θ_i (i = 1, 2, 3, and 4), was generated by adding zero-mean Gaussian errors (ε_i) to t as:

$$\theta_i = t + \varepsilon_i$$

This synthetic model was employed for the five synthetic experiments, where each assumption was violated one at a time by generating a relative error (ε_i). In addition, six value of signal-to-noise rations ($SNR_{dB} = 0.1, 0.5, 1, 2, 5, and\ 10$) were adopted for the consideration of various noise range in the five synthetic experiments as:

https://doi.org/10.5194/essd-2021-456
Preprint. Discussion started: 1 March 2022

$$SNR = 10^{\frac{SNR_{dB}}{10}}, P_n = \frac{P_t}{SNR}$$

Where P_t and P_n are signal power and noise power, respectively. While P_n is used to generated noise with a specific variance according to the given SNR.

The first experiment (denote as "Exp1") was mean to test how the sample size (N) affects the result. Here, six sample sizes were chosen (50, 200, 600, 1000, 5000, and 10000) and the errors were expressed as:

$$\varepsilon_i = N(0, P_n)$$

Where $N(0, P_n)$ represents the Gaussian distribution with zero mean and variance of P_n.

The second experiment (denoted as "Exp2") was designed to analyze the violation of stationary assumption that random error of each product was assumed to be zero-mean. Here, we increased the error mean by linearly including an additive term as:

$$\varepsilon_i = a + \frac{H - 0.5N}{N} \times s \times E[t], H = 1{:}L$$

Where a is the random error generated using Gaussian distribution ($N(0, P_n)$; H is the monotonical increment ranging from 1 to N; s is the increasing slope with relation to the exception of true signal. The sample size is fixed as 800 for Exp2 and the remaining experiments based on the analysis from Exp1.

The third experiment (denoted as "Exp3") aimed to evaluate the impact of the violation of zero error correlation covariance assumption. Non-zero ECC indicated that products were not mutual-independent. Here, we considered two conditions: (1) fully correlated, where all products were dependent; (2) partly correlated, where only two products were related. Given that IVS and IVD algorithms only require two products, evaluations of dual-input methods were all under fully correlated consideration. The errors were defined as:

$$partly\ correlated: \begin{cases} \varepsilon_{1\sim 3} = a = N(0, P_n) \\ \varepsilon_4 = \sigma_{\varepsilon_1\varepsilon_4}\varepsilon_1 + \sqrt{1 - \sigma_{\varepsilon_1\varepsilon_4}^2}a \end{cases}$$

https://doi.org/10.5194/essd-2021-456
Preprint. Discussion started: 1 March 2022

$$fully\ correlated: \begin{cases} \varepsilon_1 = a = N(0, P_n) \\ \varepsilon_{2\sim4} = \sigma_{\varepsilon_1\varepsilon_{2\sim4}}\varepsilon_1 + \sqrt{1-\sigma^2_{\varepsilon_1\varepsilon_{2\sim4}}}a \end{cases}$$

Where $\sigma_{\varepsilon_i\varepsilon_j}$ is the error correlation covariance (ECC) and set as varying values (0.1, 0.2, 0.4, 0.6, 0.8, and 0.9).

The fourth experiment (denoted as "Exp4") tested how the combination results were changed when the error-truth orthogonality assumption was violated. The errors in Exp4 were derived as:

$$\varepsilon_i = \sigma_{\varepsilon_i t} \times \frac{t}{\sqrt{SNR}} + \sqrt{1-\sigma^2_{\varepsilon_i t}}a$$

Where $\sigma_{\varepsilon_i t}$ is the cross correlation between error and true signal and set as varying values (0.1, 0.2, 0.4, 0.6, 0.8, and 0.9).

The last experiment (denoted as "Exp5") investigated the effects of violation on zero autocovariance assumption. This assumption was adopted in IVS, IVD, and EIVD algorithms. The error of the one synthetic data was generated by a first-order autoregressive process with varying autocorrelation coefficients (0.1, 0.2, 0.4, 0.6, 0.8, and 0.9) using MATLAB simulation package.

$$\varepsilon_i = simulate\left(\left[AR\left('Constant', \bar{t}, 'AR', \rho_{lag}, 'Variance', P_n\right)\right], N\right)$$

Where $\bar{t}$ is the expectation of true signal; ρ_{lag} is the autocorrelation coefficient.

In summary, the designed aim and control parameter of each experiment is shown below:

TABLE.3 Description of synthetic experiments

No.	Related Assumption	Control Parameter
Exp0	Error noise	Signal-to-noise ratios (SNR_{dB})
Exp1	Sample representative	Sample Size (N)
Exp2	Stationarity	Non-stationary slope (s)
Exp3	Zero Error correlation covariance	Error-cross-correlation ($\rho_{\varepsilon_i\varepsilon_j}$)
Exp4	Zero Error-truth orthogonality	Error-truth cross-correlation ($\sigma_{\varepsilon t}$)
Exp5	Zero Error autocorrelation	Error autocorrelation (ρ)

https://doi.org/10.5194/essd-2021-456
Preprint. Discussion started: 1 March 2022

4.1.2 Synthetic Experiments Results

The synthetic results were validated based on the data-truth correlation ($\rho_{\theta_c t}$). Given that the synthetic data could not explicitly reflect the real situation, the analysis of results should focus on patterns and tendencies with changes in control parameters. Here, we focused on the improvement of combined product based on $\Delta\rho$:

$$\Delta\rho = \rho_{\theta_c t} - \max\left[\rho_{it}\right]$$

Here, positive $\Delta\rho$ represents improvements of the combined product, and vice versa. The analysis is twofold: (1) the results of all experiments under 0.1dB SNR are investigated; (2) the results of EIVD-based experiments under six SNR value are compared to illustrate the impact of noise.

https://doi.org/10.5194/essd-2021-456
Preprint. Discussion started: 1 March 2022

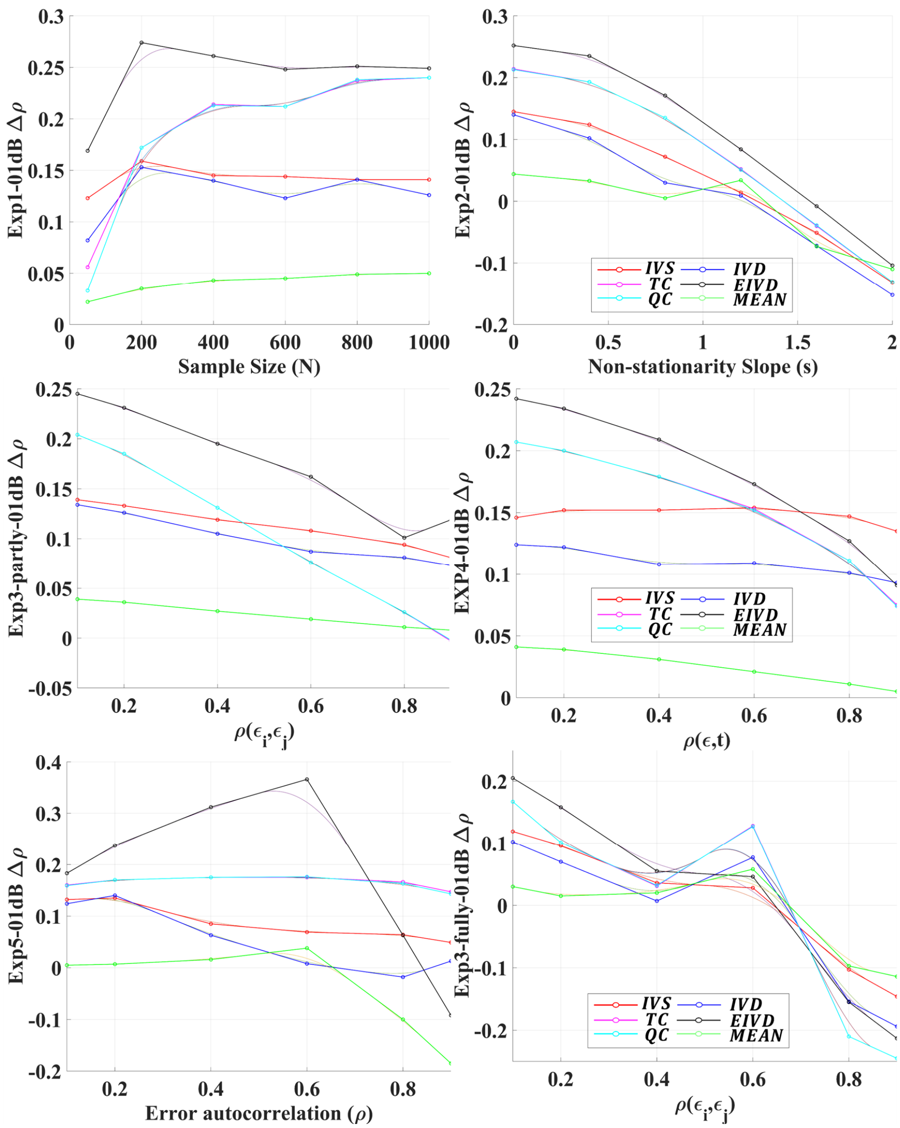


FIGURE.2 Results of all synthetic experiments with 0.1dB SNR. All methods were employed in the experiments. The legend of x axis in each figure represents the control parameter. The legend of y axis in all figures is the $\Delta\rho$. Lines in various colors corresponds to a method as marked in the legend box.

In Exp1, the value of $\Delta\rho$ raised with increasing sample size, while the variability reduced. Combined products based on all method present improvement, while EIVD-based product outperformed others. The expansion of data volume effectively reduced

https://doi.org/10.5194/essd-2021-456
Preprint. Discussion started: 1 March 2022

the uncertainty. Moreover, the results suggested that a sample size no less than 800 should be used for the combination.

Exp2 was designed to test the impact of stationarity assumption with increasing slope of error mean (s). With the augment of slope, performance of combined product based on all method turned down, including the simplest average method. When the slope was over 1.0, the combined products showed no improvement. However, this value of slope (s>1) indicated that the expectation of random error was one time greater than the mean of true signal, which was hardly possible in practice.

In Exp3, we compared the results under two different conditions: (1) fully correlated, where all products were dependent; (2) partly correlated, where only two products were related. The value of $\Delta\rho$ rapidly decreased with the augment of $\rho_{\varepsilon_i\varepsilon_j}$ under both conditions. When all products were correlated, the decreasing slope of $\Delta\rho$ was significantly steeper. Since zero-ECC assumption was hard to meet in practical situation, the finding suggested that the ECC should be carefully considered for a linear combination, as mentioned in previous studies (Gruber et al., 2020). In addition, the impact of ECC on combination was relatively lower under partly correlated condition. Thus, the selection of products was also essential for error characterization and combination.

The impact of error-truth orthogonality assumption was shown in Exp4. The improvement in $\Delta\rho$ was weakening when the true signal was more relevant with the random error. When the correlation (ρ) was over 0.9, which indicated that random error was highly correlated with true signal, the improvement in $\Delta\rho$ was nearly zero. However, given that the random error of data was usually considered independent with true signal in practice, the impact of error-truth orthogonality could not the main source of the method uncertainty (Yilmaz and Crow, 2014).

As for Exp5, we designed this experiment to investigate the impact of autocovariance. This assumption is related to IVS, IVD, and EIVD methods. The overall combination performance in $\Delta\rho$ declined compared with other experiments, and the value of $\Delta\rho$

https://doi.org/10.5194/essd-2021-456
Preprint. Discussion started: 1 March 2022

was slightly degraded when the error autocorrelation (ρ) became significant. The finding suggested that the consideration of error autocorrelations of products was necessary. However, evapotranspiration was less influenced by predominant condition (Zhang et al., 2011; Sharma et al., 2021). Thus, the impact of error-autocorrelation should not introduce much uncertainty.

Finally, comparing different methods, the results demonstrated that combination performance based on EIVD method generally outperformed other methods through synthetic experiments. As for the simplest averaging method, though previous studies recommended the average value as the proxy of reference (Pan et al., 2020; Burnett et al., 2020; Baker et al., 2021), our synthetic results demonstrated that the averaging method presented the lowest improvement. For gridded data, the equal weight for each grid assigned by averaging method ignored the spatial variability of different products, which could result in large uncertainty. In addition, since EIVD algorithm used the lag-1 series of two products, the violation of zero-error-autocorrelation assumption had the greatest impact than other methods.

https://doi.org/10.5194/essd-2021-456
Preprint. Discussion started: 1 March 2022

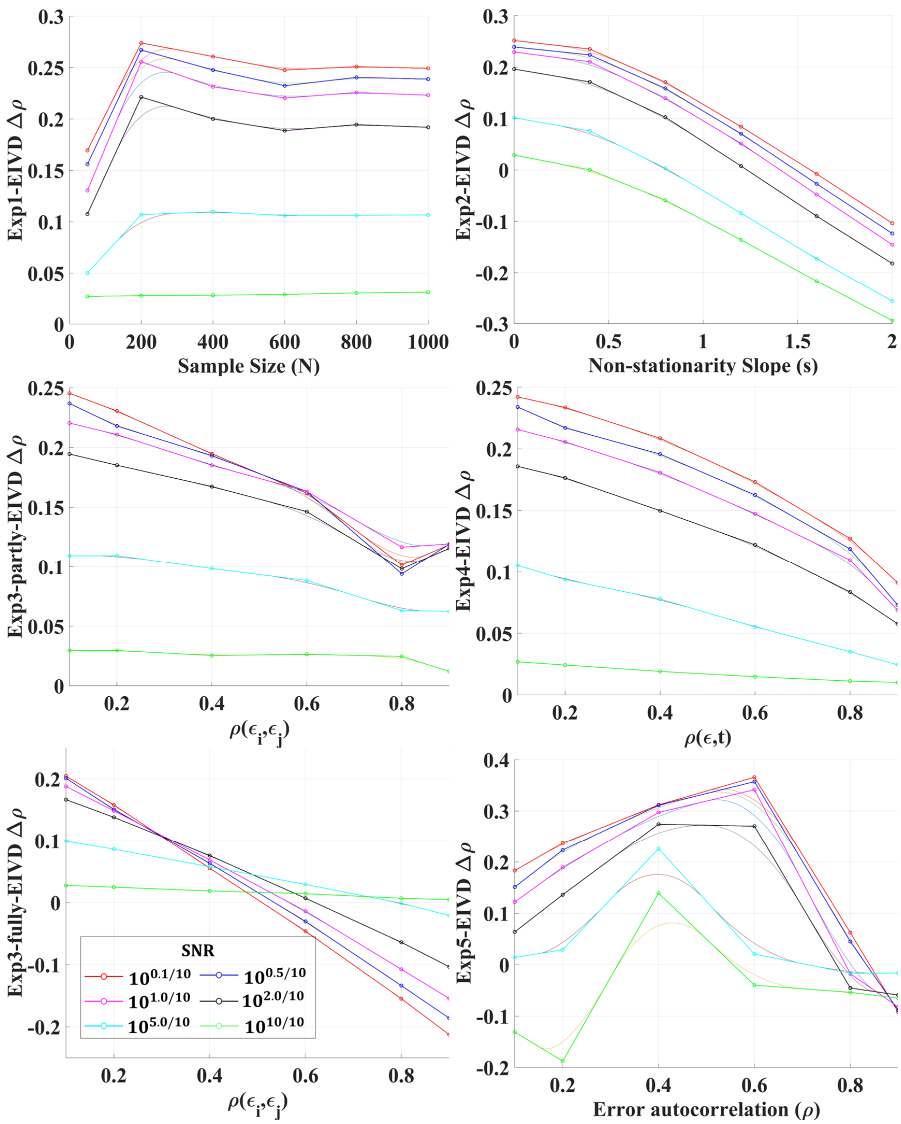


FIGURE.3 EIVD-based results of all synthetic experiments under six various SNR value to investigate the impact of error noise. The legend of x axis in each figure represents the control parameter. The legend of y axis in all figures is the $\Delta\rho$. Lines in various colors corresponds to a SNR value as marked in the legend box.

As shown in the figure, when the SNR_{dB} was over 1, the combination performance was dramatically degraded. The results illustrated that impact of noise was significant and was necessary to be taken into consideration. In practice, the SNR_{dB} is usually considered between 0.1 and 0.6 for products of geographical variables (Biscarini et al., 2021).

4.2 Site-based Validation of collocation-based evaluation results

Flux tower observation is the direct way to achieve the value of evapotranspiration, which is usually regarded as the reference for the assessments of products (Decker et al., 2012; Griebel et al., 2020). Due to the high cost for installation and maintenance of flux tower, the distribution and data period is scare at global scale. To prove that collocation analysis methods could be used as a reliable alternative when direct observations are not available, we compared the collocation-based evaluation results (as simulation) against the results based on flux tower observations (as reference). The comparisons were conducted over three resolutions: 0.1°-8Daily, 0.25°-Daily, and 0.25°-8Daily.

Table 4 presents the average value of Pearson's correlation (R^2) for five collocation analysis methods under all scenarios using multiplicative error structure. The comparison results demonstrated that collocation framework was reliable for the evaluation of ET products. Among the methods, IVD, EIVD, and QC were the three preferred methods for usage.

TABLE.4 Pearson's correlation (R^2) for different products using collocation analysis algorithms against evaluations based on in-situ observations.

	Resolution: 0.1°/8Daily			**Resolution: 0.25°/Daily**		
Methods	ERA5	FLUXCOM	PMLV2	ERA5	GLEAM	GLDAS
IVS	0.682	0.357	0.442	0.574	0.486	0.616
IVD	0.647	0.663	0.669	0.576	0.693	0.700
TC	0.712	0.035	0.574	0.649	0.645	0.691
EIVD	0.698	0.719	0.703	0.751	0.692	0.719

	Resolution: 0.25°/8Daily				
Methods	ERA5	FLUXCOM	PMLV2	GLEAM	GLDAS
IVS	0.699	0.542	0.523	0.675	0.686
IVD	0.683	0.643	0.546	0.631	0.693
TC	0.734	0.803	0.710	0.564	0.764
EIVD	0.717	0.722	0.608	0.800	0.748
QC	0.743	0.529	0.570	0.489	0.766

Furthermore, we presented the Taylor diagram (Taylor, 2001) to provide a way of

https://doi.org/10.5194/essd-2021-456
Preprint. Discussion started: 1 March 2022

graphically summarizing how closely the results match the reference. The similarity is quantified in terms of their correlation, their centered root-mean-square difference, and the amplitude of their variations (represented by their standard deviations). In this study, since calculated indexes were generally quite small (below 0.5), we multiplied the results of $RMSE$ and SD by 100 to magnify the difference for more intuitive contrast.

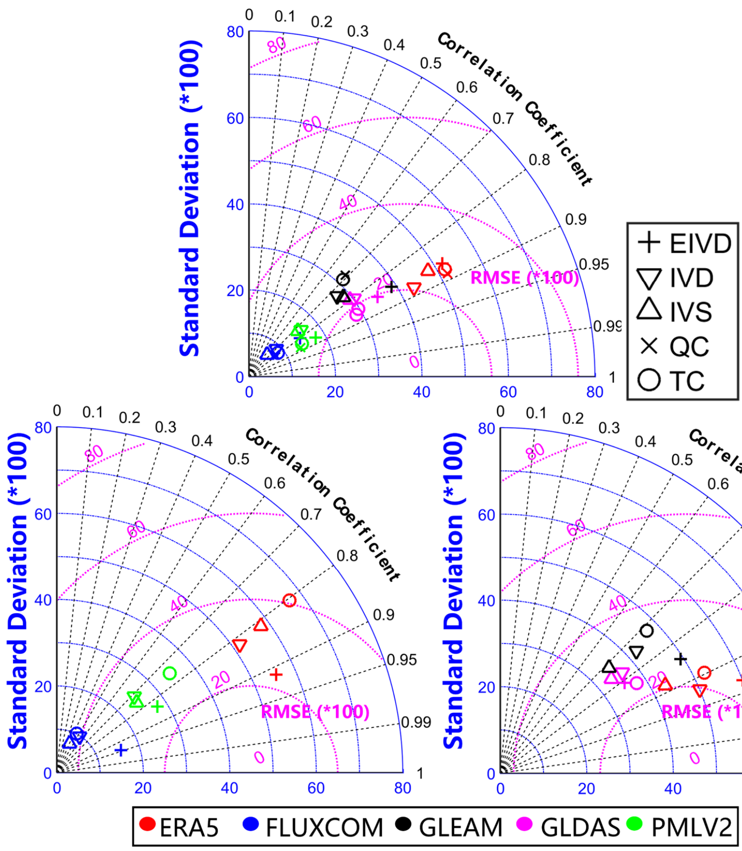


FIGURE.4 Taylor diagram for average collocation-based results against evaluations over tower observations of five products. Each color refers to one product and each shape represents one algorithm, as marked in the figure. The diagrams (from left to right) correspond to 0.1°-8Daily results, 0.25°-Daily results, and 0.25°-8Daily results, respectively.

https://doi.org/10.5194/essd-2021-456
Preprint. Discussion started: 1 March 2022

As shown in the figure, the average $RMSE$ was between 0.02-0.04 mm/d with mean R^2 over 0.8, manifesting the overall high accuracy of all products. Moreover, EIVD and IVD methods outperformed others with relatively higher correlation and lower difference, while TC showed the highest $RMSE$. However, since all the indexes were quite small, it could be concluded that results by any method matched the reference well.

In addition, to select the proper error structure and assess the performance of algorithms over various resolution, we compared the ERA5-results based on two error models over two scales. Each column referred to one resolution.

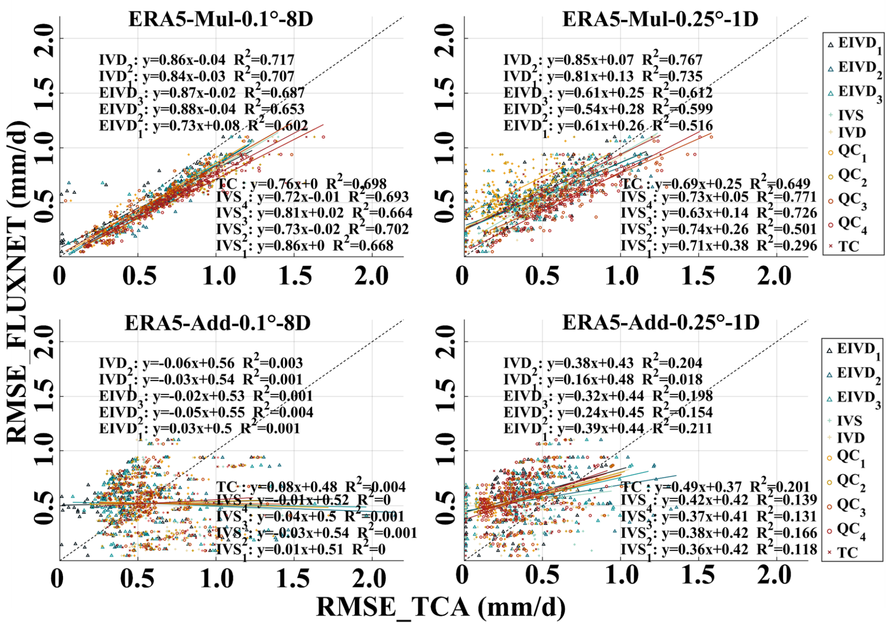


FIGURE.5 Comparison of collocation analysis results for ERA5 using multiplicative (first row) or additive (second row) error structure at various resolutions against evaluations over observations

As shown in the figure, there was a poignant contrast of performance between two error models. Multiplicative structure was proven to be a better description of the error-truth-relation, which was consistent with previous research (Yilmaz and Crow, 2014; Li et al., 2018). In the contrast, evaluations using additive model showed

https://doi.org/10.5194/essd-2021-456
Preprint. Discussion started: 1 March 2022

dramatical deviation against observation-based results. Thus, we only employed the multiplicative error model for further calculation and analysis.

In general, the comparison results demonstrated that collocation methods were reliable for the evaluation of evapotranspiration products. Multiplicative error model was more suitable for the description of error-truth-relation. Among the methods, IVD, EIVD, and QC were the three preferred methods for usage.

5. Results

In this section, we conducted the comparison between merged product and others in two steps: (1) validation at point scale using 82 selected FLUXNET sites; (2) comparison of global spatial distribution and variation trend.

5.1 Site-based Validation

The validation results of CAMELE against flux tower observations for different land cover types were shown in Figure 6 - 7. The average accuracy of CAMELE was about 0.68, 0.62 of correlation and 0.84, 1.03 mm/d of RMSE over 0.1° and 0.25°, while ERA5 and GLEAM were the second best for 0.1° and 0.25° with 0.66 and 0.61 of correlation. In general, the merged product revealed well performance over all land cover types with some variations. The results also indicated that the uncertainties of products increased over coarser resolution with obvious higher relative bias and lower correlation. Moreover, slight overestimation of merged product was found over 0.25° for generally all land cover types. Since tower or gauge could only cover the variation of geographical variables over a certain range (Tang et al., 2018), pixel-based evaluation on 0.1° and 0.25° should consider the inherent uncertainty of in-situ observations, which may explain the increased bias found in our comparison. In general, the validation against flux tower data demonstrated the overall high accuracy of CAMELE over various land cover types.

https://doi.org/10.5194/essd-2021-456
Preprint. Discussion started: 1 March 2022

Open Access Earth System Science Data Discussions

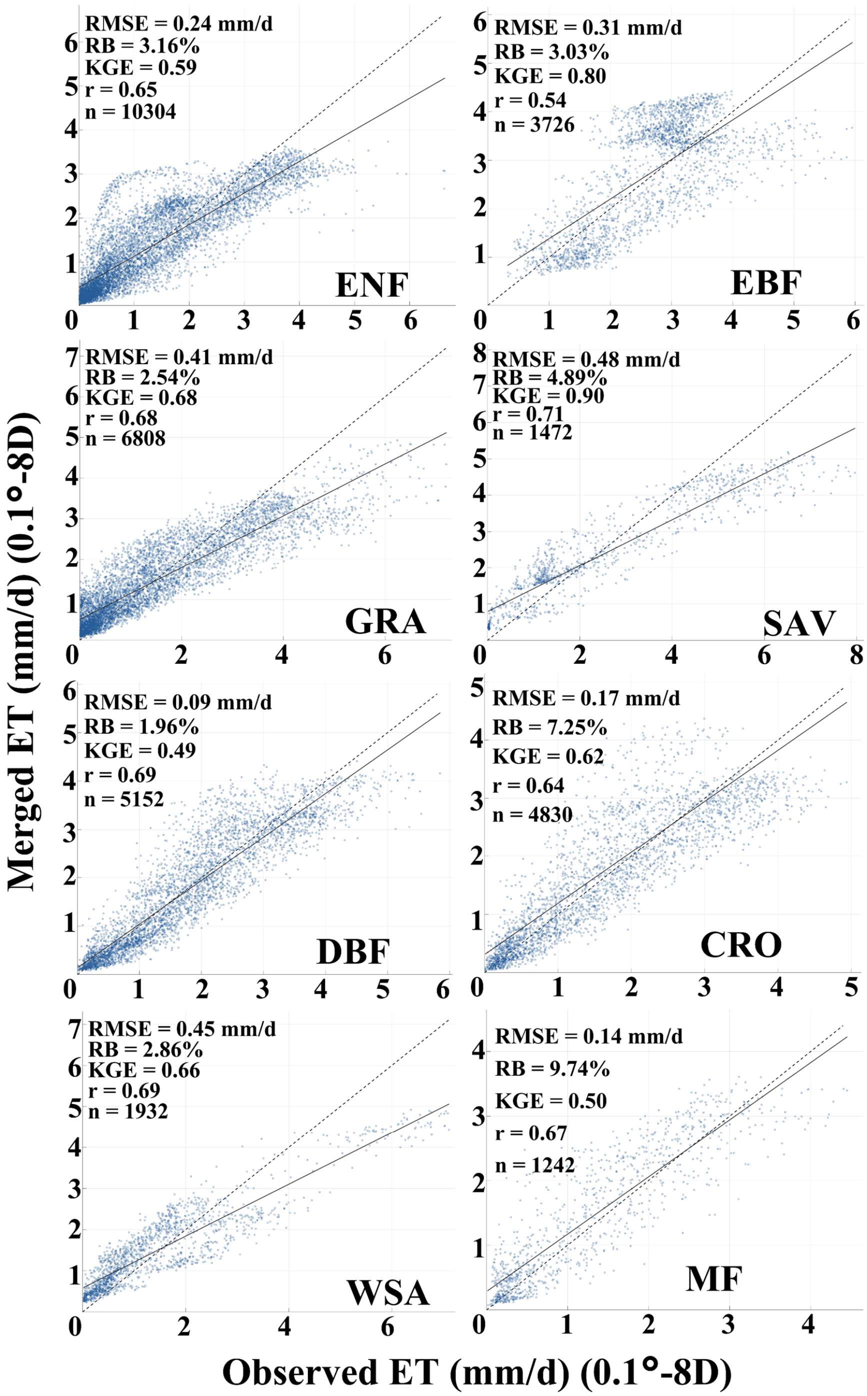
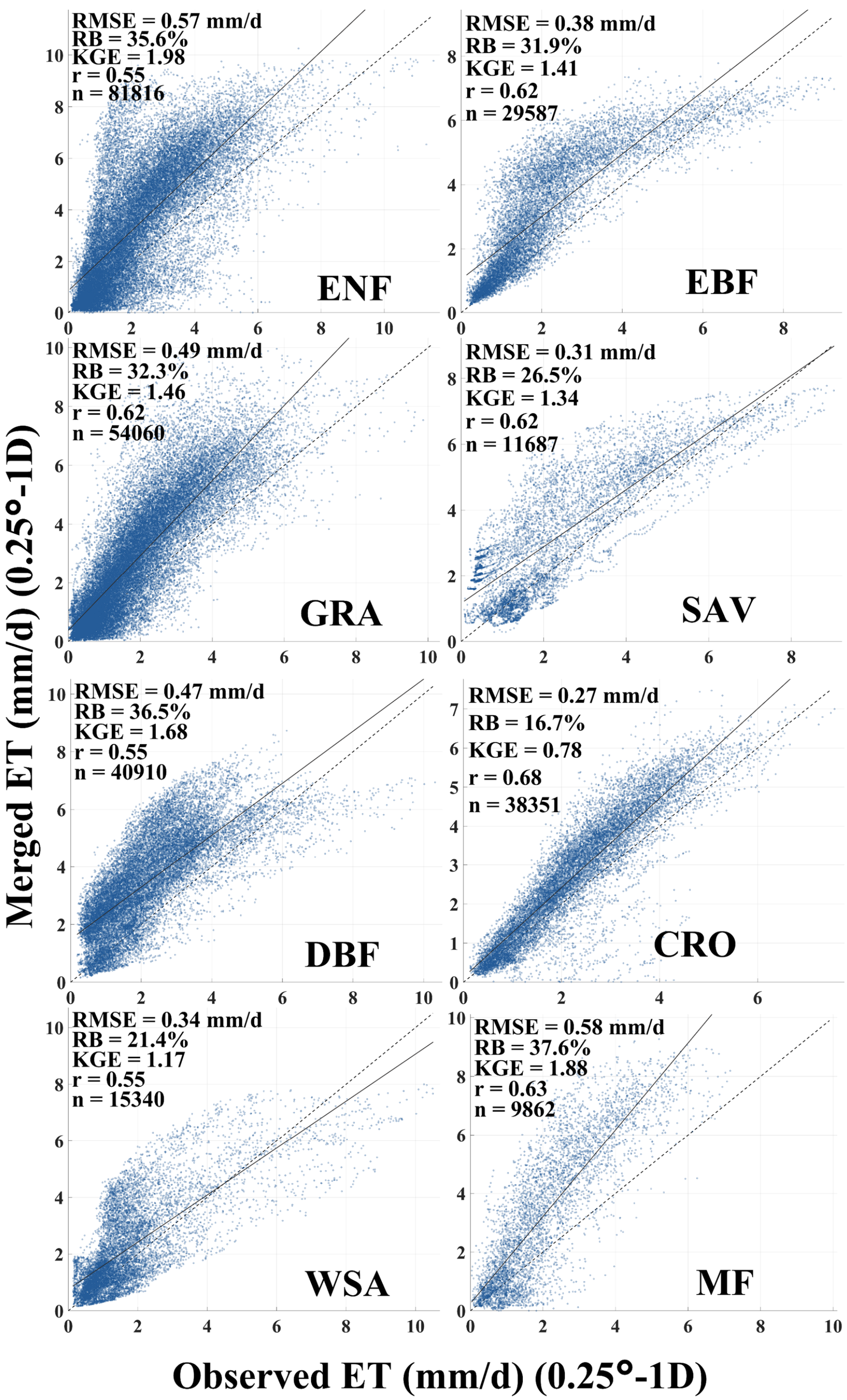


FIGURE.6 Validation results of CAMELE against flux tower observations for

https://doi.org/10.5194/essd-2021-456
Preprint. Discussion started: 1 March 2022

evergreen needleleaf forests (ENF), evergreen broadleaf forests (EBF), deciduous broadleaf forests (DBF), croplands (CRO), grasslands (GRA), savannas (SAV), woody savannas (WSA), and mixed forests (MF) over 0.1°-8Daily resolution

https://doi.org/10.5194/essd-2021-456
Preprint. Discussion started: 1 March 2022



530

https://doi.org/10.5194/essd-2021-456
Preprint. Discussion started: 1 March 2022

FIGURE.7 Validation results of CAMELE against flux tower observations for evergreen needleleaf forests (ENF), evergreen broadleaf forests (EBF), deciduous broadleaf forests (DBF), croplands (CRO), grasslands (GRA), savannas (SAV), woody savannas (WSA), and mixed forests (MF) over 0.25°-Daily resolution

Here we also presented the comparison of all products over some sites and employ the Kling-Gupta Efficiency (KGE) for better description. The KGE (Gupta et al., 2009) addressed several shortcomings in Nash-Sutcliffe Efficiency (NSE) and were increasingly used for calibration and evaluation (Knoben et al., 2019), given by:

$$KGE = 1 - \sqrt{(r-1)^2 + \left(\frac{\sigma_{sim}}{\sigma_{obs}} - 1\right)^2 + \left(\frac{\mu_{sim}}{\mu_{obs}} - 1\right)^2}$$

Where σ_{obs} and σ_{sim} are the standard deviations of observations and simulations; μ_{obs} and μ_{sim} are the mean of observations and simulations. Similar with NSE, KGE = 1 indicates perfect agreement of simulations while KGE<0 reveals that the average of observations is better than simulations (Towner et al., 2019; Knoben et al., 2019).

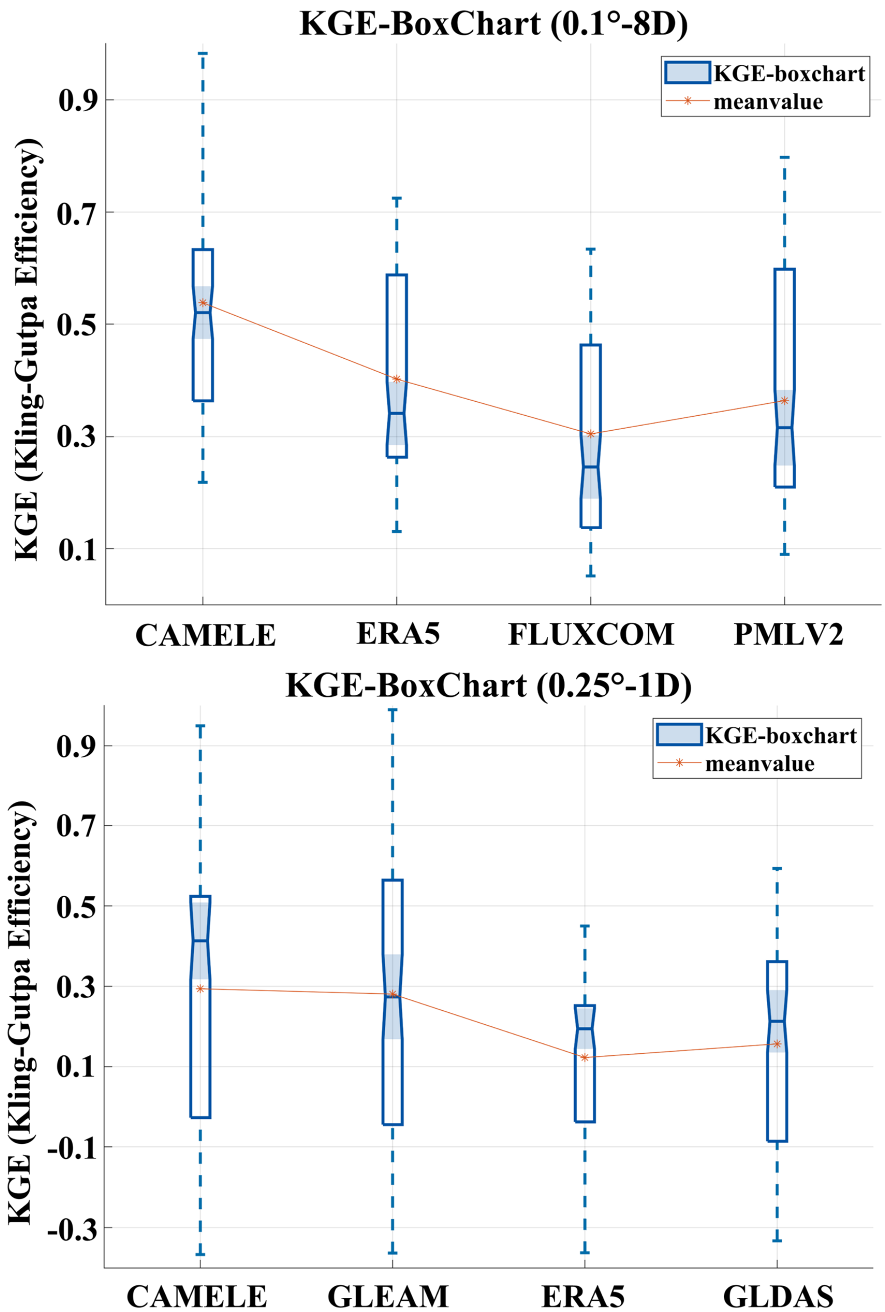


FIGURE.8 Boxplots of KGE for all products over two resolutions against 82 sites.

In terms of KGE, CAMELE obtained results superior (mean KGE=0.52) to the

https://doi.org/10.5194/essd-2021-456
Preprint. Discussion started: 1 March 2022

Open Access Earth System Science Data Discussions

second best ERA5 (mean KGE=0.44) at 0.1° basis. In general, our merged product performed well with precise reflection of the fluctuation. At FR-Gri site, some underestimation of PMLV2 and overestimation of FLUXCOM over peak value were observed over 0.1°. At US-Wkg site over 0.25°, all products showed high performance except for ERA5 with significant overestimation. To some extent, our merged product integrated the advantages of all inputs and addressed the overestimation and underestimation of peak value.

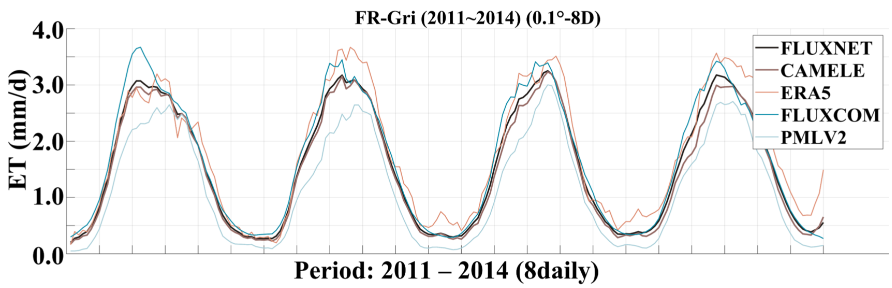


FIGURE.9 Performance of different products at 0.1°-8Daily resolution on FR-Gri site from 2011 to 2014.

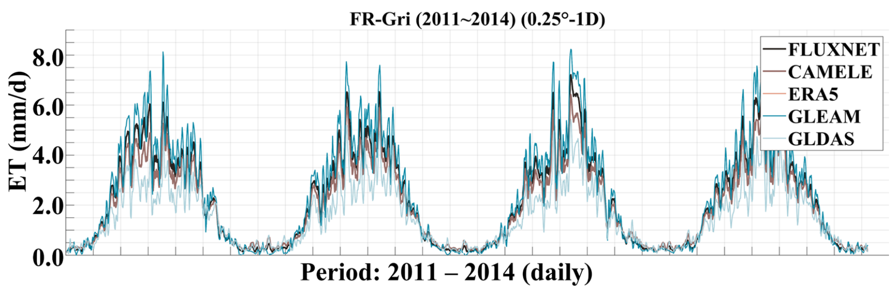


FIGURE.10 Performance of different products at 0.25°-Daily resolution on FR-Gri site from 2011 to 2014.

https://doi.org/10.5194/essd-2021-456
Preprint. Discussion started: 1 March 2022

Earth System Science Data
Open Access
Discussions

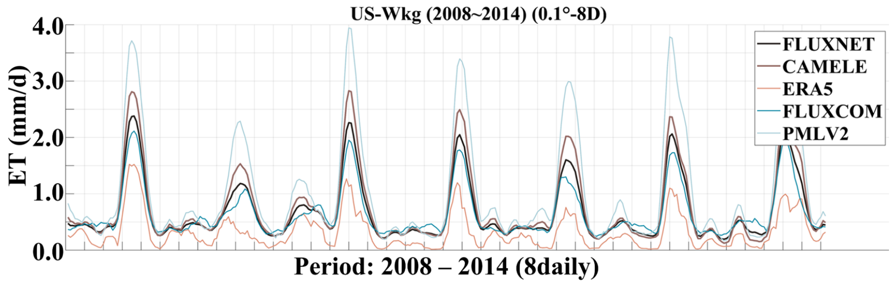


FIGURE.11 Performance of different products at 0.1°-8Daily resolution on US-Wkg site from 2008 to 2014.

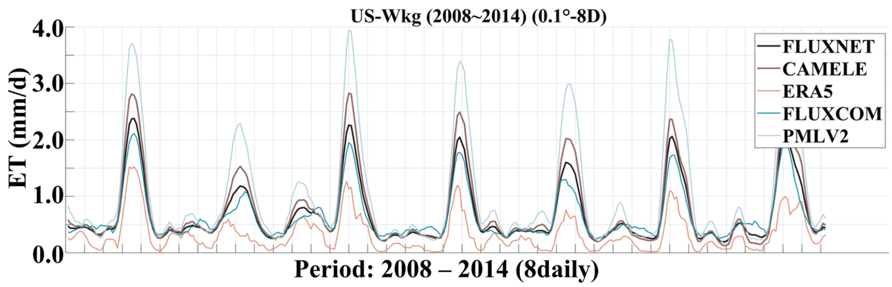


FIGURE.12 Performance of different products at 0.25°-1Daily resolution on US-Wkg site from 2008 to 2014.

5.2 Spatial distribution

Figure 13 - 14 depicted the spatial distribution of multi-year daily average ET, and the results were found consistent among different products over two resolutions. High evaporation regions were near the equators, including the Amazon Plain in South America, the Congo Basin in central Africa and the border between Asia and Oceania, with high precipitation (usually over 1000 mm per year). As for extreme low value, they were distributed in permafrost regions or dry desert, like the Sahara and Arabian deserts in North Africa, permafrost regions in North America and Eurasia. Compared to CAMELE, estimations by ERA5 and FLUXCOM were higher in wet regions near the equator while value of PMLV2 was slightly lower over 0.1°; estimations by GLEAM was significant higher near the equators and the value by GLDAS was the

https://doi.org/10.5194/essd-2021-456
Preprint. Discussion started: 1 March 2022

lowest among these wet regions. Since the available period varied among different products, we only showed a general comparison and future studies could consider more detailed regional investigation.

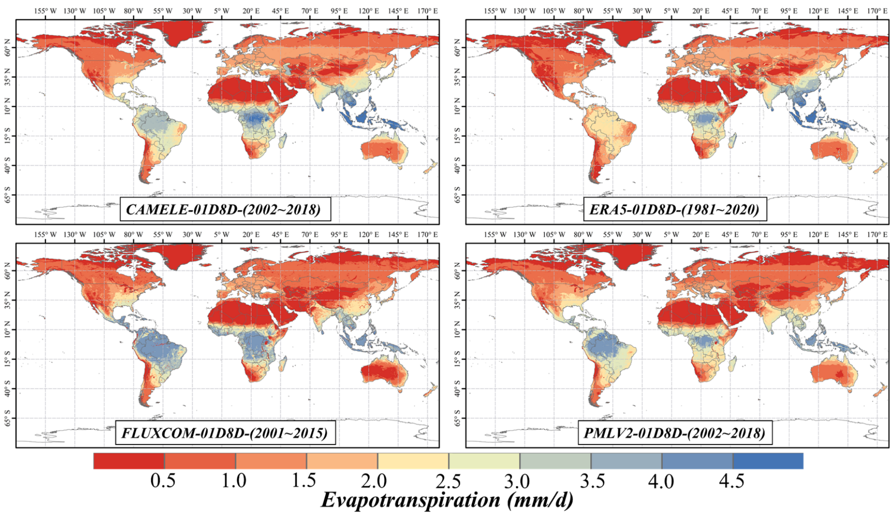


FIGURE.13 Spatial distribution of long-period average daily land evaporation for different products over relative period over 0.1°-8Daily resolution

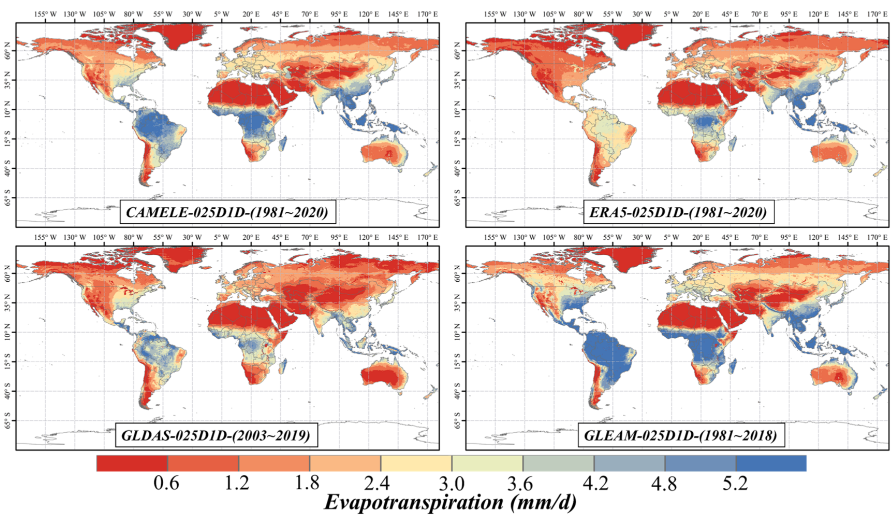


FIGURE.14 Spatial distribution of long-period average daily land evaporation for different products over relative period over 0.25°-8Daily resolution

Figure 15 - 16 presented the annual variation trends of multiple products during 2002-

https://doi.org/10.5194/essd-2021-456
Preprint. Discussion started: 1 March 2022

2015 and 2003-2017. Over 0.1°, a decrease in ET was found in Amazon Plain and Congo Basin by our merged product, while the increase regions were indicated in South Asia and the West Australia. The reduction over Amazon Plain was also found in ERA5 and FLUXCOM, while PMLV2 showed a rising trend. The decreasing trend over Congo Basin was consistent with ERA5 and FLUXCOM, still an opposite one by PMLV2. (Burnett et al., 2020) demonstrated that Congo Basin had become drier and less humid in recent years based on the analysis of environmental data. Our results showed the same trend.

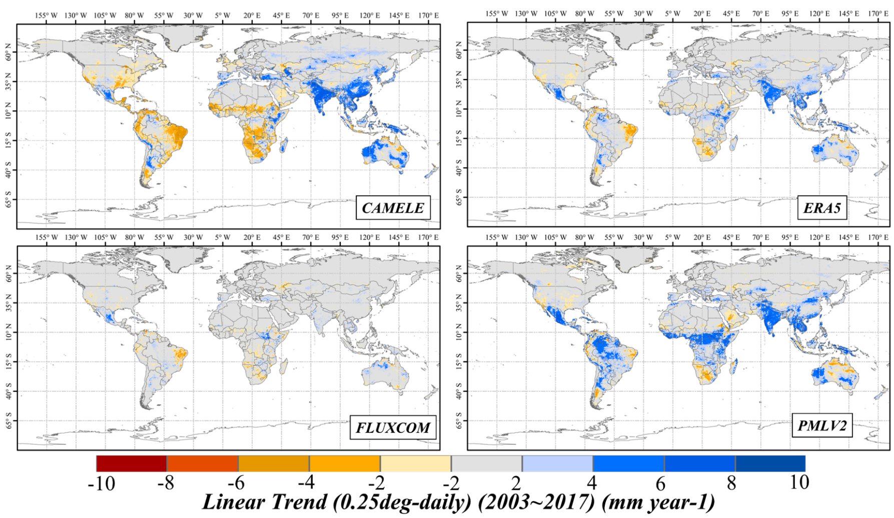


FIGURE.15 Spatial distribution of linear annual trends of land evapotranspiration of different products from 2002 to 2015 over 0.1°-8Daily resolution

https://doi.org/10.5194/essd-2021-456
Preprint. Discussion started: 1 March 2022

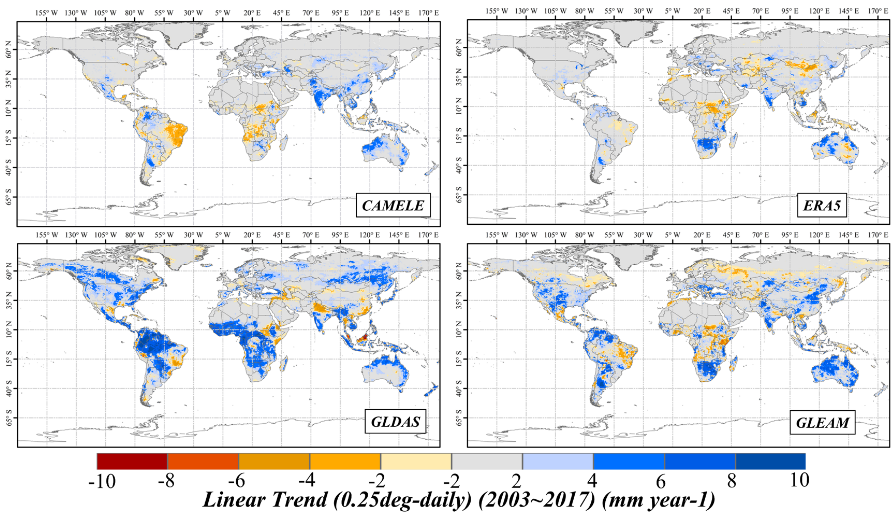


FIGURE.16 Spatial distribution of linear annual trends of land evapotranspiration of different products from 2003 to 2017 over 0.25°-8Daily resolution

Over 0.25°, the variation by merged product followed the similar patterns with the one over 0.1° with smaller value. This may be explained by the changing of resolution (from 0.1° to 0.25°), which included more pixels into one grid that neutralize the total variation. The general distribution was quite consistent with that of GLEAM, which had been proven with high accuracy, especially over tropical Africa (Liu et al., 2016; Wang et al., 2020). Decreases on Amazon Plain and Congo Basin were also revealed with opposite finding by GLDAS. (Burnett et al., 2020) found GLDAS with the maximum temporal variability among the selected products in their study, especially over Congo Basin. The increases in South Asia and the coastline of Australia were detected in all products.

6. Conclusion

In this study, we proposed a collocation-based data merging method and generated a long-period (1980-2020) CAMELE ET product over 0.1°-8Daily and 0.25°-Daily resolutions by merging five widely used datasets, including ERA5, FLUXCOM, PMLV2, GLDAS and GLEAM. The optimal weights were calculated using evaluations of inputs by collocation methods. The error characterizations were then proven be reliable against evaluations by in-situ observations. In addition, a series of synthetic experiments were design to validate our merging framework. Further, we conducted a comparison between CAMELE and other products at site-based and regional scales. To sum up, our conclusions were as follow:

1. Collocation analysis methods could serve as a reliable tool for evaluation of ET products without given reference, which provides promising future for error characterization especially over data-scare region or analysis at global scale. The evaluation results could provide important information for data merging.
2. The CAMELE product revealed general good performance at point scale. Compared to in-situ observations, the Pearson Correlation of 0.68 and 0.62 value of CAMELE over 0.1° and 0.25° resolutions are higher than the second best for relative resolution (0.66 for ERA5 and 0.61 for GLEAM). In addition to Kling-Gupta Efficiency, the merged product obtained superior mean value of 0.52, compared to 0.44 for ERA5 at 0.1° basis.
3. The spatial distributions of multi-year average daily ET and annual variation trend were generally similar to others. Results by CAMELE indicated a decrease in ET over Amazon Plain and Congo Basin, as consistent with the finding by ERA5 and GLEAM. Increases were found in South Asia and Northwest Australia. Our merged product well described the variation of global ET with combining advantages of the input products.

The optimal weight for each product was calculated using collocation-based evaluation results. Thus, the uncertainty may come from biased evaluation due to the violation of mathematical assumptions employed by collocation methods, especially the zero ECC assumption. While in our study, the ECC results by EIVD and QC demonstrated that this impact was within acceptable range since the general value of ECC was quite low as presented in the Appendix. Moreover, though random error caused by changing combinations may bring additional uncertainty, previous studies have showed that the variance over difference combination was quite small (Li et al., 2018). Thus, this may not bring much error and could be considered in further study. To sum up, our proposed collocation-based data merging method revealed promising potential for the merging of ET products. The merged CAMELE ET showed general well performance over site-based and regional scales, which could satisfy the requirement of more detailed research. In future studies, to improve the quality of merged product, dynamic weights could be calculated by adopting suitable merging period for different products and more complicated combination schemes could be considered to improve the accuracy.

Author Contribution

LC and YH designed the research. LC designed the merging method and performed most of calculation and analysis work. LC completed the manuscript and YW, LZ, YX, YH, LS and YD contributed to the revising and polishing of this paper.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

We sincerely thank a lot of researchers for their generous help as list below, in no particular order, including: (1) Prof. Stephen Sitch for providing trendy-v8 data and giving suggestion on its usage; (2) Dr. Jianzhi Dong for answering our questions on the employment of EIVD and IVD methods; (3) Dr. Dongdong Kong for providing the latest version of PMLV2; (4) Dr. Peter Weston for further instruction on the usage of MSE merging method; (5) Dr. Liming He for providing the BEPS-ET data and further suggestion on the evaluation of different products; (6) Dr.Youngryel Ryu for providing BESS-ET data. (7) Dr. Martin Jung and Dr. Ulrich Weber for addressing

https://doi.org/10.5194/essd-2021-456
Preprint. Discussion started: 1 March 2022

the data portal for the usage of FLUXCOM. Though some products provided were not used in this study, it was still very helpful for our selection of inputs.

Financial support

This research has been supported by the China National Key R&D Program (grant no. 2021YFC3000202) and the National Natural Science Foundation of China (grant no. 51979140, 42041004)

Data availability

All data used in this study could be accessed via the links described in Data Section. CAMELE products is freely available at https://doi.org/10.5281/zenodo.6283239 (Li et al., 2021) over 0.1°-8Daily and 0.25°-Daily resolutions. The data are distributed under a Creative Commons Attribution 4.0 License.

Reference

Baik, J., Park, J., Hao, Y., and Choi, M.: Integration of multiple drought indices using a triple collocation approach, Stochastic Environmental Research and Risk Assessment, 1-19, 2021.

Baker, J. C. A., Garcia-Carreras, L., Gloor, M., Marsham, J. H., Buermann, W., da Rocha, H. R., Nobre, A. D., de Araujo, A. C., and Spracklen, D. V.: Evapotranspiration in the Amazon: spatial patterns, seasonality, and recent trends in observations, reanalysis, and climate models, Hydrology and Earth System Sciences, 25, 2279-2300, 10.5194/hess-25-2279-2021, 2021.

Bates, J. M. and Granger, C. W.: The combination of forecasts, Journal of the Operational Research Society, 20, 451-468, 1969.

Biscarini, M., Vittimberga, A., De Sanctis, K., Fabio, S. D., Montagna, M., Milani, L., Tsuda, Y., and Marzano, F. S.: Optimal Stochastic Prediction and Verification of Signal-to-Noise Ratio and Data Rate for Ka-Band Spaceborne Telemetry Using Weather Forecasts, IEEE Transactions on Antennas and Propagation, 69, 1065-1077, 10.1109/tap.2020.3016865, 2021.

Burnett, M. W., Quetin, G. R., and Konings, A. G.: Data-driven estimates of evapotranspiration and its controls in the Congo Basin, Hydrology and Earth System Sciences, 24, 4189-4211, 10.5194/hess-24-4189-2020, 2020.

Chen, F., Crow, W. T., Bindlish, R., Colliander, A., Burgin, M. S., Asanuma, J., and Aida, K.: Global-scale evaluation of SMAP, SMOS and ASCAT soil moisture products using triple collocation, Remote Sensing of Environment, 214, 1-13, 10.1016/j.rse.2018.05.008, 2018.

Chen, Z., Zhu, Z., Jiang, H., and Sun, S.: Estimating daily reference evapotranspiration based on limited meteorological data using deep learning and classical machine learning methods, Journal of Hydrology, 591, 10.1016/j.jhydrol.2020.125286, 2020.

Decker, M., Brunke, M. A., Wang, Z., Sakaguchi, K., Zeng, X., and Bosilovich, M. G.: Evaluation of the Reanalysis Products from GSFC, NCEP, and ECMWF Using Flux Tower Observations, Journal of Climate, 25, 1916-1944, 10.1175/jcli-d-11-00004.1, 2012.

Dong, J., Lei, F., and Wei, L.: Triple collocation based multi-source precipitation merging, Frontiers in Water, 2, 1, 2020a.

Dong, J., Crow, W. T., Duan, Z., Wei, L., and Lu, Y.: A double instrumental variable method for geophysical product error estimation, Remote Sensing of Environment, 225, 217-228, 10.1016/j.rse.2019.03.003, 2019.

Dong, J., Wei, L., Chen, X., Duan, Z., and Lu, Y.: An instrument variable based algorithm for estimating cross-correlated hydrological remote sensing errors, Journal of Hydrology, 581, 124413, 10.1016/j.jhydrol.2019.124413, 2020b.

Dong, J., Dirmeyer, P. A., Lei, F., Anderson, M. C., Holmes, T. R. H., Hain, C., and Crow, W. T.: Soil Evaporation Stress Determines Soil Moisture-Evapotranspiration Coupling Strength in Land Surface Modeling, Geophysical Research Letters, 47, 10.1029/2020gl090391, 2020c.

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

Feng, Y., Cui, N., Zhao, L., Hu, X., and Gong, D.: Comparison of ELM, GANN, WNN and empirical models for estimating reference evapotranspiration in humid region of Southwest China, Journal of Hydrology, 536, 376-383, 10.1016/j.jhydrol.2016.02.053, 2016.

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

Gentine, P., Massmann, A., Lintner, B. R., Hamed Alemohammad, S., Fu, R., Green, J. K., Kennedy, D., and Vilà-Guerau de Arellano, J.: Land–atmosphere interactions in the tropics – a review, Hydrology and Earth System Sciences, 23, 4171-4197, 10.5194/hess-23-4171-2019, 2019.

Griebel, A., Metzen, D., Pendall, E., Burba, G., and Metzger, S.: Generating spatially robust carbon budgets from flux tower observations, Geophysical Research Letters, 47, e2019GL085942, 2020.

Gruber, A., Su, C. H., Zwieback, S., Crow, W., Dorigo, W., and Wagner, W.: Recent

https://doi.org/10.5194/essd-2021-456
Preprint. Discussion started: 1 March 2022

advances in (soil moisture) triple collocation analysis, International Journal of Applied Earth Observation and Geoinformation, 45, 200-211, 10.1016/j.jag.2015.09.002, 2016.

Gruber, A., De Lannoy, G., Albergel, C., Al-Yaari, A., Brocca, L., Calvet, J. C., Colliander, A., Cosh, M., Crow, W., Dorigo, W., Draper, C., Hirschi, M., Kerr, Y., Konings, A., Lahoz, W., McColl, K., Montzka, C., Muñoz-Sabater, J., Peng, J., Reichle, R., Richaume, P., Rüdiger, C., Scanlon, T., Van Der Schalie, R., Wigneron, J. P., and Wagner, W.: Validation practices for satellite soil moisture retrievals: What are (the) errors?, Remote Sensing of Environment, 244, 111806, 10.1016/j.rse.2020.111806, 2020.

Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, Journal of hydrology, 377, 80-91, 2009.

Han, S. and Tian, F.: A review of the complementary principle of evaporation: from the original linear relationship to generalized nonlinear functions, Hydrology and Earth System Sciences, 24, 2269-2285, 10.5194/hess-24-2269-2020, 2020.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J. N.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999-2049, 10.1002/qj.3803, 2020.

Hoareau, N., Portabella, M., Lin, W., Ballabrera-Poy, J., and Turiel, A.: Error characterization of sea surface salinity products using triple collocation analysis, IEEE Transactions on Geoscience and Remote Sensing, 56, 5160-5168, 2018.

Jiang, C., Ryu, Y., Fang, H., Myneni, R., Claverie, M., and Zhu, Z.: Inconsistencies of interannual variability and trends in long-term satellite leaf area index products, Global Change Biology, 23, 4133-4146, 2017.

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

Jung, M., Reichstein, M., Ciais, P., Seneviratne, S. I., Sheffield, J., Goulden, M. L., Bonan, G., Cescatti, A., Chen, J., and De Jeu, R.: Recent decline in the global land evapotranspiration trend due to limited moisture supply, Nature, 467, 951-954, 2010.

Kim, S., Pham, H. T., Liu, Y. Y., Marshall, L., and Sharma, A.: Improving the Combination of Satellite Soil Moisture Data Sets by Considering Error Cross Correlation: A Comparison Between Triple Collocation (TC) and Extended

https://doi.org/10.5194/essd-2021-456
Preprint. Discussion started: 1 March 2022

Double Instrumental Variable (EIVD) Alternatives, IEEE Transactions on Geoscience and Remote Sensing, 1-11, 10.1109/tgrs.2020.3032418, 2020.

Knauer, J., Zaehle, S., Medlyn, B. E., Reichstein, M., Williams, C. A., Migliavacca, M., De Kauwe, M. G., Werner, C., Keitel, C., Kolari, P., Limousin, J. M., and Linderson, M. L.: Towards physiologically meaningful water-use efficiency estimates from eddy covariance data, Glob Chang Biol, 24, 694-710, 10.1111/gcb.13893, 2018.

Knoben, W. J. M., Freer, J. E., and Woods, R. A.: Technical note: Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores, Hydrology and Earth System Sciences, 23, 4323-4331, 10.5194/hess-23-4323-2019, 2019.

Kong, D., Gu, X., Li, J., Ren, G., and Liu, J.: Contributions of Global Warming and Urbanization to the Intensification of Human‐Perceived Heatwaves Over China, Journal of Geophysical Research: Atmospheres, 125, 10.1029/2019jd032175, 2020.

Koster, R. D., Liu, Q., Reichle, R. H., and Huffman, G. J.: Improved Estimates of Pentad Precipitation Through the Merging of Independent Precipitation Data Sets, Water Resources Research, 57, 10.1029/2021wr030330, 2021.

Le, H., Yang, N., Liu, L., Chen, Y., and Zhang, H.: The latitudinal structure of nighttime ionospheric TEC and its empirical orthogonal functions model over North American sector, Journal of Geophysical Research: Space Physics, 122, 963-977, 10.1002/2016ja023361, 2017.

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

Li, C., Tang, G., and Hong, Y.: Cross-evaluation of ground-based, multi-satellite and reanalysis precipitation products: Applicability of the Triple Collocation method across Mainland China, Journal of Hydrology, 562, 71-83, 10.1016/j.jhydrol.2018.04.039, 2018.

Li Changming, Yang Hanbo, Yang Wencong, Liu Ziwei, Jia Yao, Li Sienand Yang Dawen: CAMELE: Collocation-Analyzed Multi-source Ensembled Land Evapotranspiration Data, https://doi.org/10.5281/zenodo.6283239, 2021.

Li, K., Tian, F., Khan, M. Y. A., Xu, R., He, Z., Yang, L., Lu, H., and Ma, Y.: A high-accuracy rainfall dataset by merging multiple satellites and dense gauges over the southern Tibetan Plateau for 2014–2019 warm seasons, Earth System Science Data, 13, 5455-5467, 10.5194/essd-13-5455-2021, 2021.

Li, X., Gentine, P., Lin, C., Zhou, S., Sun, Z., Zheng, Y., Liu, J., and Zheng, C.: A simple and objective method to partition evapotranspiration into transpiration and evaporation at eddy-covariance sites, Agricultural and Forest Meteorology, 265, 171-182, 10.1016/j.agrformet.2018.11.017, 2019.

Lian, X., Piao, S., Huntingford, C., Li, Y., Zeng, Z., Wang, X., Ciais, P., McVicar, T. R., Peng, S., and Ottlé, C.: Partitioning global land evapotranspiration using

CMIP5 models constrained by observations, Nature Climate Change, 8, 640-646, 2018.

Lin, C., Gentine, P., Frankenberg, C., Zhou, S., Kennedy, D., and Li, X.: Evaluation and mechanism exploration of the diurnal hysteresis of ecosystem fluxes, Agricultural and Forest Meteorology, 278, 10.1016/j.agrformet.2019.107642, 2019.

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

Loveland, T. R., Zhu, Z., Ohlen, D. O., Brown, J. F., Reed, B. C., and Yang, L.: An analysis of the IGBP global land-cover characterization process, Photogrammetric engineering and remote sensing, 65, 1021-1032, 1999.

Ma, Y., Hu, Z., Xie, Z., Ma, W., Wang, B., Chen, X., Li, M., Zhong, L., Sun, F., Gu, L., Han, C., Zhang, L., Liu, X., Ding, Z., Sun, G., Wang, S., Wang, Y., and Wang, Z.: A long-term (2005–2016) dataset of hourly integrated land–atmosphere interaction observations on the Tibetan Plateau, Earth System Science Data, 12, 2937-2957, 10.5194/essd-12-2937-2020, 2020.

Majozi, N. P., Mannaerts, C. M., Ramoelo, A., Mathieu, R., Nickless, A., and Verhoef, W.: Analysing surface energy balance closure and partitioning over a semi-arid savanna FLUXNET site in Skukuza, Kruger National Park, South Africa, Hydrology and Earth System Sciences, 21, 3401-3415, 10.5194/hess-21-3401-2017, 2017.

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., and Verhoest, N. E. C.: GLEAM v3: satellite-based land evaporation and root-zone soil moisture, Geoscientific Model Development, 10, 1903-1925, 10.5194/gmd-10-1903-2017, 2017.

McColl, K. A., Roy, A., Derksen, C., Konings, A. G., Alemohammed, S. H., and Entekhabi, D.: Triple collocation for binary and categorical variables: Application to validating landscape freeze/thaw retrievals, 176, 31-42, 10.1016/j.rse.2016.01.010, 2016.

Medlyn, B. E., De Kauwe, M. G., Lin, Y. S., Knauer, J., Duursma, R. A., Williams, C. A., Arneth, A., Clement, R., Isaac, P., Limousin, J. M., Linderson, M. L., Meir, P., Martin-StPaul, N., and Wingate, L.: How do leaf and ecosystem measures of water-use efficiency compare?, New Phytol, 216, 758-770, 10.1111/nph.14626, 2017.

Miralles, D. G., Gentine, P., Seneviratne, S. I., and Teuling, A. J.: Land–atmospheric feedbacks during droughts and heatwaves: state of the science and current challenges, Annals of the New York Academy of Sciences, 1436, 19, 2019.

Miralles, D. G., Holmes, T. R. H., De Jeu, R. A. M., Gash, J. H., Meesters, A. G. C. A., and Dolman, A. J.: Global land-surface evaporation estimated from satellite-

https://doi.org/10.5194/essd-2021-456
Preprint. Discussion started: 1 March 2022

based observations, Hydrology and Earth System Sciences, 15, 453-469, 10.5194/hess-15-453-2011, 2011.

Pan, S., Pan, N., Tian, H., Friedlingstein, P., Sitch, S., Shi, H., Arora, V. K., Haverd, V., Jain, A. K., Kato, E., Lienert, S., Lombardozzi, D., Nabel, J. E. M. S., Ottlé, C., Poulter, B., Zaehle, S., and Running, S. W.: Evaluation of global terrestrial evapotranspiration using state-of-the-art approaches in remote sensing, machine learning and land surface modeling, Hydrology and Earth System Sciences, 24, 1485-1509, 10.5194/hess-24-1485-2020, 2020.

Pastorello, G., Trotta, C., Canfora, E., Chu, H., Christianson, D., Cheah, Y.-W., Poindexter, C., Chen, J., Elbashandy, A., and Humphrey, M.: The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data, Scientific data, 7, 1-27, 2020.

Priestley, C. H. B. and TAYLOR, R. J.: On the assessment of surface heat flux and evaporation using large-scale parameters, Monthly weather review, 100, 81-92, 1972.

Restrepo-Coupe, N., Albert, L. P., Longo, M., Baker, I., Levine, N. M., Mercado, L. M., da Araujo, A. C., Christoffersen, B. O., Costa, M. H., Fitzjarrald, D. R., Galbraith, D., Imbuzeiro, H., Malhi, Y., von Randow, C., Zeng, X., Moorcroft, P., and Saleska, S. R.: Understanding water and energy fluxes in the Amazonia: Lessons from an observation-model intercomparison, Glob Chang Biol, 27, 1802-1819, 10.1111/gcb.15555, 2021.

Ribal, A. and Young, I. R.: Global Calibration and Error Estimation of Altimeter, Scatterometer, and Radiometer Wind Speed Using Triple Collocation, Remote Sensing, 12, 10.3390/rs12121997, 2020.

Rodell, M., Houser, P., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C.-J., Arsenault, K., Cosgrove, B., Radakovich, J., and Bosilovich, M.: The global land data assimilation system, Bulletin of the American Meteorological Society, 85, 381-394, 2004.

Sharma, K., Irmak, S., and Kukal, M. S.: Propagation of soil moisture sensing uncertainty into estimation of total soil water, evapotranspiration and irrigation decision-making, Agricultural Water Management, 243, 106454, 2021.

Stoffelen, A.: Toward the true near-surface wind speed: Error modeling and calibration using triple collocation, Journal of Geophysical Research: Oceans, 103, 7755-7766, 10.1029/97jc03180, 1998.

Su, C.-H., Ryu, D., Crow, W. T., and Western, A. W.: Beyond triple collocation: Applications to soil moisture monitoring, Journal of Geophysical Research: Atmospheres, 119, 6419-6439, 10.1002/2013jd021043, 2014.

Sun, J., McColl, K. A., Wang, Y., Rigden, A. J., Lu, H., Yang, K., Li, Y., and Santanello, J. A.: Global evaluation of terrestrial near-surface air temperature and specific humidity retrievals from the Atmospheric Infrared Sounder (AIRS), Remote Sensing of Environment, 252, 10.1016/j.rse.2020.112146, 2021.

Tang, G., Behrangi, A., Long, D., Li, C., and Hong, Y.: Accounting for spatiotemporal

https://doi.org/10.5194/essd-2021-456
Preprint. Discussion started: 1 March 2022

errors of gauges: A critical step to evaluate gridded precipitation products, Journal of Hydrology, 559, 294-306, 10.1016/j.jhydrol.2018.02.057, 2018.

Taylor, K. E.: Summarizing multiple aspects of model performance in a single diagram, Journal of Geophysical Research: Atmospheres, 106, 7183-7192, 2001.

Towner, J., Cloke, H. L., Zsoter, E., Flamig, Z., Hoch, J. M., Bazo, J., Coughlan de Perez, E., and Stephens, E. M.: Assessing the performance of global hydrological models for capturing peak river flows in the Amazon basin, Hydrology and Earth System Sciences, 23, 3057-3080, 10.5194/hess-23-3057-2019, 2019.

Twine, T. E., Kustas, W., Norman, J., Cook, D., Houser, P., Meyers, T., Prueger, J., Starks, P., and Wesely, M.: Correcting eddy-covariance flux underestimates over a grassland, Agricultural and forest meteorology, 103, 279-300, 2000.

Vogelzang, J. and Stoffelen, A.: Quadruple Collocation Analysis of In-Situ, Scatterometer, and NWP Winds, Journal of Geophysical Research: Oceans, 126, 10.1029/2021jc017189, 2021.

Wang, F., Liang, W., Fu, B., Jin, Z., Yan, J., Zhang, W., Fu, S., and Yan, N.: Changes of cropland evapotranspiration and its driving factors on the loess plateau of China, Sci Total Environ, 728, 138582, 10.1016/j.scitotenv.2020.138582, 2020.

Yilmaz, M. T. and Crow, W. T.: Evaluation of Assumptions in Soil Moisture Triple Collocation Analysis, Journal of Hydrometeorology, 15, 1293-1302, 10.1175/jhm-d-13-0158.1, 2014.

Yin, L., Tao, F., Chen, Y., Liu, F., and Hu, J.: Improving terrestrial evapotranspiration estimation across China during 2000–2018 with machine learning methods, Journal of Hydrology, 600, 10.1016/j.jhydrol.2021.126538, 2021.

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

Zhang, C., Abbaszadeh, P., Xu, L., Moradkhani, H., Duan, Q., and Gong, W.: A Combined Optimization-Assimilation Framework to Enhance the Predictive Skill of Community Land Model, Water Resources Research, 57, 10.1029/2021wr029879, 2021a.

Zhang, K., Kimball, J. S., Kim, Y., and McDonald, K. C.: Changing freeze-thaw seasons in northern high latitudes and associated influences on evapotranspiration, Hydrological Processes, 25, 4142-4151, 10.1002/hyp.8350, 2011.

Zhang, L., Marshall, M., Nelson, A., and Vrieling, A.: A global assessment of PT-JPL soil evaporation in agroecosystems with optical, thermal, and microwave satellite data, Agricultural and Forest Meteorology, 306, 10.1016/j.agrformet.2021.108455, 2021b.

Zhang, Y., Leuning, R., Hutley, L. B., Beringer, J., McHugh, I., and Walker, J. P.: Using long-term water balances to parameterize surface conductances and calculate evaporation at 0.05 spatial resolution, Water Resources Research, 46,

https://doi.org/10.5194/essd-2021-456
Preprint. Discussion started: 1 March 2022

2010.

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

Zhu, G., Li, X., Zhang, K., Ding, Z., Han, T., Ma, J., Huang, C., He, J., and Ma, T.: Multi-model ensemble prediction of terrestrial evapotranspiration across north China using Bayesian model averaging, Hydrological Processes, 30, 2861-2879, 10.1002/hyp.10832, 2016.