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

Changming Li[1], Ziwei Liu[1], Wencong Yang[1], Zhuoyi Tu[1], Juntai Han[1], Sien Li[2],

Hanbo 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)

13   **Abstract**

14   Land evapotranspiration (ET) plays a crucial role in Earth's water-carbon cycle, and

15   accurately estimating global land ET is vital for advancing our understanding of land-

16   atmosphere interactions. Despite the development of numerous ET products in recent

17   decades, widely used products still possess inherent uncertainties arising from using

18   different forcing inputs and imperfect model parameterizations. Furthermore, the lack

19   of sufficient global in-situ observations makes direct evaluation of ET products

20   impractical, impeding their utilization and assimilation. Therefore, establishing a

21   reliable global benchmark dataset and exploring evaluation methodologies for ET

22   products is paramount. This study aims to address these challenges by (1) proposing a

23   collocation-based method that considers non-zero error cross-correlation for merging

24   multi-source data and (2) employing this merging method to generate a long-term

25   daily global ET product at resolutions of 0.1° (2000-2020) and 0.25° (1980-2022),

26   incorporating inputs from ERA5L, FluxCom, PMLv2, GLDAS, and GLEAM. The

27   resulting product is the Collocation-Analyzed Multi-source Ensembled Land

28   Evapotranspiration Data (CAMELE). CAMELE exhibits promising performance

29   across various vegetation coverage types, as validated against in-situ observations.

30   The evaluation process yielded Pearson correlation coefficients (R) of 0.63 and 0.65,

31   root-mean-square-errors (RMSE) of 0.81 and 0.73 mm/d, unbiased root-mean-square-

32   errors (ubRMSE) of 1.20 and 1.04 mm/d, mean absolute errors (MAE) of 0.81 and

33   0.73 mm/d, and Kling-Gupta efficiency (KGE) of 0.60 and 0.65 on average over

34   resolutions of 0.1° and 0.25°, respectively. In addition, comparisons indicate that

35   CAMELE can effectively characterize the multi-year linear trend, mean average, and

36   extreme values of ET. However, it exhibits a tendency to overestimate seasonality. In

37   summary, we propose a reliable set of ET data that can aid in understanding the

38   variations in the water cycle and has the potential to serve as a benchmark for various

39   applications.

1. Introduction

Land evapotranspiration (ET) plays a critical role in the global water and energy cycles, encompassing various processes such as soil evaporation, vegetation transpiration, canopy interception, and surface water evaporation (Zhang et al., 2019; Zhao et al., 2022; Lian et al., 2018). Accurately estimating global land evapotranspiration is vital for understanding the hydrological cycle and land-atmosphere interactions, as it serves as an intermediary variable connecting soil moisture, air temperature and humidity (Miralles et al., 2019; Gentine et al., 2019). Therefore, providing a reliable ET dataset as a benchmark for further research is crucial.

In recent decades, numerous studies have focused on estimating global land evapotranspiration, resulting in many datasets (Yang et al., 2023). However, discrepancies often arise among these simulations due to algorithm and principle variations (Restrepo-Coupe et al., 2021; Han and Tian, 2020). Additionally, evaluating ET products is challenging due to the limited availability of global-scale observations, which hampers their direct use (Pan et al., 2020; Baker et al., 2021).

The fusion of multi-source data is a suitable option to address these uncertainties. Recent studies have explored several approaches to integrate multiple ET products, including Simple Average (SA) (Ershadi et al., 2014), Bayesian Model Average (BMA) (Hao et al., 2019; Ma et al., 2020; Zhu et al., 2016), Reliability Ensemble Average (REA) (Lu et al., 2021), Empirical Orthogonal Functions (EOF) (Feng et al., 2016) and machine-learning-based methods (Chen et al., 2020; Yin et al., 2021). However, the primary challenge lies in calculating reliable input weights based on a selected "truth" (Koster et al., 2021), which can involve averaging or incorporating other relevant geographical information as a benchmark.

Recently, collocation methods have emerged as promising techniques for estimating random error variances and data-truth correlations in collocated inputs (Stoffelen, 1998; Li et al., 2022, 2023c; Park et al., 2023). These methods consider the errors

associated with collocated datasets as an accurate representation of uncertainty without assuming the absence of errors in any datasets. It is important to note that while collocation methods, such as the triple collocation (TC) and the extended double instrumental variable technique (EIVD), can estimate the variance (or covariance) of random errors, they cannot evaluate the bias of the products. One primary advantage of collocation analysis is that it does not require a high-quality reference dataset (Su et al., 2014; Wu et al., 2021). However, a crucial prerequisite for applying collocation methods is the availability of many spatially and temporally corresponding datasets. For instance, the classic TC method requires a trio of independent datasets. Su et al. (2014) used the instrumental regression method and considered lag-1 time series as the third input, proposing the single instrumental variable algorithm (IVS). Dong et al. (2019) introduced the lag-1 time series from both inputs, proposing the double instrumental variable algorithm (IVD) for a more robust solution. Gruber et al. (2016a) extended the original algorithm to incorporate more datasets, partially addressing the independence assumption to calculate a portion of error cross-correlation (ECC) by using the extended collocation (EC) method. Dong et al.(2020a) further proposed the EIVD method, enabling ECC estimation using three datasets. Collocation methods have found widespread application in the evaluation of geophysical variable estimates, including soil moisture (Deng et al., 2023; Ming et al., 2022), precipitation (Dong et al., 2022; Li et al., 2018), ocean wind speed (Vogelzang et al., 2022; Ribal and Young, 2020), leaf area index (Jiang et al., 2017), total water storage (Yin and Park, 2021) sea ice thickness and surface salinity (Hoareau et al., 2018), and near-surface air temperature (Sun et al., 2021).

Recently, many studies have utilized collocation approaches to evaluate evapotranspiration products, with the TC method to assess uncertainties. For example, Barraza Bernadas et al. (2018) considered the uncertainties of ET from the Breathing Earth System Simulator, BESS (Jiang et al., 2020; Jiang and Ryu, 2016), Moderate Resolution Imaging Spectroradiometer, MOD16 (Mu et al., 2011), and a hybrid

96   model; Khan et al. (2018) utilized extended triple collocation (ETC) (McColl et al.,

97   2014) to investigate the reliability of ET from MOD16, The Global Land Data

98   Assimilation System (GLDAS) (Rodell et al., 2004) and the Global Land Evaporation

99   Amsterdam Model (GLEAM) (Martens et al., 2017) over East Asia; Li et al. (2022)

100 employed five collocation methods (e.g., IVS, IVD, TC, EIVD, and EC) to analyze

101 the uncertainties of ET from ERA5-Land (ERA5L) (Muñoz-Sabater et al., 2021),

102 GLEAM, GLDAS, FluxCom (Jung et al., 2019), and the Penman-Monteith-Leuning

103 Evapotranspiration V2 (PMLv2) (Zhang et al., 2019).

104 Moreover, error information derived from collocation analysis is valuable for merging

105 multi-source data. This was initially applied by Yilmaz et al. (2012) in the fusion of

106 multi-source soil moisture products and later improved by Gruber et al. (2017) and

107 further applied in the production of the European Space Agency Climate Change

108 Initiative (ESA CCI) global soil moisture product (Gruber et al., 2019). Dong et al.

109 (2020b) also adopted this approach to fusing multi-source precipitation products. In

110 the study of evapotranspiration, Li et al. (2023c) and Park et al.(2023) utilized a

111 weight calculation method that does not consider non-zero ECC and fused multiple

112 ET products in the Nordic and East Asia, respectively, achieving satisfactory fusion

113 results.

114 Although the above studies have demonstrated that collocation analysis can

115 effectively assess the random error variance of ET products and integrate error

116 information from multiple data sources, these studies have primarily overlooked a

117 critical aspect: non-zero ECC between ET products. Li et al. (2022) global ET product

118 evaluation research revealed clear non-zero ECC conditions between ERA5L,

119 GLEAM, PMLv2, and FluxCom. In TC analysis, non-zero ECC can result in

120 significant biases in TC-based results (Yilmaz and Crow, 2014). Furthermore, when

121 using TC-based error information for fusion, it is crucial to consider the information

122 related to ECC, as this can help improve the fusion accuracy (Dong et al., 2020b; Kim

123 et al., 2021b).

124   It is worth noting that non-zero ECC conditions pose unique challenges. Unlike other

125   violations of mathematical assumptions adopted by TC, they cannot be effectively

126   mitigated through rescaling or compensated for by equal magnitude adjustments

127   across inputs. Thus, the implications of non-zero ECC in the context of merging

128   strategies are a critical consideration often overlooked in previous research. This

129   oversight can lead to significant biases and inaccuracies. We aim to bridge this gap by

130   systematically accounting for non-zero ECC in weight calculation, contributing to a

131   more robust and accurate assessment.

132   In this study, we proposed a collocation-based data ensemble method, considering

133   non-zero ECC conditions, for merging multiple ET products to create the Collocation-

134   Analyzed Multi-source Ensembled Land Evapotranspiration data, abbreviated as

135   CAMELE. The second section of this paper presents the selected data information for

136   this study. In the third section, we explained the error calculation method for

137   collocation analysis and the weighted calculation method that considered ECC. The

138   fourth section analyzed the global errors of different ET products obtained through

139   these calculations and the distribution patterns of the corresponding weights. We

140   evaluated the accuracy of the fused products and compared them with existing

141   products using reference values from site measurements. In the fifth section, we

142   discussed the inherent errors in the methods, analyzed the ECC between the products,

143   and compared the differences between different fusion schemes. Finally, in the sixth

144   section, we summarized the results obtained from this research.

145   **2.   Datasets**

146   We selected five widely used ET products that spanned the period from 1980 to 2022.

147   When selecting these products, our aims are to ensure: (1) consistency in original

148   spatiotemporal resolution among the products: minimize potential downscaling

149   operations and avoid introducing additional errors; (2) having three or more products

150   within the same resolution or period: incorporate more information for effective

151 fusion; (3) products with extensive global observational sequences: gain basic

152 recognition from the community. While we acknowledge the existence of other

153 higher-precision products, their integration would require either downscaling or

154 upscaling other products, potentially introducing uncertainties. Therefore, we chose

155 the combination outlined in the manuscript. Despite its relatively lower resolution

156 compared to some products, it still contributes to our understanding of ET variations,

157 facilitating advantageous exploration. Furthermore, we incorporated in-situ

158 observations and Lu et al. (2021) 's global 0.25° daily-scale ET product derived using

159 Reliability Ensemble Averaging (denoted as REA) to compare our merged product

160 comprehensively. **Table 1** shows the spatial and temporal resolutions of the input

161 datasets.

162 **Table 1** Summary of evapotranspiration products involved.

Name	Schemes	Resolution		Period	Reference
ERA5-Land	H-TESSEL	0.1°	hourly	1950-present	(Muñoz-Sabater et al., 2021)
GLDAS-2	CLSM/Noah /LSM	0.25°	3-hourly daily	2.0: 1948-2014 2.1: 2000-present 2.2: 2003-present	(Li et al., 2019a; Rodell et al., 2004)
GLEAM-3.7	GLEAM model	0.25°	daily	3.7a: 1980-2022 3.7b: 2003-2022	(Martens et al., 2017)
PMLv2-v017	Penman-Monteith-Leuning	0.083°	8-day average	2000-2020	(Zhang et al., 2019)
FluxCom	Machine learning	0.083°	8-day average	2001-2015	(Jung et al., 2019)

163 **2.1. ERA5-Land**

164 The European Centre for Medium-Range Weather Forecasts (ECMWF) produces the

165 latest advanced ERA5L, a global hourly reanalysis dataset with a spatial resolution of

166   0.1°. It covers the period from January 1950 until approximately one week before the

167   present (Muñoz-Sabater et al., 2021). ERA5-Land is derived from the land component

168   of the ECMWF climate reanalysis, incorporating numerous improvements over

169   previously released versions. It is based on the Tiled ECMWF Scheme for Surface

170   Exchanges over Land incorporating land surface hydrology (H-TESSEL), utilizing

171   version CY45R1 of the ECMWF's Integrated Forecasting System (IFS). The dataset

172   benefits from atmospheric forcing data, which acts as an indirect constraint on the

173   model-based estimates (Hersbach et al., 2020). The dataset is available through the

174   Climate Change service of the Copernicus Center at http://cds.climate.copernicus.eu.

175   Evapotranspiration in ERA5L, defined as "total evaporation," represents the

176   accumulated amount of water that has evaporated from the Earth's surface, including a

177   simplified representation of transpiration from vegetation into the vapor in the air.

178   The soil water and energy balance are computed using standard soil discretization.

179   Readers could consult section 8.6.5 of the IFS documentation (ECMWF, 2014). The

180   original dataset is interpolated from (1801, 3600) to (1800, 3600) using kriging

181   interpolation and then upscaled from an hourly to a daily resolution, changing spatial

182   resolution from 0.1° to 0.25°.

183   **2.2. GLDAS**

184   The Global Land Data Assimilation System (GLDAS) product utilizes advanced data

185   assimilation methodologies, integrating model and observation datasets for land-

186   surface simulations (Rodell et al., 2004). GLDAS employs multiple land-surface

187   models (LSMs), namely Noah, Mosaic, Variable Infiltration Capacity (VIC), and the

188   Community Land Model (CLM). Together, these models generate global

189   evapotranspiration estimates at fine and coarse spatial resolutions (0.01° and 0.25°)

190   and temporal resolutions (3-hourly and monthly). The most recent iteration of

191   GLDAS, version 2, consists of three components: GLDAS-2.0, GLDAS-2.1, and

192   GLDAS-2.2. GLDAS-2.0 relies entirely on the Princeton meteorological forcing input

193 data, providing a consistent temporal series from 1948 to 2014 (Sheffield et al., 2006).

194 The GLDAS-2.1 simulation commences on January 1, 2000, utilizing the conditions

195 from the GLDAS-2.0 simulation. On the other hand, GLDAS-2.2 is simulated from

196 February 1, 2003, employing the conditions from GLDAS-2.0 and forcing with

197 meteorological analysis fields from the ECMWF Integrated Forecasting System (IFS).

198 Additionally, the GRACE satellite's total terrestrial water anomaly observation is

199 assimilated into the GLDAS-2.2 product (Li et al., 2019a).

200 This study aimed to cover the research period from 1980 to 2022. Non-zero ECC

201 between the transpiration estimates of GLDAS-2.2 and ERA5L has been reported in a

202 recent study (Li et al., 2023a). Considering the similarities in the calculation of ET

203 and transpiration of GLDAS and ERA5L, this report partially indicates a correlation.

204 Therefore, GLDAS-2.0 and GLDAS-2.1 were selected as inputs instead. The

205 "Evap_tavg" parameter representing evapotranspiration is derived from the original

206 products and aggregated to a daily scale. For more detailed information on the

207 GLDAS-2 models, please refer to NASA's Hydrology Data and Information Services

208 Center at http://disc.sci.gsfc.nasa.gov/hydrology.

209 Despite the same forcing between GLDAS-2.1 and GLDAS-2.2, significant

210 differences exist between the model results of different GLDAS versions (Qi et al.,

211 2020, 2018; Jiménez et al., 2011). The non-zero ECC will generally still be met

212 between different versions. Thus, we still need to analyze the non-zero ECC situations

213 between ERA5L and GLDAS-2.0 and 2.1, which will be assessed in the discussion

214 sections.

215 **2.3. GLEAM**

216 The version of the Global Land Evaporation Amsterdam Model 3.7 (GLEAM-3.7)

217 dataset (Martens et al., 2017; Miralles et al., 2011) at 0.25° is used. This version of

218 GLEAM provides daily estimations of actual evaporation, bare soil evaporation,

219 canopy interception, transpiration from vegetation, potential evaporation, and snow

220   sublimation. The third version of GLEAM contains a new DA scheme, an updated

221   water balance module, and evaporative stress functions. Two datasets that differ only

222   in forcing and temporal coverage are provided: GLEAMv3.7a-43-year period (1980

223   to 2022) based on satellite and reanalysis (ECMWF) data; GLEAMv3.7b-20-year

224   period (2003 to 2022) based on only satellite data. GLEAMv3.7a is used in this study.

225   The data are freely available on the GLEAM website (https://www.gleam.eu).

226   The cover-dependent potential evaporation rate (E_P) is calculated using the Priestley-

227   Taylor equation (Priestley and TAYLOR, 1972). Then a multiplicative stress factor is

228   used to convert E_P into actual transpiration and bare soil evaporation, which is the

229   function of microwave vegetation optimal depth (VOD) and root-zone soil moisture.

230   For detailed description, please refer to the paper by Martens et al., (2017). The

231   GLEAM data were validated at 43 FluxNet flux sites and have been proven to provide

232   reliable ET estimations (Majozi et al., 2017).

233   **2.4. PMLv2**

234   The Penman-Monteith-Leuning version 2 global evaporation model (PMLv2) has

235   been developed based on the Penman-Monteith-Leuning model (Zhang et al., 2019;

236   Leuning et al., 2009). Initially proposed by Leuning et al. (2008), the PML model

237   underwent further enhancements by Zhang et al. (2010). The PML version 1 (PMLv1)

238   incorporates a biophysical model that considers canopy physiological processes and

239   soil evaporation to estimate surface conductance accurately (G_s), which is the focus of

240   the PM-based method. This version was subsequently enhanced by incorporating a

241   canopy conductance (G_c) model that couples vegetation transpiration with gross

242   primary productivity, resulting in the development of PML version 2 (PMLv2) as

243   described by Gan et al. (2018). Zhang et al. (2019) applied the PMLv2 model globally.

244   The daily inputs for this model include leaf area index (LAI), broadband albedo, and

245   emissivity obtained from the Moderate Resolution Imaging Spectroradiometer

246   (MODIS), as well as temperature variables (daily maximum temperature-T_{max} , daily

247   minimum temperature-T_{min}, daily mean temperature-T_{avg}), instantaneous variables

248   (surface pressure-P_{surf}, atmosphere pressure-P_a, wind speed at 10-meter height-U,

249   specific humidity-q), and accumulated variables (precipitation-P_{rcp}, inward longwave

250   solar radiation- R_{ln} , inward shortwave solar radiation- R_s) from GLDAS-2.0.

251   Evaporation is divided into direct evaporation from bare soil (E_s), evaporation from

252   solid water sources (water bodies, snow, and ice) (ET_{water}), and vegetation

253   transpiration (E_c). To ensure its accuracy, the PMLv2-ET model was calibrated

254   against 8-daily eddy covariance data from 95 global flux towers representing ten

255   different land cover types.

256   In this study, we employ the latest version, v017. The data is freely available through

257   the   google   earth   engine   https://developers.google.com/earth-

258   engine/datasets/catalog/CAS_IGSNRR_PML_V2_v017.

259   **2.5.   FluxCom**

260   FluxCom is a machine-learning-based approach combining global land-atmosphere

261   energy flux data by combining remote sensing and meteorological data (Jung et al.,

262   2019). To achieve this, FluxCom utilizes various machine-learning regression tools,

263   including tree-based methods, regression splines, neural networks, and kernel

264   methods. The outputs of FluxCom are designed based on two complementary

265   strategies: (1) FluxCom-RS, which exclusively merges remote sensing data to

266   generate high spatial resolution flux data; and (2) FluxCom-RS+METEO, which

267   combines meteorological observations with remote sensing data at a daily temporal

268   resolution. The exclusive use of remote sensing data in the ensemble allows

269   producing gridded flux products at a spatial resolution of 500m, albeit with a

270   relatively low frequency of 8 days. It is important to note that the FluxCom-RS data

271   only covers the period after 2000 due to data availability.

272   In contrast, the merging of meteorological and remote sensing data extends the

273   coverage back to 1980 at the cost of a coarser spatial resolution of 0.5°. For more

274 detailed information about the FluxCom dataset, please refer to the FluxCom website

275 (http://FluxCom.org/). The data is freely available upon contacting the authors.

276 In this study, we utilized the FluxCom-RS 8-daily 0.0833° energy flux data and

277 converted the latent heat values to evaporation using ERA5L aggregated daily air

278 temperature. Furthermore, the original ET data were interpolated to a spatial

279 resolution of 0.1° using the MATLAB Gaussian process regression package.

280 **2.6.   Global in-situ observation: FluxNet**

281 The latest FluxNet2015 4.0 eddy-covariance data were used in our study (Pastorello et

282 al., 2020). Following the filtering process by Lin et al. (2018) and Li et al. (2019b),

283 firstly, only the measured and good-quality gap-filled data were used for quality

284 control. Secondly, we excluded days with rainfall and the subsequent day after rainy

285 events to mitigate the impact of canopy interception (Medlyn et al., 2017; Knauer et

286 al., 2018). Additionally, previous studies have indicated an energy imbalance problem

287 in FluxNet2015 data. Therefore, following the method proposed by Twine et al.

288 (2000), the measured ET data were corrected using the residual method based on

289 energy balance.

290 After data filtering and processing, 212 sites are selected as shown in **Figure 1**. The

291 selected sites are distributed globally, primarily in North America and Europe. The

292 International-Geosphere–Biosphere Program (IGBP) land cover classification system

293 (Loveland et al., 1999) was employed to distinguish the 13 Plant Functional Types

294 (PFTs) across sites. The IGBP classification was determined based on metadata from

295 the FluxNet official website, including evergreen needle leaf forests (ENF, 49 sites),

296 evergreen broadleaf forests (EBF, 15 sites), deciduous broadleaf forests (DBF, 26

297 sites), croplands (CRO, 20 sites), grasslands (GRA, 39 sites), savannas (SAV, 9 sites),

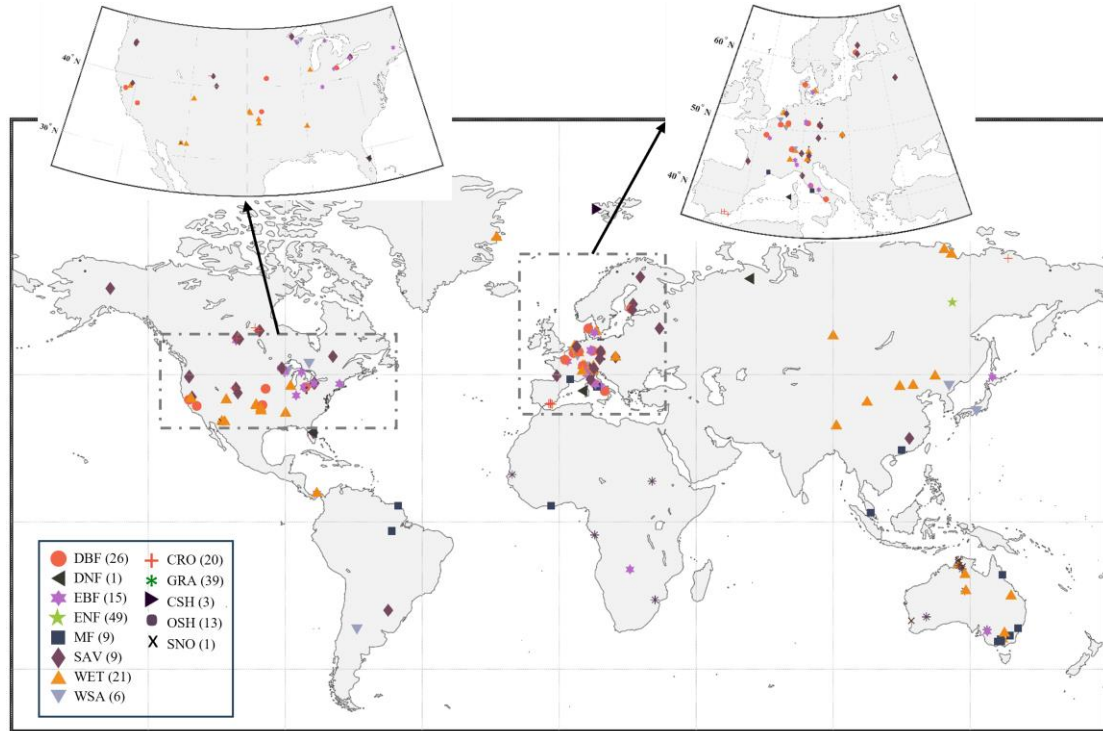
298 mixed forests (MF, 9 sites), closed shrublands (CSH, 3 sites), deciduous needle leaf

299 forests (DNF, 1 site), open shrublands (OSH, 13 sites), snow and ice (SNO, 1 site),

300 woody savannas (WSA, 6 sites) and permanent wetland (WET, 21 sites). Changes in

301   the IGBP classification during the study period are possible, but such information is

302   not publicly available. Interested parties can obtain relevant information by directly

303   contacting the site coordinators.

304

305   **Figure 1** Global distribution of selected FluxNet Sites.

306   **3. Method**

307   In this study, the fusion of products consisted of three steps: (1) the collocation

308   method (IVD and EIVD) was used to calculate the random error variance of the

309   selected input products, determine the regionally optimal products, and set an error

310   threshold; (2) aiming for minimum mean-square-error (MSE), the weights of different

311   products on each grid were calculated; (3) the products were fused according to the

312   weights to obtain a long sequence of evapotranspiration products. Since IVD and

313   EIVD were developed by combining instrumental variable regression and the

314   extended collocation system, a description of TC and EC algorithms was also

315   included.

3.1. Triple collocation analysis

Since its development in 1998, the implications and formulations of the triple collocation problem have been investigated in many studies. Here, we used difference notation for demonstration.

The commonly used error structure for triple collocation analysis (TCA) is:

$$i = \alpha_i + \beta_i \Theta + \varepsilon_i \tag{1}$$

where $i \in [X, Y, Z]$ are three spatially and temporally collocated data sets; Θ is the unknown true signal for relative geographical variable; α_i and β_i are additive and multiplicative bias factors against the true signal, respectively; ε_i is the additive zero-mean random error.

The above structure is also a typical instrumental variable (IV) regression. Thus, this provides another perspective to introduce more variables (>3) (Dong and Crow, 2017; Su et al., 2014) and polynomial models (Yilmaz and Crow, 2013; De Lannoy et al., 2007) to the standard TC. We recommend that the readers refer to Su et al. (2014) for a more detailed discussion on using the IV framework.

The basic assumptions adopted in TC are as follows: (i) Linearity between true signal and data sets, (ii) signal and error stationarity, (iii) independency between random error and true signal (error orthogonality), (iv) independence between random errors (zero ECC). Although many studies have indicated that some of these assumptions are often violated in practice (Li et al., 2018, 2022; Jia et al., 2022), the formulation based on these assumptions is still the most robust implementation (Gruber et al., 2016b). A discussion on these assumptions will be provided in the discussion section.

The data sets first need to be rescaled against an arbitrary reference (e.g., X). The others are scaled through a TC-based rescaling scheme:

$$Y^X = \beta_Y^X (Y - \overline{Y}) + \overline{X} \quad Z^X = \beta_Z^X (Z - \overline{Z}) + \overline{X} \tag{2}$$

The overbar denotes the mean value, and β_Y^X and β_Z^X are the scaling factors as:

$$\begin{cases} \beta_Y^X = \dfrac{\beta_X}{\beta_Y} = \dfrac{< \left(X - \overline{X}\right)\left(Z - \overline{Z}\right) >}{< \left(Y - \overline{Y}\right)\left(Z - \overline{Z}\right) >} = \dfrac{\sigma_{XZ}}{\sigma_{YZ}} \\[2mm] \beta_Z^X = \dfrac{\beta_X}{\beta_Z} = \dfrac{< \left(X - \overline{X}\right)\left(Y - \overline{Y}\right) >}{< \left(Z - \overline{Z}\right)\left(Y - \overline{Y}\right) >} = \dfrac{\sigma_{XY}}{\sigma_{ZY}} \end{cases} \quad (3)$$

340   where $< \cdot >$ is the average operator, σ_{ij} is the covariance of data sets i and j.

341   Subsequently, the error variances could be estimated by averaging the cross-

342   multiplied data set differences as follows:

$$\begin{cases} \sigma_{\varepsilon_X}^2 =< (X - Y^X)(X - Z^X) > \\[1mm] \sigma_{\varepsilon_Y^X}^2 = \beta_Y^{X\,2}\sigma_{\varepsilon_Y}^2 =< (Y^X - X)(Y^X - Z^X) > \\[1mm] \sigma_{\varepsilon_Z^X}^2 = \beta_Z^{X\,2}\sigma_{\varepsilon_Z}^2 =< (Z^X - X)(Z^Y - Y^X) > \end{cases} \quad (4)$$

343   Expanding the bracket and expressing the rescaling factors yields:

$$\begin{cases} \sigma_{\varepsilon_X}^2 = \sigma_X^2 - \dfrac{\sigma_{XY}\sigma_{XZ}}{\sigma_{YZ}} \\[3mm] \sigma_{\varepsilon_Y}^2 = \sigma_Y^2 - \dfrac{\sigma_{YX}\sigma_{YZ}}{\sigma_{XZ}} \\[3mm] \sigma_{\varepsilon_Z}^2 = \sigma_Z^2 - \dfrac{\sigma_{ZX}\sigma_{ZY}}{\sigma_{XY}} \end{cases} \quad (5)$$

344   When selecting various scaling references, it is essential to note that the absolute error

345   variances remain consistent. However, this choice can have an impact on the

346   estimation of data sensitivity to the actual signal ($\beta_i^2 \sigma_\Theta^2$), which serves as a crucial

347   indicator for comparing spatial error patterns. In order to address the reliance on a

348   specific scaling reference, Draper et al. (2013) introduced the fractional root-mean-

349   squared-error ($fMSE_i$). This measure is obtained by normalizing the unscaled error

350   variance with respect to the true signal variance:

$$fMSE_i = \frac{\sigma_{\varepsilon_i}^2}{\sigma_i^2} = \frac{\sigma_{\varepsilon_i}^2}{\beta_i^2 \sigma_\Theta^2 + \sigma_{\varepsilon_i}^2} = \frac{1}{1 + SNR_i} \quad (6)$$

351   where $SNR_i = \frac{\beta_i^2 \sigma_\Theta^2}{\sigma_{\varepsilon_i}^2} \in [0,1]$ is the normalized signal-to-noise ratio. $SNR = 0$

352   indicates a noise-free observation, while $SNR = 1$ corresponds that the variances of

353   estimates equal that of the true signal.

354   Following similar ideas, Mccoll et al. (2014) extended the framework to estimate the

355 data-truth correlation, known as the ETC:

$$R_i^2 = \frac{\beta_i^2 \sigma_\Theta^2}{\beta_i^2 \sigma_\Theta^2 + \sigma_{\varepsilon_i}^2} = \frac{SNR_i}{1 + SNR_i}$$ (7)

$$R_i^2 = 1 - fMSE_i$$

356 In comparison to the conventional coefficient of determination R_{ij}, which is

357 influenced by data noise and sensitivity. It is important to note that R_i^2 is merely based

358 on the data set i, whereas R_{ij} is influenced by both data set i and reference j. In other

359 words, R_i^2 incorporates the dependency on the chosen reference. Thus, TC-derived

360 $fMSE_i$ and R_i^2 serve as superior indicators for assessing the actual quality of data, as

361 discussed by Kim et al. (2021b) and Gruber et al. (2020).

362 **3.2. Double instrumental variable technique**

363 The assumed error structure in TC is also a typical instrumental variable (IV)

364 regression. In practical usage, finding three completely independent sets of products is

365 usually tricky. Su et al. (2014) effectively improve the applicability of the TC method

366 by using the lag-1 time series (e.g., $X_{t-1} = \alpha_X + \beta_X \Theta_{t-1} + \varepsilon_{X,t-1}$) from one of the

367 two sets of data as the third input for TC. In this way, we only need two independent

368 products for input.

369 Such process includes another assumption that all data sets contain serially white

370 errors (i.e., $< \varepsilon_{i,t} \varepsilon_{i,t-1} > = 0$, zero auto-correlation). Building upon this, Dong et al.

371 (2019) utilizes the lag-1 time series from both data sets as inputs and propose the

372 more stable IVD method.

373 For a double input $[X, Y$ with $\sigma_{\varepsilon_X \varepsilon_Y} = 0]$, the linear error model and related lag-1

374 time series can be expressed as:

$$\begin{cases} X = \alpha_X + \beta_X \Theta + \varepsilon_X & I = \alpha_X + \beta_X \Theta_{t-1} + \varepsilon_{X_{t-1}} \\ Y = \alpha_Y + \beta_Y \Theta + \varepsilon_Y & J = \alpha_Y + \beta_Y \Theta_{t-1} + \varepsilon_{Y_{t-1}} \end{cases}$$ (8)

375 where I and J are the lag-1 time series of X and Y, respectively.

376 Assuming product errors are mutually independent and orthogonal to the truth, the

377 covariance between the products is expressed as:

$$\begin{cases} \sigma_X^2 = \beta_X^2 \sigma_\Theta^2 + \sigma_{\varepsilon_X}^2 & \sigma_Y^2 = \beta_Y^2 \sigma_\Theta^2 + \sigma_{\varepsilon_Y}^2 \\ \qquad\qquad \sigma_{XY} = \beta_X \beta_Y \sigma_\Theta^2 \\ \quad \sigma_{IX} = \beta_X^2 L_{\Theta\Theta} & \sigma_{JY} = \beta_Y^2 L_{\Theta\Theta} \end{cases} \tag{9}$$

378 where $L_{ii} = <i_t i_{t-1}>$ is the auto-covariance. Therefore, the IVD-estimated dynamic

379 range ratio scaling factors yields:

$$s_{ivd} \equiv \frac{\beta_X}{\beta_Y} = \sqrt{\frac{\sigma_{IX}}{\sigma_{JY}}} \tag{10}$$

380 Hence, the random error variances of X and Y can be solved as:

$$\begin{cases} \sigma_{\varepsilon_X}^2 = \sigma_X^2 - \sigma_{XY} * s_{ivd} \\ \sigma_{\varepsilon_Y}^2 = \sigma_Y^2 - \dfrac{\sigma_{XY}}{s_{ivd}} \end{cases} \tag{11}$$

381 **3.3. Extended double instrumental variable technique**

382 Furthermore, by adopting the designed matrix in the EC method (Gruber et al., 2016a),

383 Dong et al. (2020a) present the EIVD method to estimate the error variance matrix

384 with only two independent data sets.

385 For a triplet input $\left[i, j, k \text{ with } \sigma_{\varepsilon_i \varepsilon_j} \neq 0\right]$. The dynamic range ratio scaling factors can

386 be estimated as follows:

$$s_{ij} \equiv \frac{\beta_i}{\beta_j} = \sqrt{\frac{L_{ii}}{L_{jj}}} \tag{12}$$

387 where $L_{ii} = <i_t i_{t-1}>$ is the auto-covariance of inputs. Subsequently, the sensitivity

388 and absolute error variance of the data set follow:

$$\beta_j^2 \sigma_\Theta^2 = \sigma_{ij} \sqrt{\frac{L_{ii}}{L_{jj}}} \qquad \sigma_{\varepsilon_j}^2 = \sigma_{ij} \sqrt{\frac{L_{ii}}{L_{jj}}} - \sigma_i^2 \tag{13}$$

389 The cross-multiplied factors can be estimated by:

$$\beta_i \beta_j \sigma_\Theta^2 = \sigma_{ik} \sqrt{\frac{L_{jj}}{L_{kk}}} = \sigma_{jk} \sqrt{\frac{L_{ii}}{L_{kk}}} \qquad \sigma_{\varepsilon_i \varepsilon_j} = \sigma_{ij} - \beta_i \beta_j \sigma_\Theta^2 \tag{14}$$

390 Hence, for a triplet with the input of $[X, Y, Z$ with $\sigma_{\varepsilon_X \varepsilon_Y} \neq 0]$: the matrix notation of

391 the above system with $\mathbf{y} = \mathbf{Ax}$ is given as:

$$
\mathbf{y} = \begin{pmatrix} \sigma_X^2 \\ \sigma_Y^2 \\ \sigma_Z^2 \\ \sigma_{XY} \\ \sigma_{XZ}\sqrt{\dfrac{L_{XX}}{L_{ZZ}}} \\ \sigma_{YZ}\sqrt{\dfrac{L_{YY}}{L_{ZZ}}} \\ \sigma_{ZX}\sqrt{\dfrac{L_{ZZ}}{L_{XX}}} \\ \sigma_{ZY}\sqrt{\dfrac{L_{ZZ}}{L_{YY}}} \\ \sigma_{XZ}\sqrt{\dfrac{L_{YY}}{L_{ZZ}}} \\ \sigma_{YZ}\sqrt{\dfrac{L_{XX}}{L_{ZZ}}} \end{pmatrix}_{10\times1} \quad \mathbf{A} = \begin{pmatrix} \mathbf{I_{4x4}} & \mathbf{I_{4x4}} \\ \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}_{6x4} & \mathbf{0_{6x4}} \end{pmatrix}_{10x8} \quad \mathbf{x} = \begin{pmatrix} \beta_X^2\sigma_\Theta^2 \\ \beta_Y^2\sigma_\Theta^2 \\ \beta_Z^2\sigma_\Theta^2 \\ \beta_X\beta_Y\sigma_\Theta^2 \\ \sigma_{\varepsilon_X}^2 \\ \sigma_{\varepsilon_Y}^2 \\ \sigma_{\varepsilon_Z}^2 \\ \sigma_{\varepsilon_X\varepsilon_Y} \end{pmatrix}_{8x1} \tag{15}
$$

392 Likewise, the least-squared solution for unknown $\mathbf{x}$ is then solved by:

$$
\mathbf{x} = (\mathbf{A^T A})^{-1}\mathbf{A^T y} \tag{16}
$$

3.4. Weight Estimation

394 Our objective is to predict an uncertain variable, such as evapotranspiration (ET) over

395 time at a specific location, by utilizing parent products that may contain random errors.

396 The underlying concept of weighted averaging is to extract independent information

397 from multiple data sources to enhance prediction accuracy by mitigating the effects of

398 random errors. The effectiveness of this approach relies on the independence of the

399 individual data sources under consideration. Weighted averaging has found

400 applications in various fields following the influential work of Bates and Granger

401 (1969), who proposed the optimal combination of forecasts based on a minimum

402 MSE criterion. In this context, the term "optimal" refers to minimizing the variance of

403 residual random errors in the least squares sense. Mathematically, this weighted

404 average can be expressed as follows:

$$\overline{x} = \vec{\mathbf{W}}^{\mathrm{T}}\vec{\mathbf{X}} = \sum_{i=1}^{N} \omega_i x_i \tag{17}$$

405 where $\overline{x}$ is the merged estimate; $\vec{\mathbf{X}} = [x_1, \dots, x_n]^{\mathrm{T}}$ contains the temporally collocated

406 estimates from N different parent products, which are merged with relative zero-mean

407 random error $\vec{\boldsymbol{e}} = [\varepsilon_1, \dots, \varepsilon_n]^{\mathrm{T}}$; and $\vec{\mathbf{W}} = [\omega_1, \dots, \omega_n]^{\mathrm{T}}$ contains the weights assigned

408 to these estimates, where $\omega_i \in [0,1]$ and $\sum \omega_i = 1$ ensuring an unbiased prediction.

409 The averaging weights can be expressed as the solution to the problem:

$$\min f(\vec{\mathbf{W}}) = \mathbb{E}(\vec{\boldsymbol{e}}^{\mathrm{T}}\vec{\mathbf{W}})^2 \tag{18}$$

410 where $\mathbb{E}()$ is the operator for mathematical expectation, the solution of this problem is

411 determined by the individual random error characteristics of the input data sets and

412 can be derived from their covariance matrix (Bates and Granger, 1969; Gruber et al.,

413 2017; Kim et al., 2021b):

$$\vec{\mathbf{W}} = \left(\vec{\mathbf{I}}^{\mathrm{T}}\mathbb{E}(\vec{\boldsymbol{e}}\vec{\boldsymbol{e}}^{\mathrm{T}})^{-1}\vec{\mathbf{I}}\right)^{-1}\mathbb{E}(\vec{\boldsymbol{e}}\vec{\boldsymbol{e}}^{\mathrm{T}})^{-1}\vec{\mathbf{I}}$$
$$\sigma_{\varepsilon_{\overline{x}}}^2 = \left(\vec{\mathbf{I}}^{\mathrm{T}}\mathbb{E}(\vec{\boldsymbol{e}}\vec{\boldsymbol{e}}^{\mathrm{T}})^{-1}\vec{\mathbf{I}}\right)^{-1} \tag{19}$$

414 where $\mathbb{E}(\vec{\boldsymbol{e}}\vec{\boldsymbol{e}}^{\mathrm{T}})$ is the $N \times N$ error covariance matrix that holds the random error

415 variance $\sigma_{\varepsilon_i}^2$ of the parent products in the diagonals and relative error covariances $\sigma_{\varepsilon_i \varepsilon_j}$

416 in the off-diagonals; $\vec{\mathbf{I}} = [1, \dots, 1]^{\mathrm{T}}$ is an ones-vector of length N; and $\sigma_{\varepsilon_{\overline{x}}}^2$ is the

417 resulting random error variances of the merged estimate.

418 When only two groups of products are used as input ($N = 2$), it is generally assumed

419 that the errors between them are independent. In this case, the weights are as follows:

$$\mathbb{E}(\vec{\boldsymbol{e}}\vec{\boldsymbol{e}}^{\mathrm{T}}) = \begin{bmatrix} \sigma_{\varepsilon_1}^2 & 0 \\ 0 & \sigma_{\varepsilon_2}^2 \end{bmatrix}$$
$$\omega_1 = \frac{\sigma_{\varepsilon_2}^2}{\sigma_{\varepsilon_1}^2 + \sigma_{\varepsilon_2}^2} \qquad \omega_1 = \frac{\sigma_{\varepsilon_1}^2}{\sigma_{\varepsilon_1}^2 + \sigma_{\varepsilon_2}^2} \tag{20}$$

420 In most cases, we can identify three sets of products as inputs ($N = 3$). In this

421 scenario, we consider the possibility of error homogeneity, assuming a non-zero ECC

422   exists between inputs 1 and 2. In this case, the error matrix can be represented as:

$$\mathbb{E}(\vec{\boldsymbol{e}}\vec{\boldsymbol{e}}^{\mathrm{T}}) = \begin{bmatrix} \sigma_{\varepsilon_1}^2 & \sigma_{\varepsilon_1\varepsilon_2} & 0 \\ \sigma_{\varepsilon_1\varepsilon_2} & \sigma_{\varepsilon_2}^2 & 0 \\ 0 & 0 & \sigma_{\varepsilon_3}^2 \end{bmatrix} \tag{21}$$

423   The weights can then be written as:

$$\vec{\mathbf{W}} = \begin{cases} \dfrac{\sigma_{\varepsilon_2}^2 - \sigma_{\varepsilon_1\varepsilon_2}}{\left(\sigma_{\varepsilon_1}^2 \sigma_{\varepsilon_2}^2 - \sigma_{\varepsilon_1\varepsilon_2}^2\right) * \mathbb{Z}} \\ \dfrac{\sigma_{\varepsilon_1}^2 - \sigma_{\varepsilon_1\varepsilon_2}}{\left(\sigma_{\varepsilon_1}^2 \sigma_{\varepsilon_2}^2 - \sigma_{\varepsilon_1\varepsilon_2}^2\right) * \mathbb{Z}} \\ \dfrac{1}{\sigma_{\varepsilon_3}^2 * \mathbb{Z}} \end{cases} \tag{22}$$

$$\mathbb{Z} = \frac{\sigma_{\varepsilon_1}^2 + \sigma_{\varepsilon_2}^2 - 2\sigma_{\varepsilon_1\varepsilon_2}}{\sigma_{\varepsilon_1}^2 \sigma_{\varepsilon_2}^2 - \sigma_{\varepsilon_1\varepsilon_2}^2} + \frac{1}{\sigma_{\varepsilon_3}^2}$$

424   It is essential to acknowledge that before applying these weights for merging the data

425   sets, it is necessary to address any existing systematic differences. Typically, this is

426   achieved by rescaling the data sets to a standardized data space. Consequently, the

427   weights can be derived from the rescaled data sets using Eq (2)-(3) and converge

428   accordingly. This procedure ensures the accuracy and reliability of the merged data

429   sets for further analysis.

430   If ECC is not considered (i.e., setting $\sigma_{\varepsilon_1\varepsilon_2} = 0$), Eq (22) represents the weight

431   calculation method commonly used in most TC fusion studies. In contrast to the

432   fusion studies mentioned above for evapotranspiration products, for the first time, the

433   consideration of non-zero ECC is incorporated into the fusion process and integrated

434   into the weight calculation. Yilmaz and Crow (2014) have demonstrated that TC

435   underestimates error variances when the zero ECC assumption is violated. Li et al.

436   (2022), in their evaluation study of global ET products using the collocation method,

437   also indicated the existence of error homogeneity issues between commonly used ET

438   products (such as ERA5L and GLEAM), necessitating the consideration of the

439   influence of non-zero ECC. The merging technique employed in this study provides a

440   more explicit characterization of product errors and facilitates the derivation of more

441   reliable weight coefficients, thereby achieving promising fusion outcomes.

442   The differences in results are evaluated at the site scale by contrasting the scenarios

443   without considering non-zero ECC and directly using simple averages to compare and

444   validate the advantages of the weight calculation method used in our study.

445   **3.5.   Merging combination**

446   In this study, we employ five commonly used global land surface ET products as

447   described in the datasets section. PMLv2 and FluxCom-RS have an original resolution

448   of 0.083° and an 8-day average. In this research, they are interpolated to 0.1°

449   resolution, and the values for each data period of 8 days are kept consistent. For

450   example, the values for March 5 to March 12, 2000, are the same. ET values often

451   exhibit variability over an 8-day period, making the use of an 8-day average to

452   represent temporal dynamics potentially introducing further uncertainties. This

453   operation is performed to ensure adequate data for the collocation analysis (Kim et al.,

454   2021a). We openly acknowledge the possible sources of error and express our

455   commitment to addressing and improving them in future work.

456   As mentioned in the methodology section, it is vital to consider the issue of random

457   error homogeneity among different products before applying the collocation method.

458   Although EC or EIVD methods can be used to calculate the ECC between specific

459   pairs of products, it is necessary to determine which pairs of products have non-zero

460   ECC conditions. In previous research, Li et al.(2022) employed five collocation

461   methods (IVS/IVD/TC/EIVD/EC) to analyze the performance of five sets of ET

462   products (ERA5L/ PMLv2/FluxCom/GLDAS2/GLEAMv3) at the global scale, and

463   applied EC and EIVD methods to calculate the ECC between different products. The

464   results indicated a relatively significant error homogeneity between PMLv2 and

465   FluxCom at a resolution of 0.1° (with a global average ECC of approximately 0.3).

466   The error homogeneity could be attributed to both products utilizing GLDAS

467   meteorological data as input, despite their different methods for ET estimation. At a

468   resolution of 0.25°, ERA5L and GLEAM exhibited a more apparent error correlation

469   (with a global average ECC of approximately 0.4). Considering the long temporal

470   data of GLEAMv3 version a, ECMWF meteorological data was chosen as the driving

471   force, making the error correlation between the two products predictable.

472   Therefore, this study assumes that non-zero ECC situations occur between PMLv2-

473   FluxCom and ERA5L-GLEAM. We also calculated the possible ECC situations

474   among other products, presented in the discussion section and the appendix. Based on

475   the analysis, our assumed non-zero ECC situations align reasonably well with the

476   actual circumstances.

477   In addition, previous research suggests that the IVD method outperforms the IVS

478   method in scenarios involving two sets of inputs, while the EIVD method is

479   considered more reliable than the TC method in situations with three sets of inputs (Li

480   et al., 2022; Kim et al., 2021a). Therefore, in this study, the IVD and EIVD methods

481   are selected for computation based on different combinations of inputs. **Table 2**

482   presents the data and methods used during corresponding periods. When only two sets

483   of products are available, we employ the IVD method for fusion and calculate weights

484   using Eq. (20). When three sets of products are available, we utilize the EIVD method

485   for fusion and calculate weights using Eq. (22).

486   **Table 2** Combination of inputs and accessible methods

Scenario 1 (0.1°)		
Period	**Selected Inputs**	**Method**
(2000.02.26-2000.12.31)	ERA5L/ PMLv2	IVD
(2001.01.01-2015.12.27)	ERA5L/ FluxCom/ PMLv2	EIVD
(2015.12.28-2020.12.26)	ERA5L/ PMLv2	IVD
Scenario 2 (0.25°)		
Period	**Selected Inputs**	**Method**
(1980.01.01-1999.12.31)	ERA5L/ GLDAS20/ GLEAMv3.7a	EIVD
(2000.01.01-2022.12.31)	ERA5L/ GLDAS21/ GLEAMv3.7a	

487   It should be noted that the same product can have different versions. In this study,

488 appropriate versions are selected based on the following principles: (1) Selection

489 based on the corresponding data coverage duration and ensuring more products to

490 gain more information; (2) Choosing the latest version while considering the

491 assumption of non-zero ECC conditions; (3) Making efforts to select the exact

492 product versions for different periods, to avoid uncertainties caused by version

493 changes. We selected a subset of sites to compare the fusion results using different

494 versions, and the corresponding details will be presented in the discussion section.

495 **3.6. Evaluation indices**

496 Five statistical indicators, namely Root-mean-squared-error ($RMSE$) , Pearson's

497 correlation coefficient (R), Mean-absolute-error (MAE), unbiased $RMSE$ ($ubRMSE$)

498 and Kling-Gupta Efficiency (KGE), are selected for comparison with existing

499 products. The relative equations are shown as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(sim_i - obs_i)^2}{n}} \tag{23}$$

$$R = \frac{\sum_{i=1}^{n}\left(sim_i - \overline{sim}\right)\left(obs_i - \overline{obs}\right)}{\sqrt{\sum_{i=1}^{n}\left(sim_i - \overline{sim}\right)^2 \sum_{i=1}^{n}\left(obs_i - \overline{obs}\right)^2}} \tag{24}$$

$$-1 \leq R \leq 1$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|sim_i - obs_i| \tag{25}$$

$$ubRMSE = \sqrt{\frac{\sum_{i=1}^{n}\left[\left(sim_i - \overline{sim}\right) - \left(obs_i - \overline{obs}\right)\right]^2}{n}} \tag{26}$$

500 Where sim is the simulations, obs is the observation as reference.

501 The modified KGE (Kling et al., 2012) offers insights into reproducing temporal

502 dynamics and preserving the distribution of time series, which are increasingly used

503 to calibrate and evaluate hydrological models (Knoben et al., 2019). For a better

504 understanding of the KGE statistic and its advantages over the Nash-Sutcliffe

505 Efficiency (NSE), please refer to Gupta et al. (2009). The equation is given by:

$$KGE = 1 - \sqrt{(r-1)^2 + \left(\frac{\sigma_{sim}}{\sigma_{obs}} - 1\right)^2 + \left(\frac{\sigma_{sim}/\mu_{sim}}{\sigma_{obs}/\mu_{obs}} - 1\right)^2} \qquad (27)$$

506   Where σ_{obs} and σ_{sim} are the standard deviations of observations and simulations;

507   μ_{obs} and μ_{sim} are the mean of observations and simulations. Similar to NSE, $KGE = 1$

508   indicates perfect agreement of simulations, while $KGE < 0$ reveals that the average of

509   observations is better than simulations (Towner et al., 2019).

510   **4. Results**

511   In this study, we aimed to compare and evaluate the performance of fused products at

512   both site and global scales. At the site scale, the performance of the fused products

513   was evaluated against 212 FluxNet observations and compared with other products,

514   including the simple average. At the global scale, the mean and temporal variations of

515   the land surface ET calculated by the fused products were compared with those of

516   other products.

517   **4.1. Analysis of error variances and weights**

518   This section examines the random error variances and identifies the predominant

519   product based on assigned weights for the 0.1° and 0.25° inputs obtained through the

520   EIVD method.

521

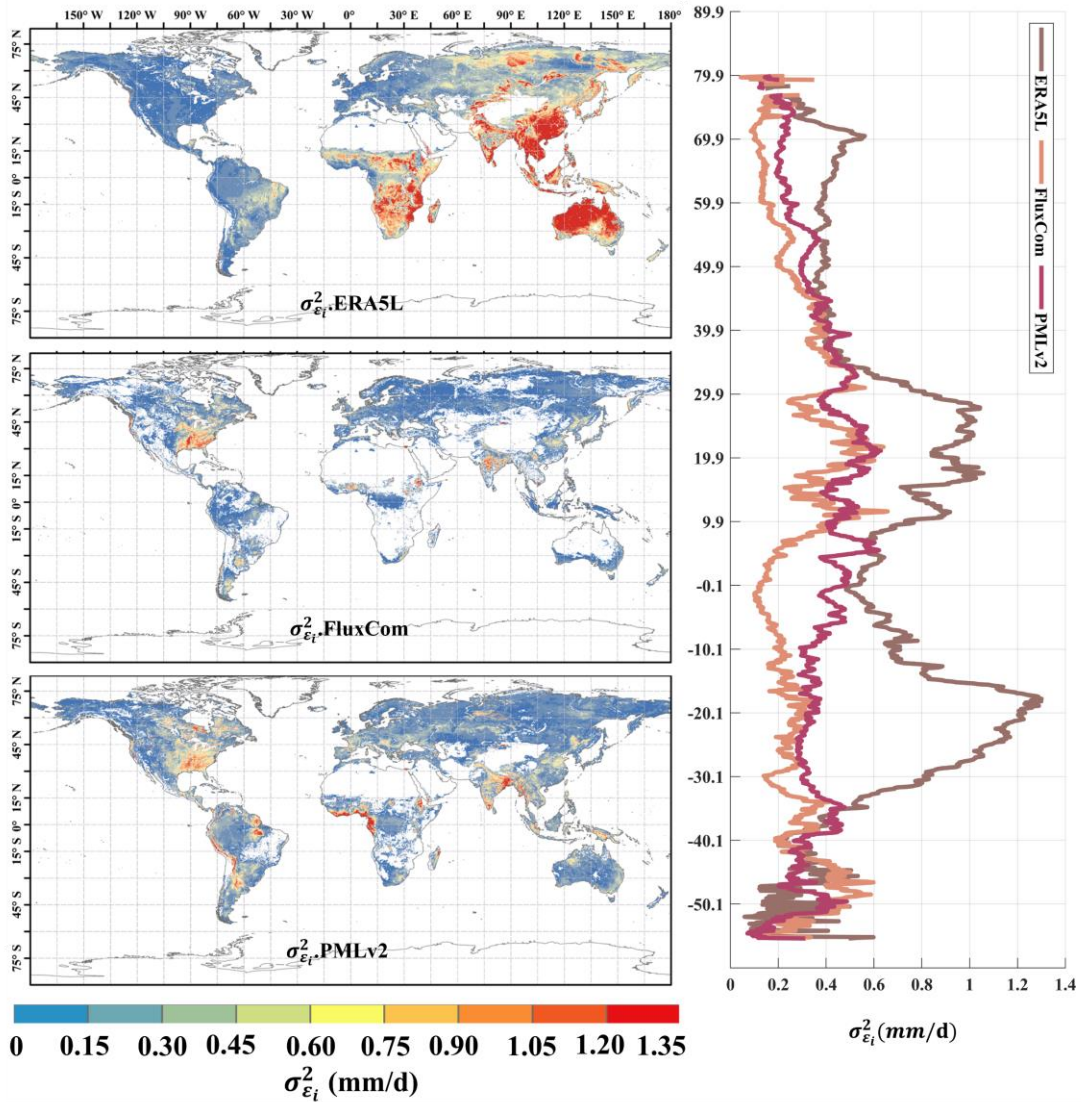



Figure 2 Global distribution of absolute error variances ($\sigma^2_{\varepsilon_i}$) of ERA5L, FluxCom,

and PMLv2 using EIVD at 0.1° from 2001 to 2015, depicted alongside corresponding

variation curves of average $\sigma^2_{\varepsilon_i}$ with latitude.

Figure 2 represents the random errors of the correlation products calculated using the

EIVD method from 2001 to 2015 at 0.1°, where a non-zero ECC is assumed between

FluxCom and PMLv2. The areas with missing values are due to the absence of data

from either FluxCom or PMLv2 in those regions. The global random error variances

(mean±standard deviation) obtained using the EIVD method are as follows: ERA5L:

0.58 ± 0.53 mm/day, FluxCom: 0.12 ± 0.13 mm/day, PMLv2: 0.17 ± 0.14 mm/day.

These results indicate that FluxCom performs best overall, while ERA5L performs the

poorest. Regarding spatial distribution, regions with more significant random errors in

533   ERA5L are mainly located in East Asia, Australia, and southern Africa. On the other

534   hand, FluxCom and PMLv2 show relatively more considerable uncertainties in the

535   southeastern United States. The latitude distribution reveals that ERA5L has the

536   highest uncertainty, primarily in the vicinity of 20° to 30° north and south, consistent

537   with its spatial distribution.

538   It is important to note that due to missing data in specific regions at 0.1°, such as

539   Northern Africa, the Sahara Desert region, Northwestern China, and Australia, the

540   error results obtained may not accurately reflect the performance of FluxCom and

541   PMLv2 in these areas. Considering the current results, we can cautiously conclude

542   that FluxCom and PMLv2 demonstrate better performance. Future data

543   supplementation in these regions would further enhance our ability to analyze the

544   products' accuracy.

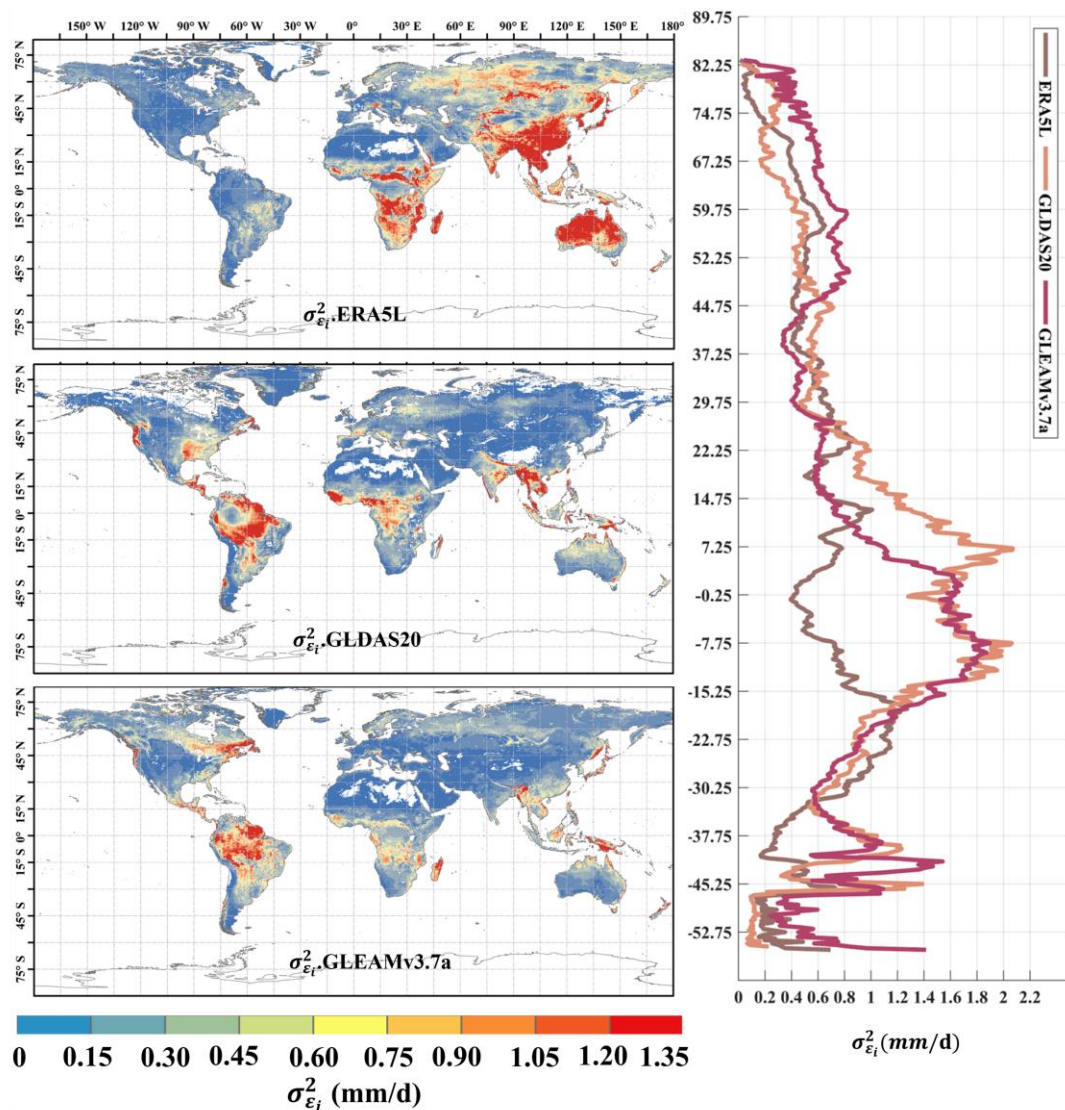



Figure 3 Global distribution of absolute error variances ($\sigma_{\varepsilon_i}^2$) of ERA5L, GLDAS2.0, and GLEAMv3.7a using EIVD at 0.25° from 1980 to 1999, depicted alongside corresponding variation curves of average with latitude.

The distribution of random error variance for ERA5L (0.59±0.58 mm/d), GLDAS2.0 (0.37±0.44 mm/d), and GLEAMv3.7a (0.38±0.36 mm/d) from 1980 to 1999 at 0.25° is shown in **Figure 3**. Here, we assumed a non-zero ECC between ERA5L and GLEAM. The ERA5L data was resampled from a 0.1° resolution to 0.25°, and its error distribution pattern is like that of the 0.1° resolution. It exhibits higher uncertainties in East Asia, Australia, and southern Africa. GLDAS and GLEAM exhibit relatively higher uncertainty over the southeastern United States and the Amazon Plain. GLDAS and GLEAM show similar performance among the three

557 products, while ERA5L performs relatively worse. Regarding the average distribution

558 with latitude, ERA5L demonstrates a more even distribution, whereas GLDAS and

559 GLEAM exhibit relatively higher uncertainties in tropical regions.

560 The ET calculations in both GLDAS and GLEAM involve complex surface

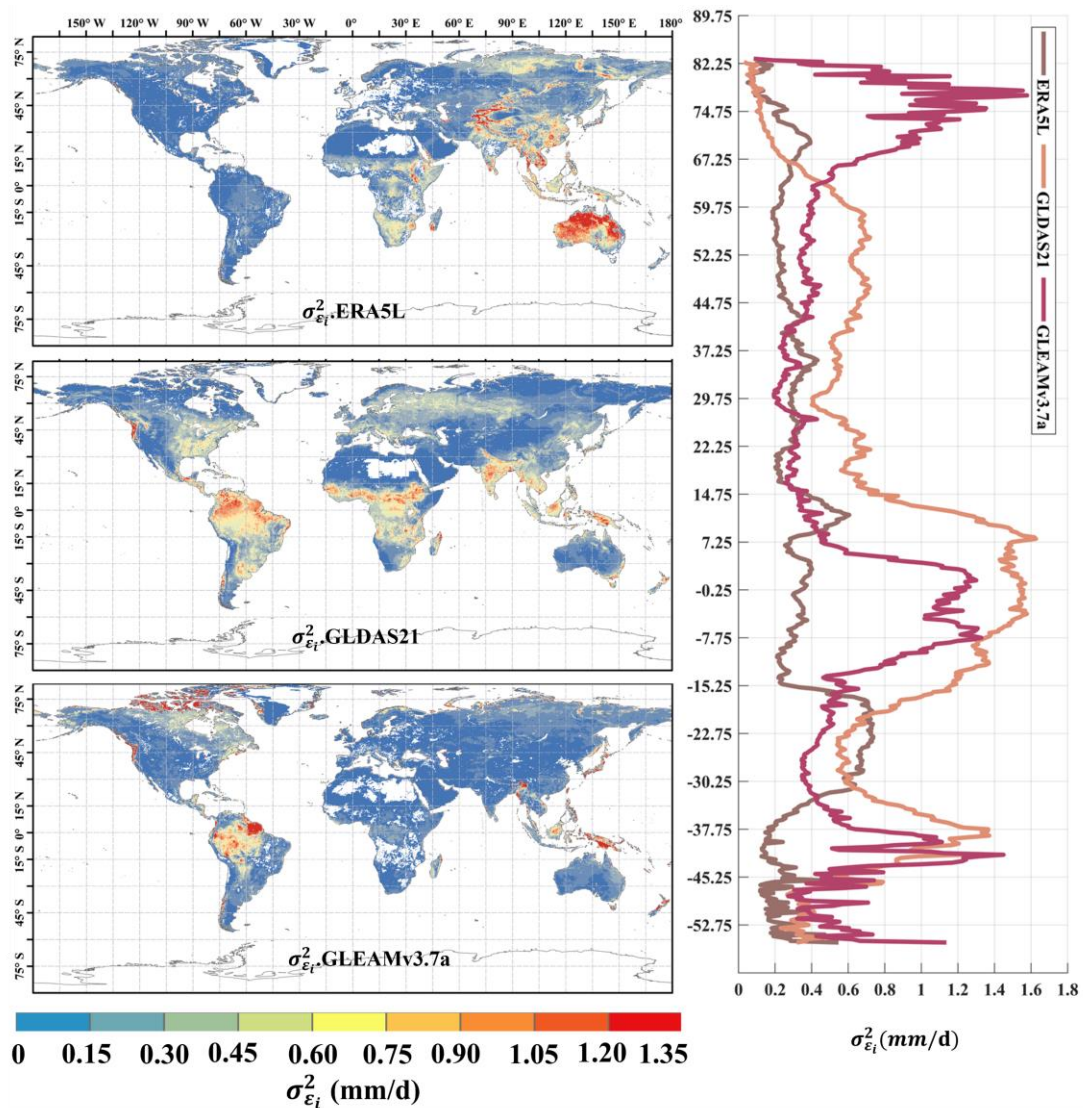
561 parameterization processes. In tropical regions, the high non-heterogeneity in land

562 covers poses a challenge, and the 0.25° resolution grid may not capture the intricacies

563 of the underlying surface conditions. This mismatch could impact the

564 parameterization process, leading to errors. Future work could involve in-depth model

565 analyses or sensitivity experiments to identify sources of error in complex ET models,

566 facilitating improvements.

567

568 **Figure 4** Global distribution of absolute error variances ($\sigma^2_{\varepsilon_i}$) of ERA5L, GLDAS2.1,

569   and GLEAMv3.7a using EIVD at 0.25° from 2000 to 2022, depicted alongside

570   corresponding variation curves of average with latitude.

571   In addition, **Figure 4** presents the distribution of random error variance for ERA5L

572   (0.32±0.33 mm/d), GLDAS2.1 (0.35±0.29 mm/d), and GLEAMv3.7a (0.38±0.36

573   mm/d) from 2000 to 2022 at a resolution of 0.25°. The non-zero ECC assumption was

574   made between ERA5L and GLEAM. In this combination, ERA5L shows significantly

575   lower errors than in previous periods, indicating improved ERA5L performance

576   during this time frame. However, ERA5L still exhibits more significant errors in the

577   East Asia and Australia regions compared to the other two datasets. The overall errors

578   for GLDAS and GLEAM have also decreased, but there are still random error

579   variances exceeding 1.0 mm/d in the Amazon plain and Indonesia region. Regarding

580   the latitudinal distribution, ERA5L shows relatively smooth changes, while GLDAS

581   and GLEAM exhibit similar trends. However, GLEAM demonstrates a noticeable

582   increase in errors near the Arctic.

583   Next, in **Figure 5**, we present the dominant product for each grid cell in the three

584   scenarios, where dominance refers to the product with the highest assigned weight.

585   The results in **Figure 5** indicate that at 0.1° resolution, the weights for FluxCom and

586   PMLv2 are significantly higher than ERA5L, aligning with the error calculations

587   presented in **Figure 2**. This underscores the effectiveness of error and weight analysis

588   based on collocation in reflecting product performance, thereby allowing for a rational

589   adaptation of weights. At 0.25° resolution, the dominant regions for ERA5L,

590   GLDAS-2, and GLEAM products are relatively balanced. In the fusion scenario from

591   1980 to 1999, GLDAS20 predominantly covers the Northern Hemisphere, while

592   GLEAM dominates the Southern Hemisphere, with ERA5L prevalent in the Amazon

593   region. However, in the fusion scenario from 2000 to 2022, GLEAM's dominant

594   region significantly expanded, primarily covering the central United States and

595   southeastern China. The Amazon region continues to be dominated by ERA5L. The

596   variation in dominant products highlights that the calculation of product weights

597   evolves with changes in the fusion scenario. The error and weight computation

598   methods based on collocation can only provide the minimum MSE solution for a

599   given combination of inputs. It is important to note that changes in inputs will impact

600   the results.

601

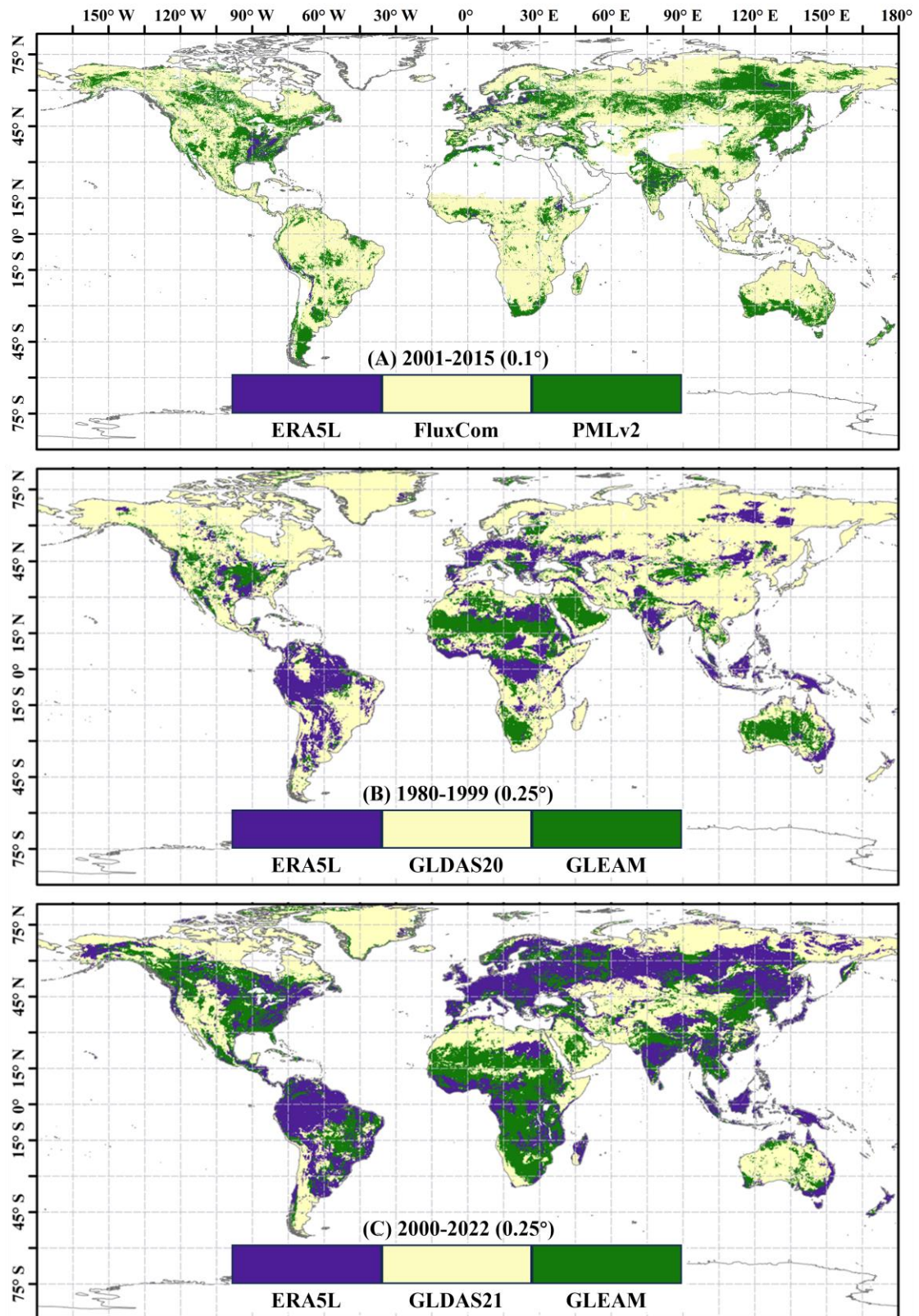



Figure 5 Map of the prevailing product at individual pixels based on scenario-specific

weights.

For the analysis at a resolution of 0.1°, we also applied the IVD method to calculate

605 the errors between ERA5L and PMLv2 for two time periods: 2000 and 2015 to 2020.

606 Since the analysis of product errors is not the focus of this paper, we provide the

607 results of the IVD in the appendix. Grids with higher random error variances

608 correspond to smaller weights when calculating the weights. The weight distribution

609 calculated at different time intervals is available in the appendix.

610 4.2. Site-scale evaluation and comparison

611 At the site scale, the performance of CAMELE was compared with FluxNet as the

612 reference. In this subsection, **Figure 6** and **Table 3** correspond to each other, as they

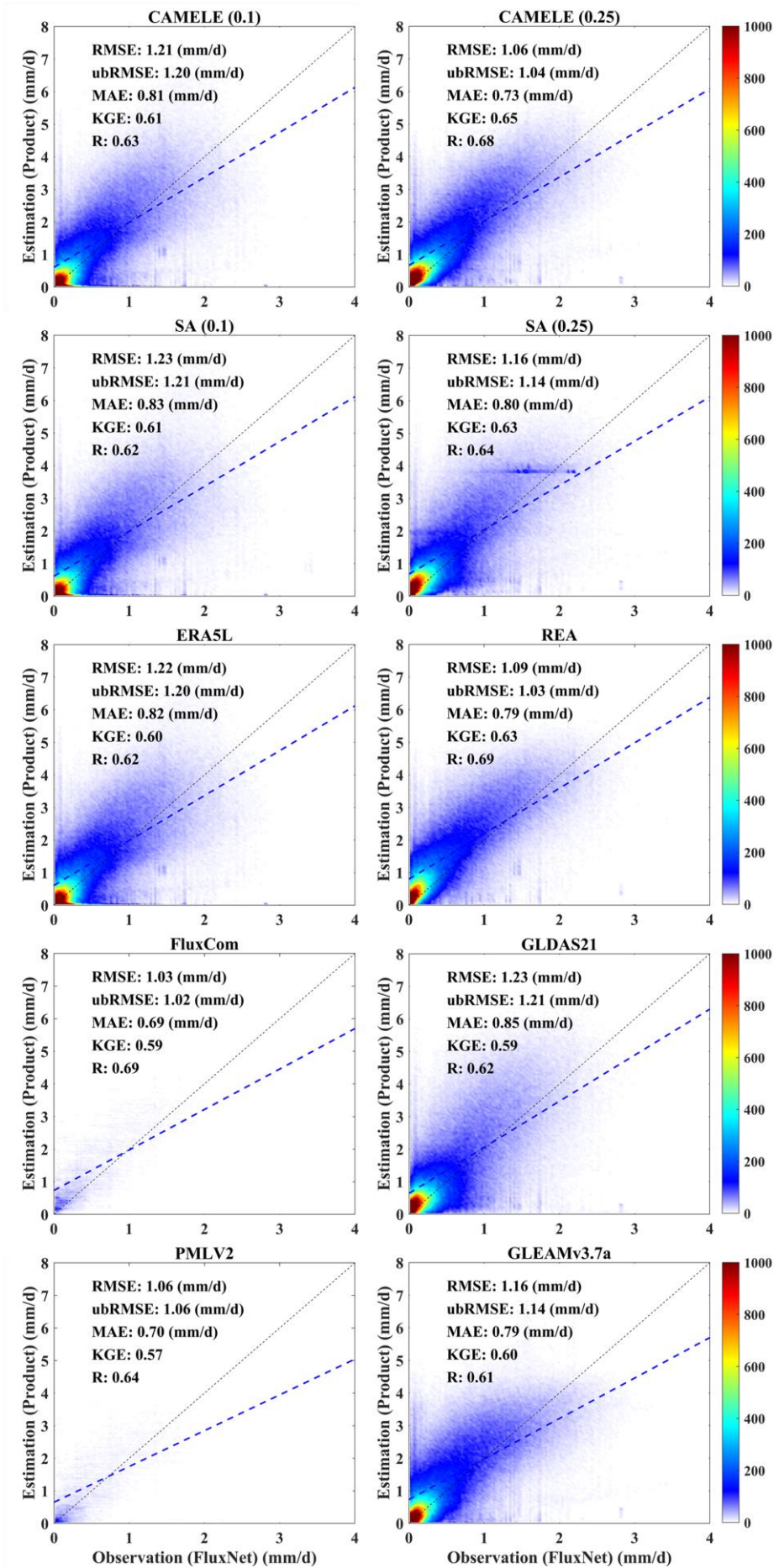
613 integrate data from 212 sites for all available periods, allowing for a comparative

614 analysis of the performance of different products at different times. Similarly, **Figure**

615 **7** and **Table 4** correspond to each other, where different product metrics were

616 calculated for each site, and the calculated metric results were subjected to statistical

617 analysis.

618

619   **Figure 6** Scatter plots of product corresponding to the available period data from 212

620   FluxNet sites. The colorbar represents the density, with darker colors indicating

621   higher concentration. The left and right columns present results for 0.1° and 0.25°

622   resolutions, respectively, with "SA" indicating the results for simple average.

623   Relevant statistical metrics are annotated in their respective figures.

624   The scatter plots in **Figure 6** demonstrate that CAMELE consistently performs at 0.1°

625   and 0.25° resolutions. At 0.1° resolution, FluxCom and PMLv2 showed superior

626   performance with fewer data points due to their original 8-day average resolution.

627   CAMELE exhibited a performance like ERA5L. At 0.25° resolution, CAMELE

628   performed comparably to the other datasets, demonstrating reasonable accuracy.

629   Notably, there was an improvement in the KGE and R indices. The fitted line closely

630   approximated the 1:1 line, indicating a solid agreement with the observed values.

631   Moreover, the results obtained from the simple average were also acceptable, but SA

632   (0.25°) had a concentration of data points between (2-4 mm/d), possibly due to the

633   inputs having a high concentration within that range. The assumption that a simple

634   average implies equal performance of each product on every grid cell is inaccurate;

635   variations in performance exist among different products across distinct grid cells

636   (regions).

637   **Table 3** Average values of different metrics for CAMELE and other fusion schemes

638   corresponding to the available period data from 212 FluxNet sites. The bolded

639   sections indicate the schemes with the best performance in their respective metrics.

Product		RMSE (mm/d)	ubRMSE (mm/d)	MAE (mm/d)	KGE	R
	CAMELE	1.21	1.20	0.81	**0.61**	0.63
	SA	1.23	1.21	0.83	0.61	0.62
0.1°-daily	ERA5L	1.22	1.20	0.82	0.60	0.62
	FluxCom	**1.03**	**1.02**	**0.69**	0.59	**0.69**
	PMLv2	1.06	1.06	0.70	0.57	0.64
0.25°-daily	**CAMELE**	**1.06**	1.04	**0.73**	**0.65**	0.68
	SA	1.16	1.14	0.80	0.63	0.64

REA	1.09	**1.03**	0.79	0.63	**0.69**
GLDAS21	1.23	1.21	0.85	0.59	0.62
GLEAMv3.7a	1.16	1.14	0.79	0.60	0.61

640 The information in **Table 3** corresponds to **Figure 6** and presents the results of

641 various product indicators. The bolded parts indicate the products with the best

642 corresponding indicators. The results indicate that CAMELE performed well at both

643 0.1° and 0.25° resolutions, mainly showing improvements in the KGE and R

644 indicators. FluxCom exhibited the best performance; however, considering that this

645 product utilized FluxNet sites for result calibration, this phenomenon is reasonable. In

646 this study, we pooled the data from all 212 available periods at the stations as a

647 reference without considering the differences between individual sites. This approach

648 provided an initial validation of the reliability of CAMELE at all sites.

649 The information in **Figure 7** corresponds to the data presented in **Table 4**, which

650 involves the calculation of five indicators at each site, followed by statistical analysis

651 of these indicators. From the distribution of the violin plots, it can be observed that a

652 violin plot with a closer belly to 1 indicates better results in terms of the R and KGE

653 indicators. CAMELE performs well overall, closely resembling PMLv2 and FluxCom.

654 On the other hand, the results obtained from the Simple Average are relatively poorer.

655 Regarding the RMSE, ubRMSE, and MAE indicators, a violin plot with a closer belly

656 to 0 suggests less errors. CAMELE demonstrates a notable enhancement in

657 performance at the 0.1° level. This suggests that the fusion method effectively reduces

658 errors, aligning with the original intention of weight calculation, and it compares

659 favorably with the products used in the merging scheme.

660 Additionally, FluxCom and PMLv2 also exhibit minimal errors, which is expected

661 considering their utilization of FluxNet sites for error correction. Furthermore, SA

662 shows significantly larger errors. Although the simple average method can

663 compensate for positive and negative errors between inputs in some instances, it can

664 also lead to error accumulation, as evidenced by the results in the violin plots.

665

666

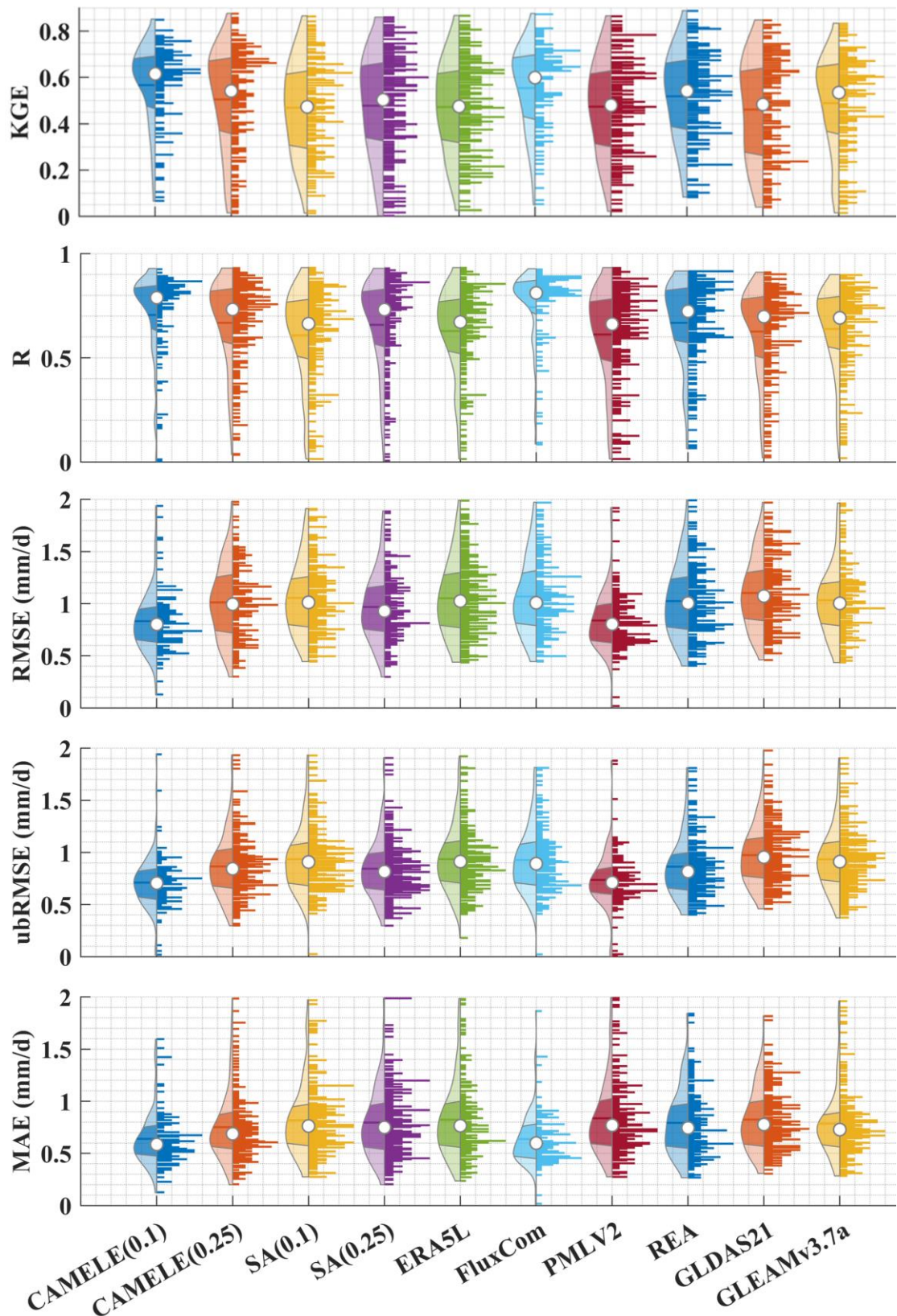



Figure 7 Violin plots obtained by aggregating five different statistical indicators,

calculated separately for each site. In each violin plot, the left side represents the

669   distribution, with the shaded area indicating the box plot, the dot representing the

670   mean, and the right side showing the histogram.

671   **Table 4** Average values of indicators corresponding to different products, calculated

672   based on the comprehensive results obtained for each site. The bolded sections

673   indicate the schemes with the best performance in their respective metrics.

Product		RMSE (mm/d)	ubRMSE (mm/d)	MAE (mm/d)	KGE	R
	CAMELE	**0.83**	**0.71**	**0.64**	**0.57**	0.71
	SA	1.05	0.93	0.82	0.47	0.61
0.1°-daily	ERA5L	1.05	0.94	0.82	0.47	0.63
	FluxCom	1.07	0.93	0.64	0.55	**0.74**
	PMLv2	0.84	0.74	0.84	0.47	0.61
	CAMELE	**1.03**	0.87	**0.75**	**0.51**	**0.67**
	SA	0.97	**0.84**	0.80	0.48	0.66
0.25°-daily	REA	1.02	0.86	0.80	0.48	0.67
	GLDAS21	1.10	0.97	0.83	0.46	0.63
	GLEAMv3.7a	1.03	0.93	0.79	0.49	0.64

674   **Table 4** presents the average values of different metrics in **Figure 7**, boldly

675   highlighting the optimal products corresponding to each metric. It can be observed

676   that CAMELE exhibits significant improvements in performance at a resolution of

677   0.1°, particularly in terms of the error metrics RMSE and ubRMSE, surpassing other

678   products. This further confirms the effectiveness of our fusion scheme in reducing

679   product errors. Additionally, although the performance of CAMELE at a resolution of

680   0.25° is comparable to other products, there is still a slight decline compared to its

681   performance at 0.1°. This can be attributed partly to the inherent errors in the input

682   products and partly to the decreasing representativeness of FluxNet, which serves as

683   the reference at the 0.25° grid. Nevertheless, we can still consider CAMELE to have

684   good accuracy.

Figure 8 Heatmaps of five statistical indicators, where each row corresponds to the mean value for all sites of the specific PFT, and each column corresponds to a product. The product with the best performance for that PFT is highlighted in bold within each row. (a)-(c) represent three error indicators: RMSE, ubRMSE, and MAE; (d)-(e) represent two goodness-of-fit indicators: KGE and R.

Table 5 Optimal product corresponding to different PFTs under various statistical indicators against observations from FluxNet sites

IGBP (n-sites)	RMSE (mm/d)	ubRMSE (mm/d)	MAE (mm/d)	KGE	R
CRO (20)	**CAMELE**	**CAMELE**	**CAMELE**	PMLv2	**CAMELE**

CSH (3)		PMLv2		FluxCom	FluxCom
DBF (26)				REA	
DNF (1)			FluxCom	**CAMELE**	
EBF (15)			**CAMELE**	GLEAM	
ENF (49)		**CAMELE**	FluxCom	**CAMELE**	
GRA (39)	PMLv2				**CAMELE**
MF (9)			**CAMELE**	REA	
OSH (13)			FluxCom	**CAMELE**	
SAV (9)	**CAMELE**				FluxCom
SNO (1)			CAMELE	REA	
WET (21)		PMLv2	FluxCom	**CAMELE**	
WSA (6)		**CAMELE**			

693   Furthermore, we classified 212 sites according to PFTs and analyzed the statistical

694   indicators of different PFTs corresponding to each site. The results are represented in

695   **Figure 8** as a heatmap, and the corresponding optimal products for other PFTs sites

696   are marked in **Table 5**. The results show that CAMELE performs the best in almost

697   all PFTs categories, as indicated by various indicators. While on sites where other

698   products perform better, CAMELE's indicators are comparable to the optimal

699   products, albeit slightly inferior. This indicates that our fusion approach effectively

700   combines the advantages of different products, resulting in superior fusion results

701   across different vegetation types.

702   From the results, it is evident that CAMELE performs well across various vegetation

703   types. To delve deeper into the reasons behind this performance, we conduct site-scale

704   analyses at two resolutions, evaluating errors and computed weights for different

705   PFTs sites. These are visualized in radar chart format in **Figure 9**.

706

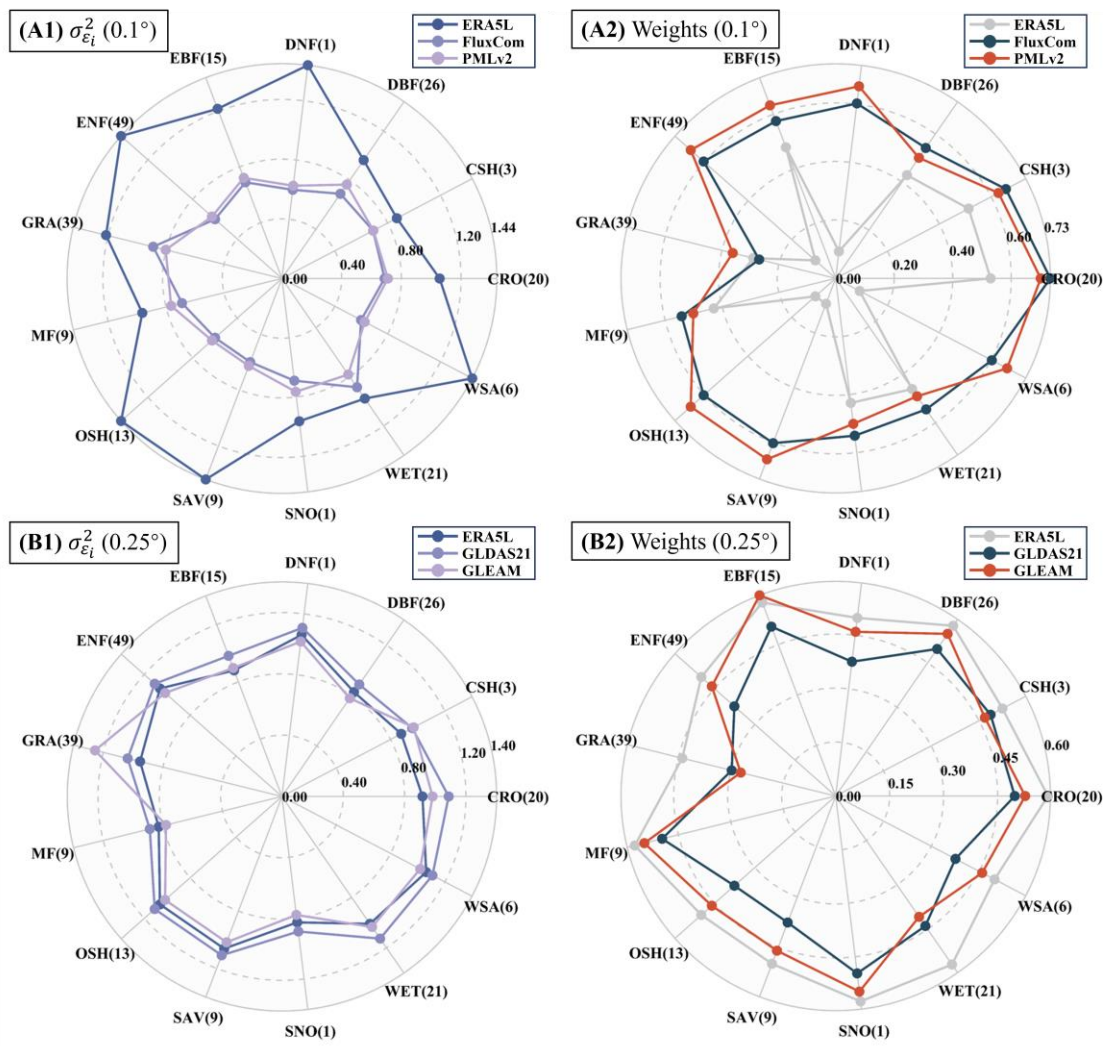



Figure 9 Mean collocation-based errors and weights of different products at various

PFTs sites at (A) 0.1° and (B) 0.25° resolutions. The parentheses next to each PFTs

name denote the corresponding number of sites.

The results from **Figure 9** demonstrate that the error-weighting calculation method

based on collocation effectively considers the error situation of inputs, thereby

providing reasonable weight assignments. At 0.1° resolution, ERA5L's error is

significantly higher across all PFTs than FluxCom and PMLv2, resulting in relatively

lower corresponding weights. FluxCom and PMLv2 exhibit closer performance, with

higher weights at most PFT sites. At 0.25° resolution, ERA5L, GLDAS21, and

GLEAM perform more evenly, with minimal differences, resulting in closer weights.

The weights for different inputs vary noticeably with changes in PFTs, depending on

the performance of other products within the same combination. Products with more

719   significant errors correspondingly have lower weights, affirming the rationale behind

720   the fusion method. However, it is essential to note that the presented results depict the

721   mean values of errors and weights across all sites; there might be variations among

722   sites with the same PFTs.

723   In summary, using the filtered daily-scale data from 212 FluxNet sites as a reference,

724   we conducted a benchmark analysis with CAMELE and demonstrated its good fit

725   with the observed data. Additionally, by comparing the performance of different

726   products at each site, we further illustrated that CAMELE exhibits similar or slightly

727   improved accuracy and minor errors compared to existing products.

728   **4.3. Assessment and comparison of multi-year average**

729   In this section, we will first analyze and compare the performance of CAMELE with

730   other products in estimating the multi-year mean and extreme values of ET at the site

731   scale. Subsequently, a global-scale analysis will be conducted for the same periods

732   ($0.1°$: 2001 to 2015; $0.25°$: 2000 to 2017) to examine the distribution of multi-year

733   daily average ET calculated by different products. For site comparisons, we have

734   selected monthly mean ET values and three quantiles (5th, 50th, and 95th) to

735   represent the products' performance in estimating ET's average and extreme values.

736

737

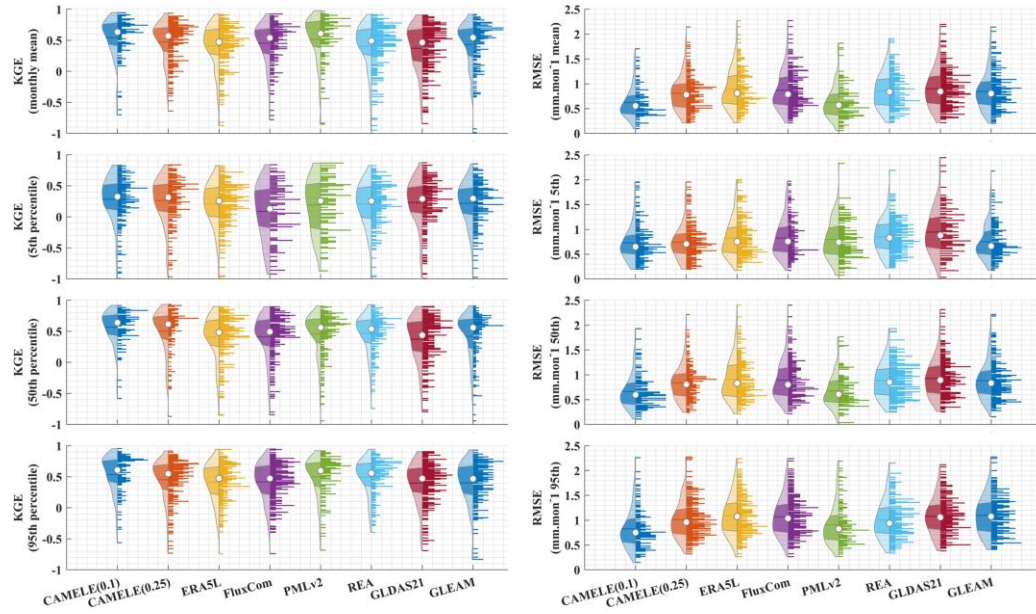



Figure 10 Violin plots depicting the KGE and RMSE metrics calculated for
CAMELE and other products based on the monthly mean, 5th, 50th, and 95th
percentiles at each FluxNet site. The left four columns represent KGE plots, while the
right four columns represent RMSE plots. The dots in the violin plots represent the
median, and the horizontal lines represent the mean.

Table 6 Average values of KGE and RMSE corresponding to different products,
calculated based on the results obtained for each site. The bolded sections indicate the
schemes with the best performance in their respective metrics.

Product		KGE			
		Mean	**5th**	**50th**	**95th**
0.1°-daily	**CAMELE**	**0.54**	**0.28**	**0.57**	**0.54**
	ERA5L	0.41	0.21	0.40	0.42
	FluxCom	0.45	0.09	0.42	0.42
	PMLv2	0.52	0.19	0.46	0.50
0.25°-daily	**CAMELE**	**0.47**	**0.26**	**0.50**	0.45
	REA	0.40	0.21	0.46	**0.50**
	GLDAS21	0.37	0.23	0.37	0.40
	GLEAMv3.7a	0.43	0.22	0.42	0.40
Product		RMSE (mm/mon)			
		Mean	**5th**	**50th**	**95th**
0.1°-daily	**CAMELE**	0.63	**0.73**	**0.66**	**0.83**

	ERA5L	0.89	0.83	0.91	1.09
	FluxCom	0.87	0.83	0.89	1.07
	PMLv2	**0.63**	0.80	0.68	0.91
	CAMELE	**0.81**	**0.74**	**0.84**	1.01
0.25°-daily	REA	0.86	0.85	0.88	**1.01**
	GLDAS21	0.90	0.95	0.93	1.08
	GLEAMv3.7a	0.85	0.75	0.88	1.10

746 The information in **Figure 10** corresponds to the data presented in **Table 6**, which

747 involves the calculation of KGE and RMSE at each site, followed by statistical

748 analysis. From the distribution of the violin plots, it can be observed that a violin plot

749 with a closer belly to 1 indicates better results in terms of the KGE.

750 The results show that CAMELE outperforms other products in the estimation of

751 monthly averages and the 5th, 50th, and 95th percentiles at both 0.1° and 0.25°

752 resolutions. Its performance in capturing monthly averages is noteworthy, with a

753 noticeable improvement in the KGE and RMSE metrics relative to the inputs.

754 Examining the results for percentiles, CAMELE shows a relatively poorer estimation

755 for shallow values (5th percentile) but still demonstrates some improvement

756 compared to the input data, albeit influenced by input errors.

757 At 0.1°, PMLv2 and FluxCom perform just below the fusion result, aligning with the

758 previous error and weight analysis. At 0.25°, GLEAM and REA closely follow

759 CAMELE, with REA exhibiting slightly better estimation results for extremely high

760 values (95th percentile) than CAMELE. Despite this, the analysis results still indicate

761 that the products obtained reflect well the multi-year averages and extremes of ET,

762 holding promise as reliable products for analyzing ET variations.

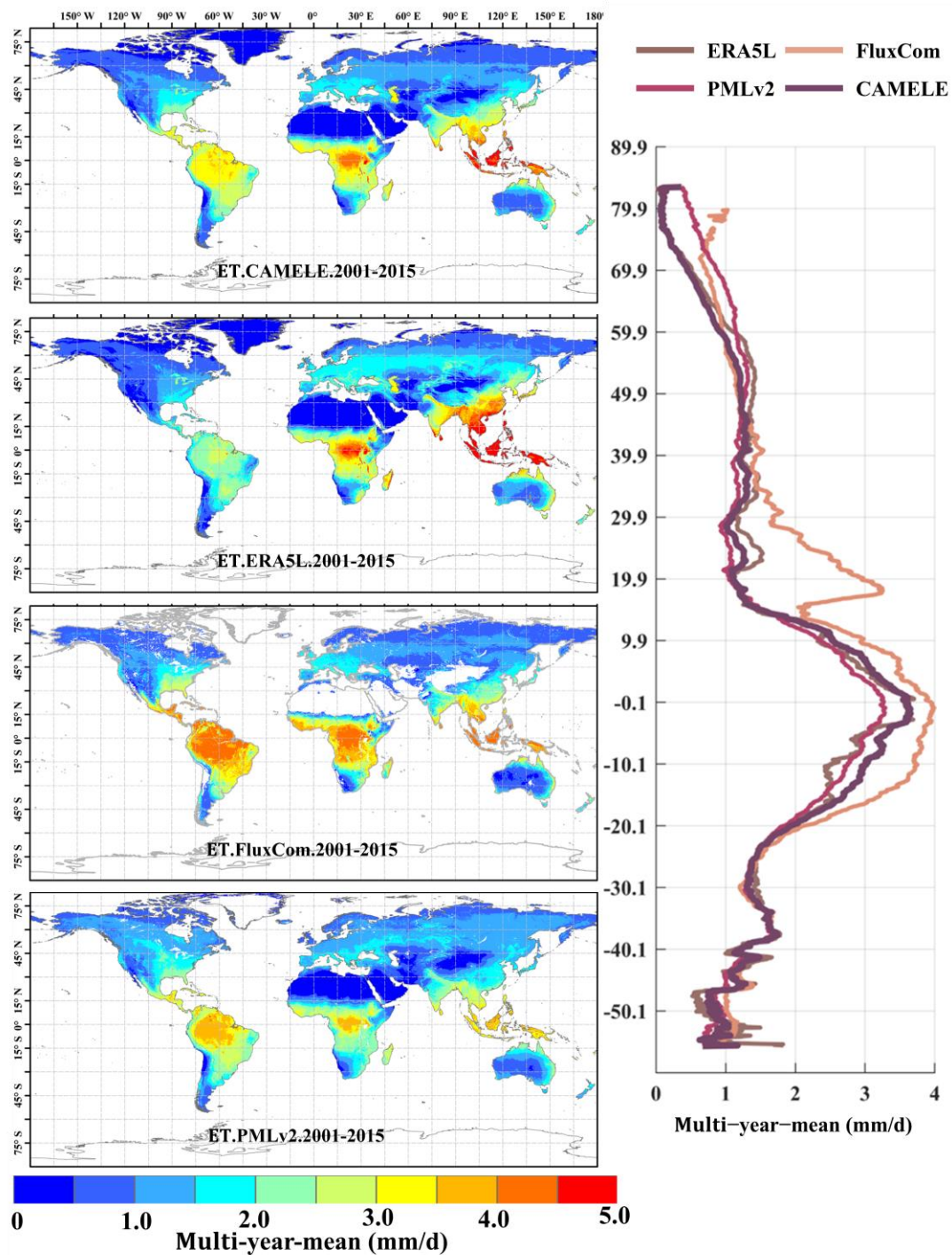



Figure 11 Global distribution of multi-year daily average ET at 0.1° for CAMELE, ERA5L, FluxCom, and PMLv2, depicted alongside corresponding variation curves of average with latitude.

The results in **Figure 11** indicate significant differences in the multi-year daily average distribution of global evapotranspiration (ET) among different products. Specifically, ERA5L shows noticeably higher values in East Asia than other products,

770   while FluxCom and PMLv2 exhibit higher values in the Amazon rainforest and

771   southern Africa regions. This distribution pattern is consistent with the error results

772   obtained from the EIVD calculation, indicating that these products possess certain

773   uncertainties in the regions. In terms of the latitudinal distribution pattern, except for

774   FluxCom, which displays distinct fluctuations, the variability among the other

775   products is relatively similar. This suggests that despite spatial differences among the

776   different products, they maintain consistency in the overall quantity.

777   **Figure 12** presents the results with a resolution of 0.25°. It can be observed that

778   compared to the 0.1° distribution, the spatial distribution of annual average

779   evapotranspiration (ET) is more consistent among different products at 0.25°,

780   showing larger ET values in tropical regions. The main differences are concentrated

781   in the Amazon rainforest and the Congo Basin, where GLEAM and GLDAS results

782   are higher than REA's. The assigned weights for REA's inputs (MERRA2, GLDAS,

783   and GLEAM.) are approximately equal in these two regions, each contributing about

784   one-third to the overall calculation (Lu et al., 2021). This balanced allocation results

785   in the REA being distributed among them roughly equally over multiple years in these

786   two regions. The latitude variation plots show that the results from each product are

787   very close, providing additional evidence for the reliability of CAMELE.

788

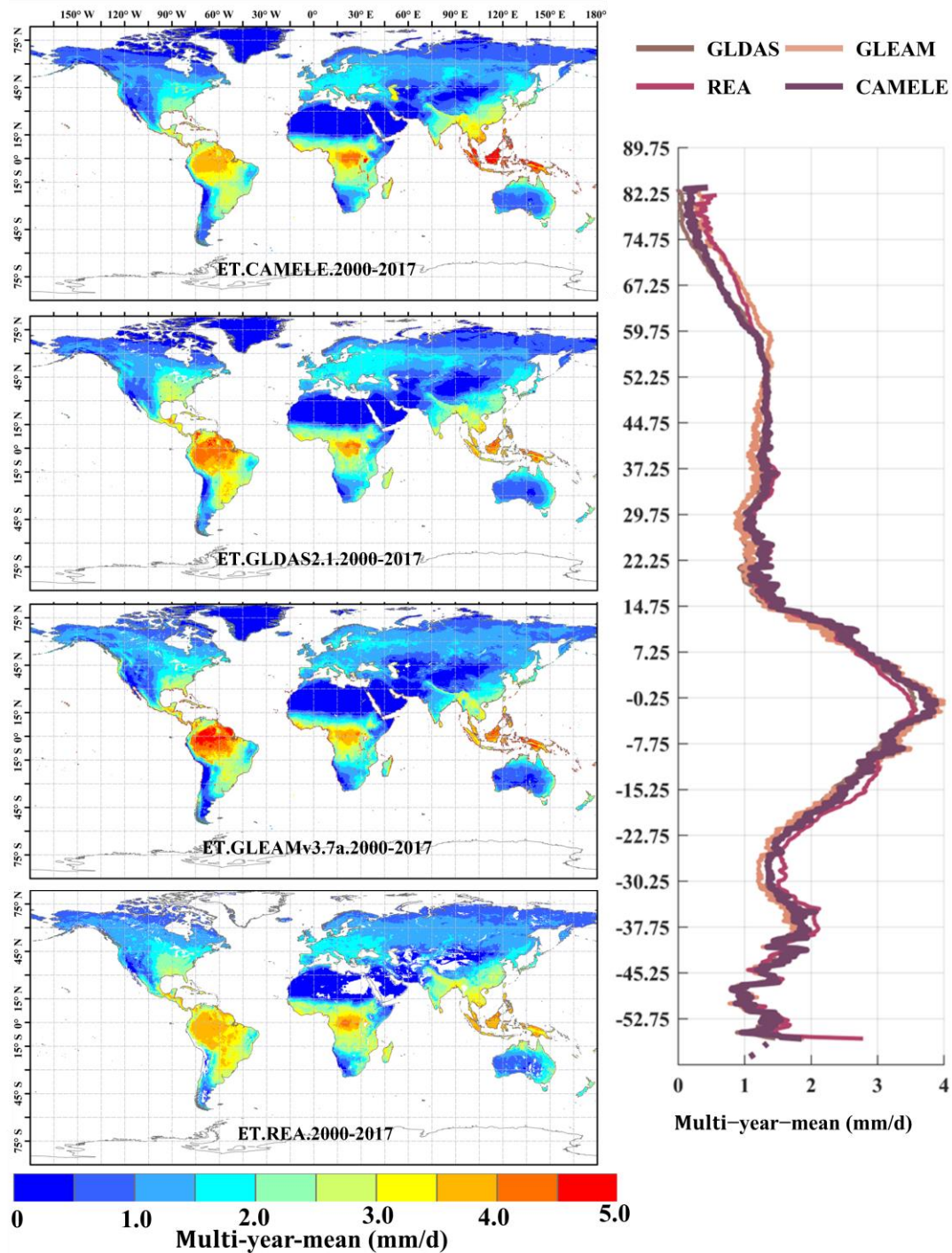



Figure 12 Global distribution of multi-year daily average ET at 0.25° for CAMELE,

GLDAS2.1, GLEAMv3.7a, and REA, depicted alongside corresponding variation

curves of average with latitude.

In parallel, it is worth noting that, despite the regional disparities that may arise when

contrasting the trends by CAMELE with inputs, a noteworthy consistency emerges

when examining these trends along latitudinal gradients. This notable alignment

796 signifies the robustness of CAMELE to some extent. It underscores the capacity of

797 CAMELE to capture ET patterns, providing further insights for the scientific

798 community.

4.4. Assessment and comparison of linear trend and seasonality

800 In this section, we first validate and compare the performance of CAMELE with other

801 products in estimating multi-year trends and seasonality at the site scale. Due to the

802 inconsistent time lengths of FluxNet sites, trends at many sites are not significant.

803 Therefore, we deliberately selected 13 sites with continuous evapotranspiration (ET)

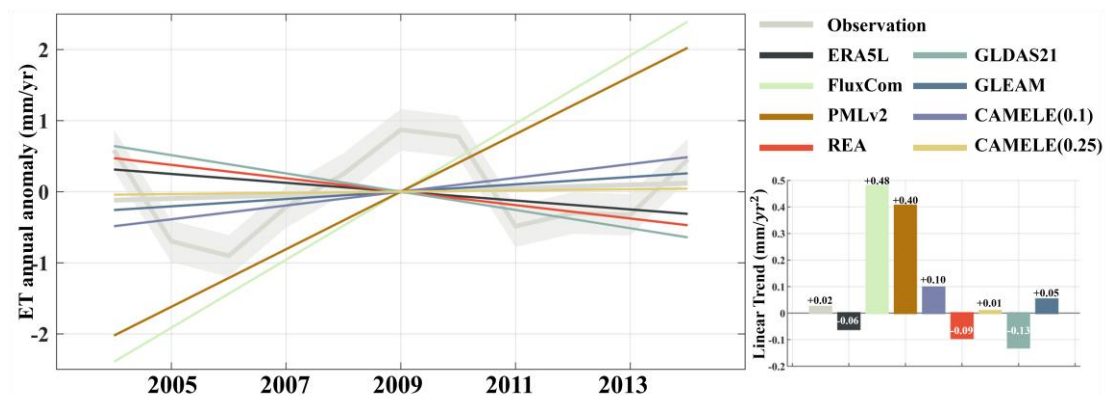
804 observations for the same 11-year period (2004 to 2014) and with significant trends.

805 The annual ET values for each year were calculated as the mean of the 13 sites for

806 that year, allowing the computation of linear trends and seasonality. We employed

807 singular spectrum analysis (SSA), which assumes an additive decomposition A = LT

808 + ST + R. In this decomposition, LT represents the long-term trend in the data, ST is

809 the seasonal or oscillatory trend (or trends), and R is the remainder.

810

811 **Figure 13** Comparison of linear trend from 2004 to 2014 among 13 FluxNet sites

812 using CAMELE and other products. The trends have been subjected to SSA

813 decomposition, removing seasonality. The gray enveloping line represents the mean

814 plus the standard deviation of the 13 sites.

815

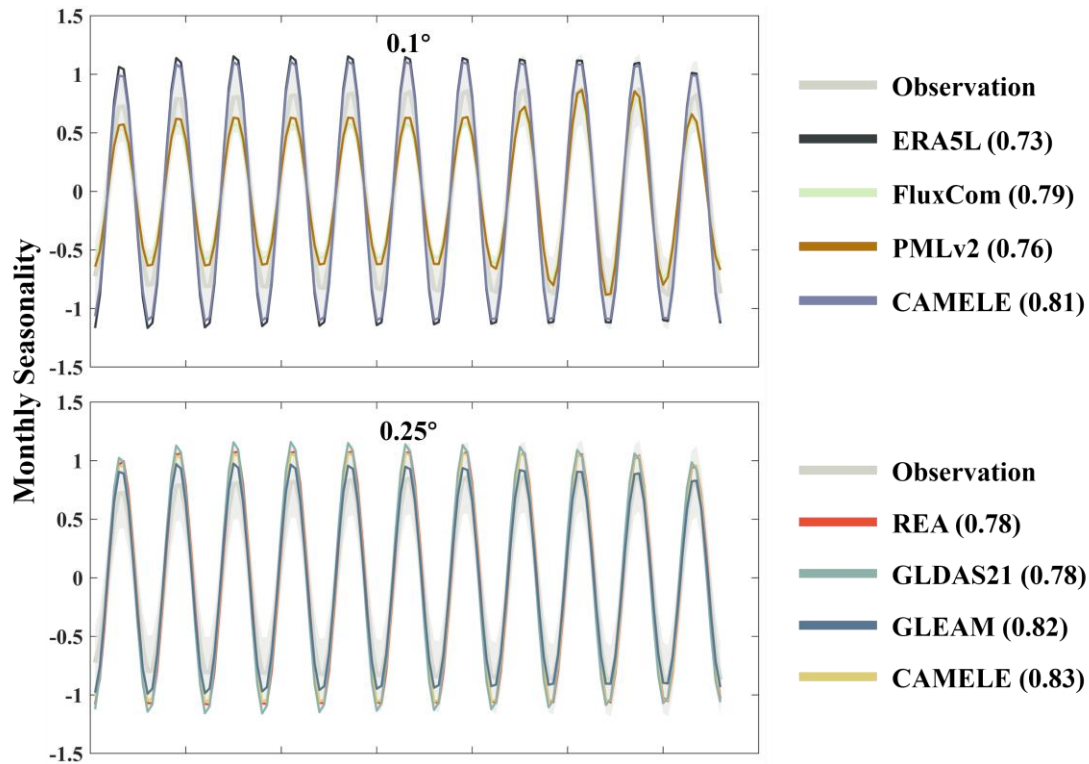



Figure 14 Comparison of seasonal variations from 2004 to 2014 among 13 FluxNet

sites using CAMELE and other products. The seasonality has been obtained through

SSA decomposition, with the gray area representing the observed values. The

parentheses in each product name indicate the KGE coefficient comparing with the

observed values.

In **Figure 13** and **Figure 14**, based on observations from FluxNet sites, we analyzed

the performance of CAMELE and other products in estimating the linear trend and

seasonality of ET over multiple years. It is important to note that we only present the

analysis results for 13 sites with continuous 11-year observations, and the

performance of different ET products in trend estimation at individual sites still varies,

not fully reflecting the overall performance on all grids in terms of trend and

seasonality. Nevertheless, such a comparison can still provide valuable insights.

Examining the results of the linear trend, both PMLv2 and FluxCom exhibit a

significant upward trend, well above the observations. On the contrary, ERA5L,

GLDAS, and REA show a noticeable downward trend, while CAMELE demonstrates

a gradual upward trend closer to the observations. Additionally, GLEAM slightly

832   outperforming CAMELE at a resolution of 0.25°. Overall, CAMELE shows good

833   agreement with site observations in capturing the multi-year linear trend of ET.

834   Continuing with the analysis of seasonality, the KGE index comparing each product's

835   results with observed values is provided in parentheses next to the product name.

836   Generally, all products exhibit a good representation of ET's seasonal variations.

837   CAMELE's 0.1° seasonal results closely match FluxCom (with the two lines almost

838   overlapping). However, the fluctuations it reflects are higher than the observed values.

839   This is likely due to keeping the 8-day average results of FluxCom consistent with

840   PMLv2 every 8 days, and the variability in ET primarily originates from ERA5L

841   results. This aspect may need improvement in subsequent research. At 0.25°,

842   CAMELE's seasonal representation is closer to the observed results. The differences

843   in CAMELE's performance at the two resolutions are mainly attributed to input

844   variations, which we discuss in the following section as potential areas for

845   improvement.

846   The results indicate that CAMELE effectively captures the multi-year changes in ET,

847   but at 0.1°, it tends to overestimate seasonal fluctuations. We further generated global

848   maps of multi-year linear trends in ET, estimating trends using Theil–Sen's slope

849   method and testing significance with the Mann–Kendall method. The dotted areas

850   indicate trends passing a significance test at a 5% level.

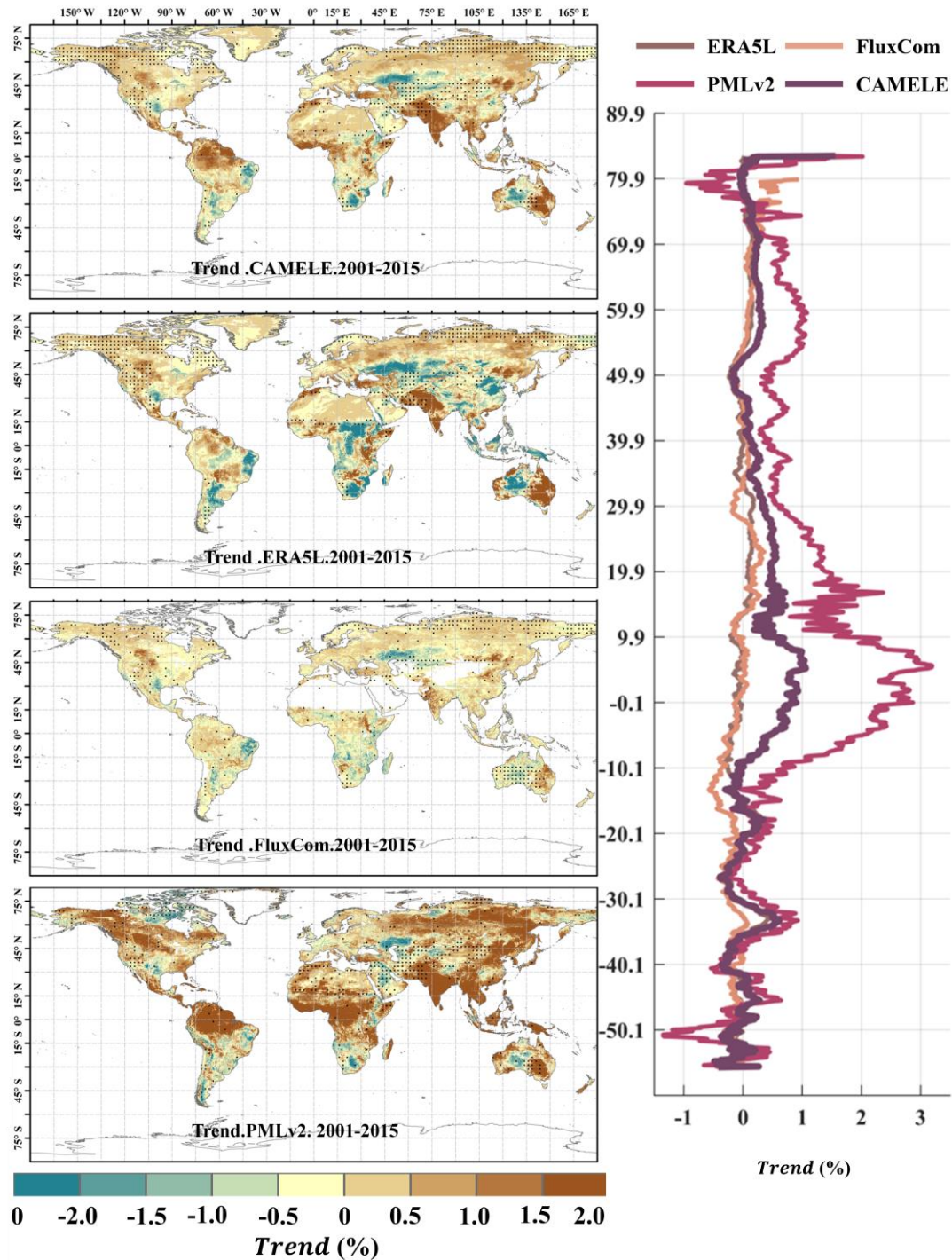



Figure 15 Global distribution of multi-year linear trend at 0.1° for CAMELE, ERA5L,

FluxCom, and PMLv2, depicted alongside corresponding average trend with latitude.

The trend is estimated with Theil–Sen's slope method, and the significance level is

tested with the Mann–Kendall method. The dotted area indicates that the trend has

passed the significance test at 5 % level.

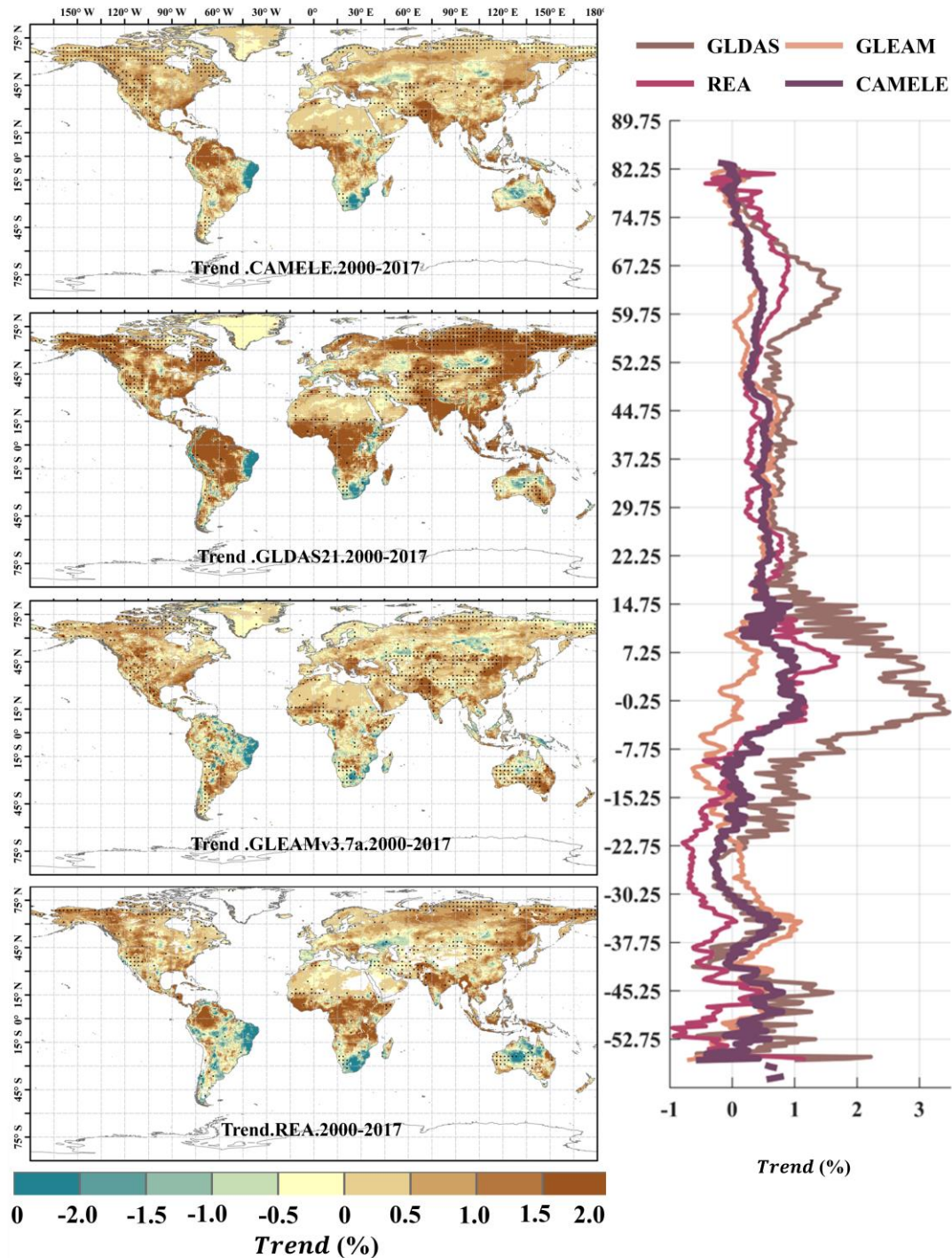



Figure 16 Global distribution of multi-year linear trend at 0.25° for CAMELE,

GLDAS2.1, GLEAMv3.7a, and REA, depicted alongside corresponding average

trend with latitude. The trend is estimated with Theil–Sen's slope method, and the

significance level is tested with the Mann–Kendall method. The dotted area indicates

that the trend has passed the significance test at 5 % level.

Figure 15 and **Figure 16** present the linear trends of multi-year daily scale

865 evapotranspiration (ET) calculated for different products at resolutions of 0.1° and

866 0.25°, respectively. The corresponding latitude-dependent variations of the rate of

867 change are shown on the right side. It can be observed that the differences in linear

868 trends among the different products are more significant than the multi-year averages,

869 and in some regions, they even exhibit opposite trends. For example, at 0.1°

870 resolution, PMLv2 shows a global increase of 1.0% in ET in most regions, while the

871 results from CAMELE, ERA5L, and PMLv2 indicate a milder increase in ET in the

872 Amazon rainforest, southern Africa, and northwestern Australia. At 0.25° resolution,

873 except for GLDAS2.1, which shows an apparent global increase in ET, the results

874 from CAMELE, GLEAMv3.7a, and REA indicate milder variations in global ET.

875 **5. Discussion**

876 **5.1. Impact of underlying assumptions in collocation analysis**

877 The collocation analysis system relies on key assumptions, including linearity (linear

878 regression model), stationarity (unchanged probability distribution over time), error

879 orthogonality (independence between random error and true signal), and zero error

880 cross-correlation (independence between random errors). Potential error

881 autocorrelation is considered with lag-1 [day] series. Various studies have examined

882 the validity and impact of these assumptions. Numerous studies have examined the

883 validity of these assumptions and their impact on the outcomes if violated (Tsamalis,

884 2022; Duan et al., 2021; Gruber et al., 2020).

885 The linearity assumption shapes the error model by including additive and

886 multiplicative biases and zero-mean random error. Although some studies have

887 explored the application of a nonlinear rescaling technique (Yilmaz and Crow, 2013;

888 Zwieback et al., 2016), those efforts are primarily limited to soil moisture signals and

889 often fail to accurately represent the true signal unless all datasets share a similar

890 signal-to-noise ratio (SNR). However, it is worth noting that after rescaling processes,

891   such as cumulative distribution function (CDF) matching or climatology removal, the

892   resulting time series (anomalies) are often considered linearly related to the truth since

893   higher-order error terms are removed. In addition, multiplicative relationships have

894   been more commonly identified in rainfall products (Li et al., 2018). In contrast,

895   collocation analysis within the context of ET products frequently suggests that linear

896   relationships are reasonable (Li et al., 2022; Park et al., 2023). Therefore, the linear

897   error model remains a robust implementation, though it has the potential for

898   improvement through rescaling techniques.

899   Regarding violating the stationarity assumption, the evapotranspiration signal does

900   not strictly adhere to this characteristic. However, by collocating triplets with similar

901   magnitude variations, the influence of this violation is minimized. Nonetheless,

902   disparities in climatology between datasets can still arise for various reasons (Su and

903   Ryu, 2015). Several proposed alternatives aim to address this issue, such as removing

904   the climatology of inputs (Stoffelen, 1998; Yilmaz and Crow, 2014; Draper et al.,

905   2013) and subsequently analyzing the random error variance of the anomalies (Dong

906   et al., 2020b). Nevertheless, obtaining a reliable estimation of climatology proves

907   challenging in practice.

908   The assumption of error orthogonality assumes independence between random error

909   and true signal, i.e., $\sigma_{\varepsilon_i \Theta} = 0$. A few studies have examined this assumption. Yilmaz

910   and Crow (2014) investigated such violations using four in situ sites and concluded

911   that the impact is negligible since rescaling mitigates or compensates for bias.

912   Additionally, non-orthogonality results in non-zero error cross-correlation (ECC),

913   although the latter is considered more important. Vogelzang et al. (2022) also

914   investigated this violation recently and demonstrated minimal second-order impact.

915   Non-zero ECC conditions introduce more substantial bias in the results compared to

916   other violations mainly due to two reasons: (1) they cannot be mitigated by rescaling;

917   (2) they cannot be compensated even with equal magnitude for all inputs; and (3) they

918   have been frequently reported in recent studies for various variables (Li et al., 2018,

919   2022; Gruber et al., 2016b). Gruber et al. (2016a) proposed the extended collocation

920   method, which effectively addresses the ECC of selected pairs. Moreover, the EIVD

921   method adopts the error cross-correlation framework. In the following section, we

922   will analyze the ECC between pairs.

923   **5.2. Analysis of error cross-correlation**

924   This study assumes non-zero ECC (Error-Correction Coefficient) conditions exist

925   between FluxCom and PMLv2 at 0.1° and between ERA5L and GLEAM at 0.25°.

926   However, non-zero ECC conditions were also possible between other pairs. Therefore,

927   we presented the EIVD-based ECC results of various pairs.

928

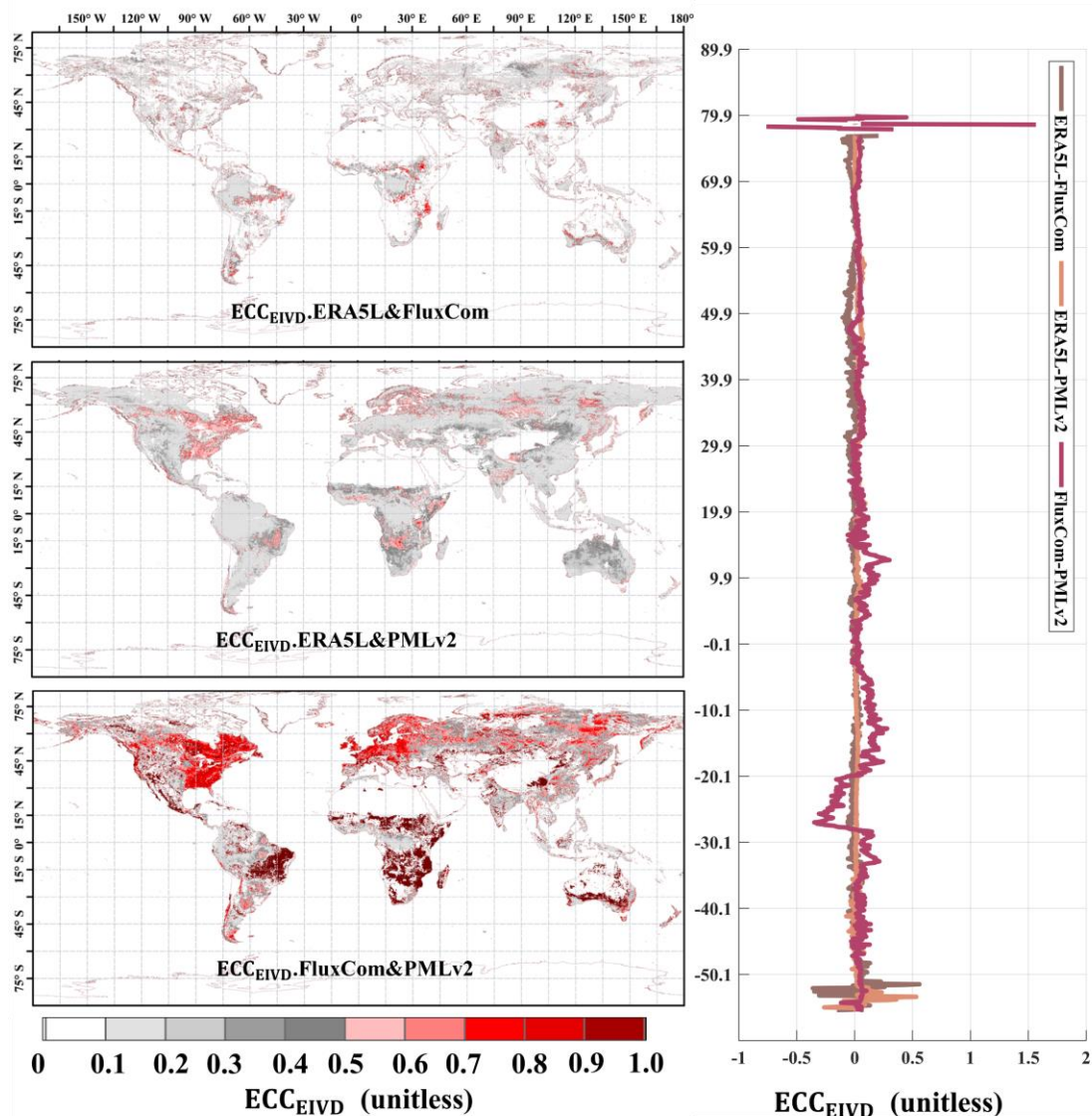



Figure 17 Global Distribution of estimated error cross-correlation (ECC) between ERA5L, FluxCom, and PMLv2 pairwise using EIVD alongside relevant variation curves of average with latitude.

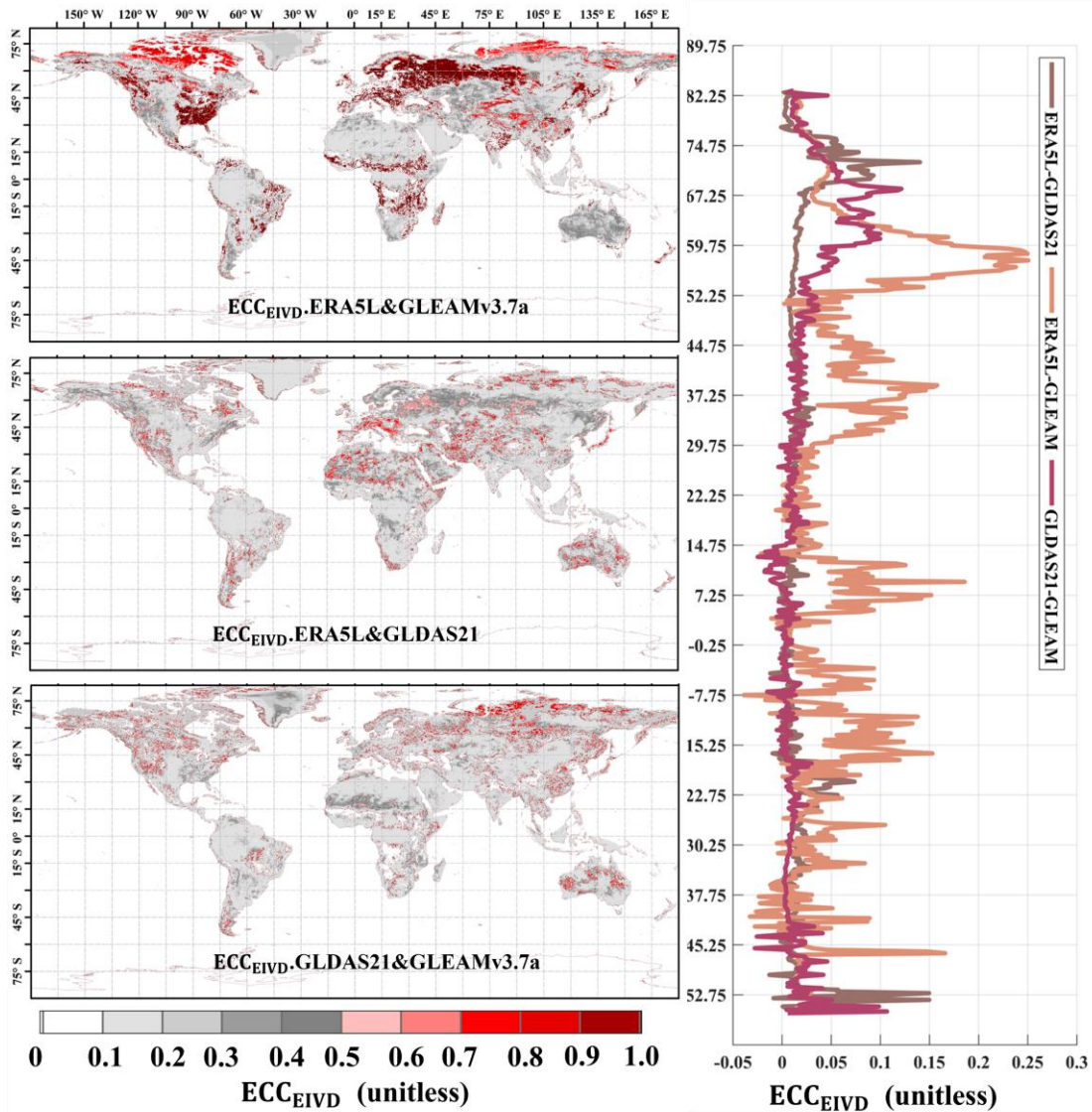



Figure 18 Global Distribution of estimated error cross-correlation (ECC) between ERA5L, GLEAMv3.7a, and GLDAS21 pairwise using EIVD alongside relevant variation curves of average with latitude.

As depicted in **Figure 17** and **Figure 18**, at a resolution of 0.1°, the ECC values of FluxCom and PMLv2 were notably higher than those of ERA5L-FluxCom and ERA5L-PMLv2. The global average ECC value for FluxCom-PMLv2 was 0.16, and

940   regions with high ECC values were identified in the eastern United States, most of

941   Europe, and the western Amazon, areas densely covered by measurement sites. Since

942   both FluxCom and PMLv2 incorporated corrections based on FluxNet measurement

943   sites, there is likely some overlap between the sites used by both products in the high

944   ECC regions. This partially explains the shared source of random errors between the

945   two datasets.

946   The global error correlations of GLEAM-GLDAS and ERA5L-GLDAS are relatively

947   low. The random error of ERA5L correlates with that of GLEAM, primarily in arid

948   regions such as the Sahara Desert, Northwest China, and central Australia, where the

949   average ECC exceeds 0.20. The global average ECC of ERA5L-GLEAM is

950   approximately 0.14. A higher error correlation is observed for ERA5L-GLEAM, with

951   a mean ECC value of 0.26, which is expected since meteorological information from

952   ECMWF is reanalyzed for both datasets. However, ECC values for GLEAM-GLDAS

953   and ERA5L-GLDAS are generally low globally, supporting the assumption of zero

954   ECC for these two pairs.

955   Our findings highlight the significant impact of Error Cross Correlation (ECC)

956   between FluxCom-PMLv2 and ERA5L-GLEAM at 0.1° and 0.25° resolutions,

957   respectively. Mathematically, when a triplet exhibits a high ECC value (>0.3)

958   between two sets, it indicates a preference for the remaining independent product as

959   the "better" one, potentially leading to an underestimation of its error variance.

960   However, it is essential to note that the overall ECC values for other pairs are

961   relatively small, suggesting that the zero ECC assumptions can be considered valid

962   for these pairs across most areas. Therefore, these assumptions are unlikely to affect

963   the relevant results of uncertainties significantly. Nevertheless, we have considered

964   the non-zero ECC condition between FluxCom-PMLv2 and ERA5L-GLEAM in this

965   study, as it requires careful consideration.

5.3. Comparison of different fusion schemes

In this section, we conducted comparisons in three aspects: (1) comparing the performance of CAMELE at different resolutions; (2) comparing the performance of different change fusion schemes, explicitly changing the input products' versions (GLDAS21 to GLDAS20 or GLDAS22, GLEAMv3.7a to v3.7b); and (3) comparing the performance of the results obtained without considering the ECC impact.

We conducted a comprehensive comparison of our fusion approach with several alternative schemes. Specifically, these schemes encompassed utilizing only ERA5L and PMLV2 at 0.1° based on the IVD method (Comb1), changing the versions of GLDAS2 and GLEAM at 0.25° based on the EIVD method (Comb2-5), and two TC fusion approaches at 0.1° and 0.25°, which did not incorporate ECC.

It should be noted that the Comb2 scheme, which includes GLDAS20, covers the period from 1980 to 2014, while the other 0.25° comparison schemes (Comb3-5) span from 2003 to 2022. The combinations based on TC (assuming zero ECC) had the same inputs as CAMELE at both resolutions.

Table 7 Average metrics for CAMELE and other fusion schemes at all sites. The bolded sections indicate the schemes with the best performance in their respective metrics.

Product	RMSE (mm/d)	ubRMSE (mm/d)	MAE (mm/d)	KGE	R
CAMELE (0.1)	**0.83**	**0.71**	**0.64**	**0.57**	**0.71**
CAMELE (0.25)	1.03	0.87	0.75	0.51	0.67
ERA5L+PMLV2 *(Comb1-0.1 | IVD)*	1.13	1.00	0.89	0.46	0.61
ERA5L+GLDAS20+GLEAMv3.7a *(Comb2-0.25 | EIVD)*	1.09	0.89	0.87	0.44	0.66
ERA5L+GLDAS22+GLEAMv3.7a *(Comb3-0.25 | EIVD)*	1.20	0.95	0.94	0.44	0.68
ERA5L+GLDAS22+GLEAMv3.7b *(Comb4-0.25 | EIVD)*	1.19	0.94	0.93	0.44	0.69
ERA5L+GLDAS21+GLEAMv3.7b *(Comb5-0.25 | EIVD)*	1.05	0.90	0.80	0.49	0.69
ERA5L+FluxCom+PMLv2 *(Zero-ECC-0.1 | TC)*	1.06	0.91	0.80	0.46	0.60

ERA5L+GLDAS21+GLEAMv3.7a *(Zero-ECC-0.25 / TC)*	1.26	1.03	0.99	0.39	0.61

984 According to the information in the table, CAMELE (0.1°) results were superior in all

985 indicators. Firstly, when comparing the performance of CAMELE at resolutions of

986 0.1° and 0.25°, it was observed that the fused product performed slightly worse at the

987 0.25° resolution. Additionally, the representative of FluxNet sites at the 0.25°

988 resolution decreased, leading to degraded statistical indicators.

989 At the 0.1° resolution, we conducted a comparison of results obtained by exclusively

990 fusing ERA5-Land and PMLv2. Multiple indicators indicated that this approach did

991 not enhance the accuracy of ET estimates and fell significantly short of the scheme

992 employed in CAMELE. This implies that using only two product sets as input did not

993 allow for effective error analysis through collocation analysis, resulting in suboptimal

994 fusion results. More importantly, the limitation of employing only two datasets

995 prevented us from effectively acquiring error information through collocation analysis

996 (Dong et al., 2020a, 2019). Consequently, we made the strategic decision to ensure

997 the inclusion of three datasets as inputs, facilitating the utilization of the EIVD

998 method and maintaining methodological consistency between the 0.1° and 0.25°

999 resolutions.

1000 Furthermore, when comparing the results of different fusion schemes between

1001 CAMELE and Comb2-5 at the 0.25° resolution, CAMELE performed better regarding

1002 error metrics (RMSE, ubRMSE, MAE). The differences in fitting metrics (KGE, R)

1003 were insignificant, indicating that the choice of fusion scheme primarily affected the

1004 errors of the fusion results. The relatively poorer performance of other fusion schemes

1005 could be due to the lack of consideration for non-zero ECC. For example, non-zero

1006 ECC between GLDAS-2.2 and ERA5L has been reported in a recent study (Li et al.,

1007 2023a).

1008 For the comparative analysis of the GLDAS2.0 and GLDAS2.1 schemes, the usage of

1009 GLDAS2.1 yielded better performance. The GLDAS-2.1 simulation leverages

1010 conditions from the GLDAS-2.0 simulation, with improved models driven by a

1011 combination of datasets. Previous research has demonstrated that GLDAS-2.1 offers

1012 improvements in the regional-scale simulation of hydrological variables compared to

1013 GLDAS-2.0 (Qi et al., 2018, 2020). Consequently, we chose to incorporate GLDAS-

1014 2.1 data for as much of the time series as possible.

1015 Moreover, when comparing the fusion effects with and without considering non-zero

1016 ECC conditions, it was evident that considering ECC information could effectively

1017 improve the performance of the fused product, which further demonstrated the

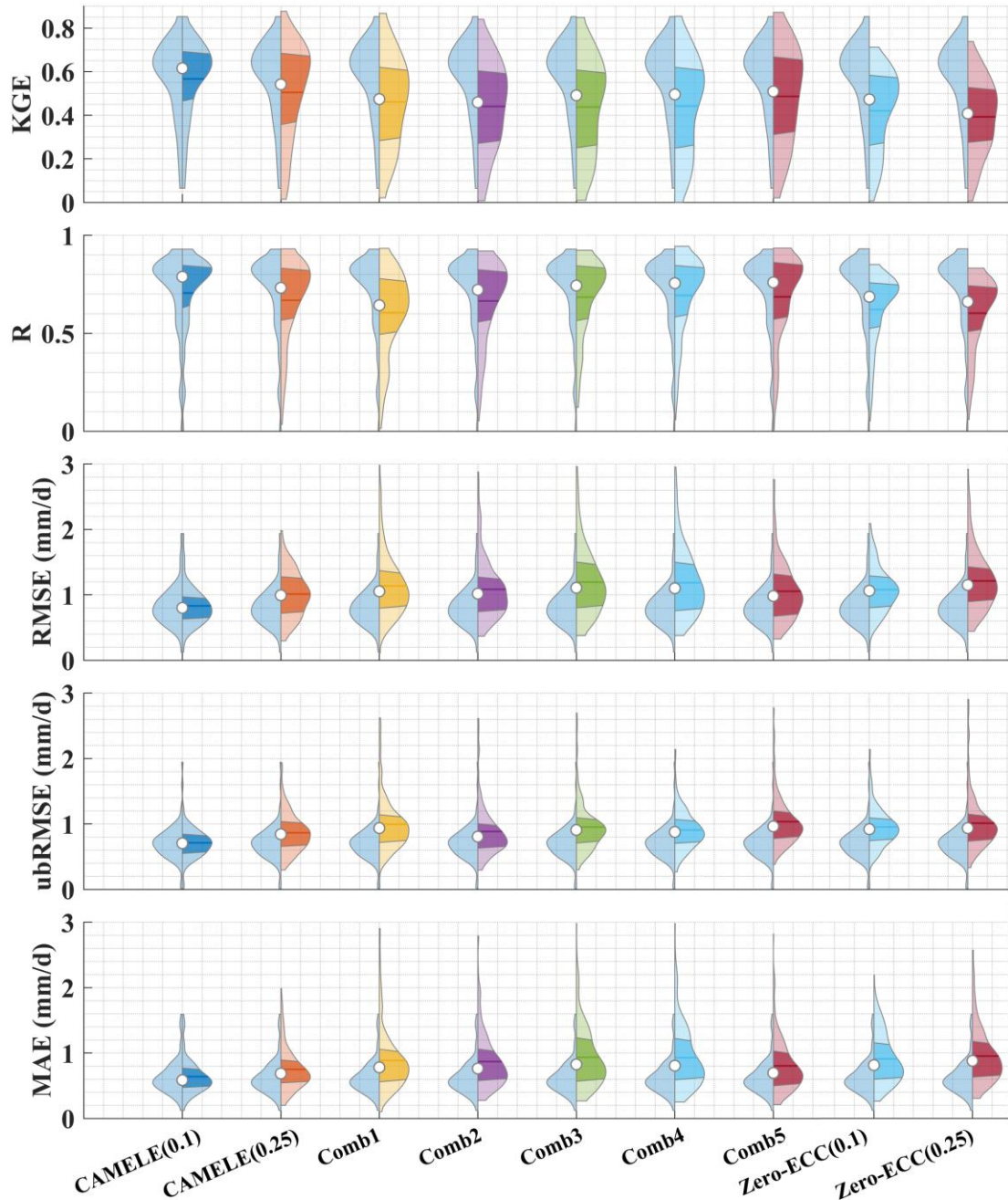
1018 reliability and advantages of the fusion method employed in this study.

1019

Figure 19 Violin plot comparing KGE, R, RMSE, ubRMSE and MAE of CAMELE

with other fusion schemes. The right half of each violin plot represents the

distribution, with shaded areas indicating the box plot, where the horizontal line

corresponds to the median and the dot represents the mean. The left half represents

the results of CAMELE (0.1°) for comparison.

We further provided violin plots for different metrics, comparing the results of each

fusion scheme to CAMELE (0.1°) as shown in **Figure 19**. The results indicated that

1027 the fusion schemes adopted were significantly superior to other schemes based on the

1028 distribution of results for all metrics across all sites. Regarding KGE and R,

1029 CAMELE's results were concentrated near 1 for most sites. Regarding RMSE,

1030 ubRMSE, and MAE, their results were concentrated below one mm/d. The results in

1031 the plots also suggested that CAMELE performed slightly worse at 0.25° compared to

1032 0.1° but still outperformed other combination results. Additionally, comparing

1033 CAMELE and the zero-ECC scheme in the plots further highlighted the importance of

1034 considering non-zero ECC conditions.

1035 **5.4. Potential Applications and Future Enhancements**

1036 In this section, we delve into the potential applications of our product and outline our

1037 commitment to future enhancements to maintain its accuracy and relevance.

1038 Here, we identify three potential applications for our transpiration product: (1) Global

1039 ET Trends: Our product facilitates global-scale analysis of current ET patterns and

1040 long-term trends, essential for comprehending ecosystem responses to evolving

1041 environmental conditions in a warming climate; (2) Transpiration-to-

1042 Evapotranspiration Ratio: Our merging approach can fuse multi-source global gridded

1043 transpiration data, allowing for the examination of the transpiration-to-

1044 evapotranspiration ratio. This analysis can enhance water resource management and

1045 water availability predictions in diverse regions; (3) Attribution analysis: Our product

1046 is a valuable tool for attribution analysis, helping researchers identify the drivers of

1047 patterns. This knowledge is crucial for understanding the roles of climate variability,

1048 land-use changes, and other factors in shaping terrestrial water fluxes.

1049 Furthermore, we are committed to enhancing our product proactively. Key strategies

1050 include: (1) Data Update and Validation: To ensure our product's continued accuracy

1051 and reliability, we will prioritize regularly updating the data used in this study to the

1052 latest versions. By adopting this approach, we aim to provide users with results that

1053 reflect the latest advancements in scientific knowledge; (2) Enhanced Integration and

1054 Error Reduction: We continually refine estimates by incorporating additional data

1055 sources and implementing extended collocation method to minimize errors; (3)

1056 Integration of High-Resolution Regional ET Data: Recognizing the significance of

1057 regional-scale insights, we will focus on improving the accuracy of CAMELE by

1058 integrating higher-resolution regional ET data. This integration will enable more

1059 precise regional estimation.

1060 In summary, these endeavors collectively represent our commitment to maintaining

1061 our product's quality and relevance, ensuring its value for the scientific community.

1062 **6. Conclusion**

1063 This study used a collocation-based approach for merging data considering non-zero

1064 conditions. We successfully generated a long-term daily CAMELE evapotranspiration

1065 (ET) product at resolutions of 0.1° (2000 to 2020) and 0.25° (1980 to 2022) by

1066 integrating five widely used datasets: ERA5L, FluxCom, PMLv2, GLDAS, and

1067 GLEAM. The key findings of our study are as follows:

1068 1. Collocation analysis methods proved to be a reliable tool for evaluating ET

1069   products without a reference dataset. This approach shows promising potential for

1070   error characterization, especially in regions with limited data availability or on a

1071   global scale. The evaluation results provided valuable insights into the data

1072   merging process.

1073 2. Compared to five input products, REA, and simple average, the CAMELE product

1074   performed well when evaluated against FluxNet flux tower data. While CAMELE

1075   may not excel in all individual metrics, it effectively reduces errors associated

1076   with the input products. The result showed Pearson correlation coefficients (R) of

1077   0.63 and 0.65, root-mean-square errors (RMSE) of 0.81 and 0.73 mm/d, unbiased

1078   root-mean-square errors (ubRMSE) of 1.20 and 1.04 mm/d, mean absolute errors

1079   (MAE) of 0.81 and 0.73 mm/d, and Kling-Gupta efficiency (KGE) of 0.60 and

1080   0.65 on average over resolutions of 0.1° and 0.25°, respectively. This robust

1081   performance is especially evident when assessing its comprehensive station-scale

1082   evaluation.

1083   3. For different plant functional types (PFTs), the CAMELE product outperformed

1084   the five input products, REA, and simple average in most PFTs. Although

1085   FluxCom and PMLv2 performed slightly better than CAMELE at some PFT sites,

1086   considering that both utilized FluxNet sites for product calibration, it indirectly

1087   demonstrates the promising and robust performance of CAMELE.

1088   4. Based on site-scale observations, CAMELE effectively captures the multi-year

1089   linear trend of ET. The accuracy of the multi-year mean value depicted by

1090   CAMELE is improved compared to the input data. Moreover, it accurately

1091   characterizes extreme ET values. However, there is a slight overestimation in

1092   representing the seasonality, which needs further improvement in future research.

1093   5. When utilizing the error information derived from collocation analysis for

1094   merging, it is crucial to consider the potential presence of non-zero error

1095   compensation conditions (ECC). Comparing the merging schemes with and

1096   without considering non-zero ECC, it was found that considering ECC improves

1097   the accuracy of the merging process. Additionally, when using collocation

1098   analysis, it is necessary to identify which products may have ECC in advance,

1099   providing more effective support for data merging and obtaining more accurate

1100   product error information.

1101   In conclusion, our proposed collocation-based data merging approach demonstrates

1102   the promising potential for merging ET products. The resulting CAMELE product

1103   exhibited good overall performance at site-based and regional scales, meeting the

1104   requirements for more detailed research. Furthermore, further evaluation of the

1105   merged product in specific regions is necessary to improve its accuracy. In future

1106   studies, dynamic weights could be computed by considering suitable merging periods

1107   for different products to enhance the quality of the merged product, and more

1108   sophisticated combination schemes could be explored to improve accuracy.

Author Contribution

C.L. conceived and designed the study, collected and analyzed the data, and wrote the manuscript. H.Y participated in the study design, provided intellectual insights, and reviewed the manuscript for important intellectual content. Z.L. and W.Y provided substantial input in the study design and data interpretation and revised the manuscript. Z.T., J.H., and S.L. guided the research process and critically reviewed the manuscript. All authors have read and approved the final version of the manuscript.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgments

This research was supported by the China National Key R&D Program (grant no. 2022YFC3002802), the National Natural Science Foundation of China (grant nos. 51979140, 52309022 and 42041004), and the Key Research and Development Program of Yunnan Province, China (grant no. 202203AA080010).

Data and code availability

The datasets utilized in this research can be accessed through the links provided in the Dataset Section. The CAMELE products are available via https://doi.org/10.5281/zenodo.5704736 (Li et al., 2023b). The data is distributed under a Creative Commons Attribution 4.0 License. Additionally, we provide example MATLAB codes to read and plot CAMELE data and employ IVD and EIVD methods to merge the inputs. Please refer to the latest version., 202306.

1131 **Reference**

1132 Baker, J. C. A., Garcia-Carreras, L., Gloor, M., Marsham, J. H., Buermann, W., da
1133 Rocha, H. R., Nobre, A. D., de Araujo, A. C., and Spracklen, D. V.:
1134 Evapotranspiration in the Amazon: spatial patterns, seasonality, and recent trends in
1135 observations, reanalysis, and climate models, Hydrology and Earth System Sciences,
1136 25, 2279–2300, https://doi.org/10.5194/hess-25-2279-2021, 2021.
1137 Barraza Bernadas, V., Grings, F., Restrepo-Coupe, N., and Huete, A.: Comparison of
1138 the performance of latent heat flux products over southern hemisphere forest
1139 ecosystems: estimating latent heat flux error structure using in situ measurements and
1140 the triple collocation method, International Journal of Remote Sensing, 39, 6300–
1141 6315, 2018.
1142 Bates, J. M. and Granger, C. W.: The combination of forecasts, Journal of the
1143 Operational Research Society, 20, 451–468, 1969.
1144 Chen, Z., Zhu, Z., Jiang, H., and Sun, S.: Estimating daily reference
1145 evapotranspiration based on limited meteorological data using deep learning and
1146 classical machine learning methods, Journal of Hydrology, 591,
1147 https://doi.org/10.1016/j.jhydrol.2020.125286, 2020.
1148 De Lannoy, G. J., Houser, P. R., Verhoest, N. E., Pauwels, V. R., and Gish, T. J.:
1149 Upscaling of point soil moisture measurements to field averages at the OPE3 test site,
1150 Journal of Hydrology, 343, 1–11, 2007.
1151 Deng, X., Zhu, L., Wang, H., Zhang, X., Tong, C., Li, S., and Wang, K.: Triple
1152 Collocation Analysis and In Situ Validation of the CYGNSS Soil Moisture Product,
1153 IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,
1154 16, 1883–1899, https://doi.org/10.1109/jstars.2023.3235111, 2023.
1155 Dong, J. and Crow, W. T.: An Improved Triple Collocation Analysis Algorithm for
1156 Decomposing Autocorrelated and White Soil Moisture Retrieval Errors, Journal of
1157 Geophysical Research: Atmospheres, 122, 13,081-13,094,
1158 https://doi.org/10.1002/2017jd027387, 2017.
1159 Dong, J., Crow, W. T., Duan, Z., Wei, L., and Lu, Y.: A double instrumental variable
1160 method for geophysical product error estimation, Remote Sensing of Environment,
1161 225, 217–228, https://doi.org/10.1016/j.rse.2019.03.003, 2019.
1162 Dong, J., Wei, L., Chen, X., Duan, Z., and Lu, Y.: An instrument variable based
1163 algorithm for estimating cross-correlated hydrological remote sensing errors, Journal
1164 of Hydrology, 581, 124413, https://doi.org/10.1016/j.jhydrol.2019.124413, 2020a.
1165 Dong, J., Lei, F., and Wei, L.: Triple Collocation Based Multi-Source Precipitation
1166 Merging, Frontiers in Water, 2, https://doi.org/10.3389/frwa.2020.00001, 2020b.
1167 Dong, J., Crow, W. T., Chen, X., Tangdamrongsub, N., Gao, M., Sun, S., Qiu, J., Wei,
1168 L., Gao, H., and Duan, Z.: Statistical uncertainty analysis-based precipitation merging
1169 (SUPER): A new framework for improved global precipitation estimation, Remote
1170 Sensing of Environment, 283, 113299, https://doi.org/10.1016/j.rse.2022.113299,

1171 2022.

1172 Draper, C., Reichle, R., de Jeu, R., Naeimi, V., Parinussa, R., and Wagner, W.:
1173 Estimating root mean square errors in remotely sensed soil moisture over continental
1174 scale domains, Remote Sensing of Environment, 137, 288–298, 2013.

1175 Duan, Z., Duggan, E., Chen, C., Gao, H., Dong, J., and Liu, J.: Comparison of
1176 traditional method and triple collocation analysis for evaluation of multiple gridded
1177 precipitation products across Germany, Journal of Hydrometeorology,
1178 https://doi.org/10.1175/JHM-D-21-0049.1, 2021.

1179 ECMWF, S. P.: In IFS documentation CY40R1 Part IV: Physical Processes, ECMWF:
1180 Reading, UK, 111–113, 2014.

1181 Ershadi, A., McCabe, M. F., Evans, J. P., Chaney, N. W., and Wood, E. F.: Multi-site
1182 evaluation of terrestrial evaporation models using FLUXNET data, Agricultural and
1183 Forest Meteorology, 187, 46–61, https://doi.org/10.1016/j.agrformet.2013.11.008,
1184 2014.

1185 Feng, Y., Cui, N., Zhao, L., Hu, X., and Gong, D.: Comparison of ELM, GANN,
1186 WNN and empirical models for estimating reference evapotranspiration in humid
1187 region of Southwest China, Journal of Hydrology, 536, 376–383,
1188 https://doi.org/10.1016/j.jhydrol.2016.02.053, 2016.

1189 Gan, R., Zhang, Y., Shi, H., Yang, Y., Eamus, D., Cheng, L., Chiew, F. H., and Yu,
1190 Q.: Use of satellite leaf area index estimating evapotranspiration and gross
1191 assimilation for Australian ecosystems, Ecohydrology, 11, e1974, 2018.

1192 Gentine, P., Massmann, A., Lintner, B. R., Hamed Alemohammad, S., Fu, R., Green,
1193 J. K., Kennedy, D., and Vilà-Guerau de Arellano, J.: Land–atmosphere interactions in
1194 the tropics – a review, Hydrology and Earth System Sciences, 23, 4171–4197,
1195 https://doi.org/10.5194/hess-23-4171-2019, 2019.

1196 Gruber, A., Su, C., Crow, W. T., Zwieback, S., Dorigo, W., and Wagner, W.:
1197 Estimating error cross-correlations in soil moisture data sets using extended
1198 collocation analysis, Journal of Geophysical Research: Atmospheres, 121, 1208–1219,
1199 2016a.

1200 Gruber, A., Su, C.-H., Zwieback, S., Crow, W., Dorigo, W., and Wagner, W.: Recent
1201 advances in (soil moisture) triple collocation analysis, International Journal of
1202 Applied Earth Observation and Geoinformation, 45, 200–211,
1203 https://doi.org/10.1016/j.jag.2015.09.002, 2016b.

1204 Gruber, A., Dorigo, W. A., Crow, W., and Wagner, W.: Triple Collocation-Based
1205 Merging of Satellite Soil Moisture Retrievals, IEEE Trans. Geosci. Remote Sensing,
1206 55, 6780–6792, https://doi.org/10.1109/TGRS.2017.2734070, 2017.

1207 Gruber, A., Scanlon, T., van der Schalie, R., Wagner, W., and Dorigo, W.: Evolution
1208 of the ESA CCI Soil Moisture climate data records and their underlying merging
1209 methodology, Earth System Science Data, 11, 717–739, https://doi.org/10.5194/essd-
1210 11-717-2019, 2019.

1211 Gruber, A., De Lannoy, G., Albergel, C., Al-Yaari, A., Brocca, L., Calvet, J. C.,
1212 Colliander, A., Cosh, M., Crow, W., Dorigo, W., Draper, C., Hirschi, M., Kerr, Y.,

1213 Konings, A., Lahoz, W., McColl, K., Montzka, C., Muñoz-Sabater, J., Peng, J.,
1214 Reichle, R., Richaume, P., Rüdiger, C., Scanlon, T., van der Schalie, R., Wigneron, J.
1215 P., and Wagner, W.: Validation practices for satellite soil moisture retrievals: What
1216 are (the) errors?, Remote Sensing of Environment, 244, 111806,
1217 https://doi.org/10.1016/j.rse.2020.111806, 2020.

1218 Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the
1219 mean squared error and NSE performance criteria: Implications for improving
1220 hydrological modelling, Journal of hydrology, 377, 80–91, 2009.

1221 Han, S. and Tian, F.: A review of the complementary principle of evaporation: from
1222 the original linear relationship to generalized nonlinear functions, Hydrology and
1223 Earth System Sciences, 24, 2269–2285, https://doi.org/10.5194/hess-24-2269-2020,
1224 2020.

1225 Hao, Y., Baik, J., and Choi, M.: Combining generalized complementary relationship
1226 models with the Bayesian Model Averaging method to estimate actual
1227 evapotranspiration over China, Agricultural and Forest Meteorology, 279, 107759,
1228 https://doi.org/10.1016/j.agrformet.2019.107759, 2019.

1229 Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J.,
1230 Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S.,
1231 Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., Chiara,
1232 G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R.,
1233 Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková,
1234 M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., Rosnay, P., Rozum, I.,
1235 Vamborg, F., Villaume, S., and Thépaut, J.: The ERA5 global reanalysis, Quarterly
1236 Journal of the Royal Meteorological Society, 146, 1999–2049,
1237 https://doi.org/10.1002/qj.3803, 2020.

1238 Hoareau, N., Portabella, M., Lin, W., Ballabrera-Poy, J., and Turiel, A.: Error
1239 characterization of sea surface salinity products using triple collocation analysis,
1240 IEEE Transactions on Geoscience and Remote Sensing, 56, 5160–5168, 2018.

1241 Jia, Y., Li, C., Yang, H., Yang, W., and Liu, Z.: Assessments of three
1242 evapotranspiration products over China using extended triple collocation and water
1243 balance methods, Journal of Hydrology, 614, 128594, 2022.

1244 Jiang, C. and Ryu, Y.: Multi-scale evaluation of global gross primary productivity and
1245 evapotranspiration products derived from Breathing Earth System Simulator (BESS),
1246 Remote Sensing of Environment, 186, 528–547, 2016.

1247 Jiang, C., Ryu, Y., Fang, H., Myneni, R., Claverie, M., and Zhu, Z.: Inconsistencies of
1248 interannual variability and trends in long-term satellite leaf area index products, Glob
1249 Chang Biol, 23, 4133–4146, https://doi.org/10.1111/gcb.13787, 2017.

1250 Jiang, C., Guan, K., Pan, M., Ryu, Y., Peng, B., and Wang, S.: BESS-STAIR: a
1251 framework to estimate daily, 30 m, and all-weather crop evapotranspiration using
1252 multi-source satellite data for the US Corn Belt, Hydrology and Earth System
1253 Sciences, 24, 1251–1273, https://doi.org/10.5194/hess-24-1251-2020, 2020.

1254 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., and Wang, K.: Global intercomparison of 12 land surface heat flux estimates, J. Geophys. Res., 116, D02102, https://doi.org/10.1029/2010JD014545, 2011.

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, Sci Data, 6, 74, https://doi.org/10.1038/s41597-019-0076-8, 2019.

Khan, M. S., Liaqat, U. W., Baik, J., and Choi, M.: Stand-alone uncertainty characterization of GLEAM, GLDAS and MOD16 evapotranspiration products using an extended triple collocation approach, Agricultural and Forest Meteorology, 252, 256–268, https://doi.org/10.1016/j.agrformet.2018.01.022, 2018.

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 Double Instrumental Variable (EIVD) Alternatives, IEEE Transactions on Geoscience and Remote Sensing, 59, 7285–7295, https://doi.org/10.1109/tgrs.2020.3032418, 2021a.

Kim, S., Sharma, A., Liu, Y. Y., and Young, S. I.: Rethinking satellite data merging: from averaging to SNR optimization, IEEE Transactions on Geoscience and Remote Sensing, 60, 1–15, 2021b.

Kling, H., Fuchs, M., and Paulin, M.: Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios, Journal of Hydrology, 424–425, 264–277, https://doi.org/10.1016/j.jhydrol.2012.01.011, 2012.

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, https://doi.org/10.5194/hess-23-4323-2019, 2019.

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, https://doi.org/10.1029/2021wr030330, 2021.

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.

Leuning, R., Zhang, Y. Q., Rajaud, A., Cleugh, H., and Tu, K.: Correction to "A simple surface conductance model to estimate regional evaporation using MODIS leaf area index and the Penman-Monteith equation," Water Resources Research, 45, https://doi.org/10.1029/2008wr007631, 2009.

Li, B., Rodell, M., Kumar, S., Beaudoing, H. K., Getirana, A., Zaitchik, B. F., Goncalves, L. G., Cossetin, C., Bhanja, S., Mukherjee, A., Tian, S., Tangdamrongsub, N., Long, D., Nanteza, J., Lee, J., Policelli, F., Goni, I. B., Daira, D., Bila, M., Lannoy, G., Mocko, D., Steele-Dunne, S. C., Save, H., and Bettadpur, S.: Global GRACE Data Assimilation for Groundwater and Drought Monitoring: Advances and Challenges,

Water Resour. Res., 55, 7564–7586, https://doi.org/10.1029/2018WR024618, 2019a.

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, https://doi.org/10.1016/j.jhydrol.2018.04.039, 2018.

Li, C., Yang, H., Yang, W., Liu, Z., Jia, Y., Li, S., and Yang, D.: Error Characterization of Global Land Evapotranspiration Products: Collocation-based approach, Journal of Hydrology, 128102, 2022.

Li, C., Liu, Z., Tu, Z., Shen, J., He, Y., and Yang, H.: Assessment of global gridded transpiration products using the extended instrumental variable technique (EIVD), Journal of Hydrology, 623, 129880, https://doi.org/10.1016/j.jhydrol.2023.129880, 2023a.

Li, C., Liu, Z., Yang, W., Tu, Z., Han, J., Sien, L., and Hanbo, Y.: CAMELE: Collocation-Analyzed Multi-source Ensembled Land Evapotranspiration Data, , https://doi.org/10.5281/zenodo.8047038, 2023b.

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, https://doi.org/10.1016/j.agrformet.2018.11.017, 2019b.

Li, X., Zhang, W., Vermeulen, A., Dong, J., and Duan, Z.: Triple collocation-based merging of multi-source gridded evapotranspiration data in the Nordic Region, Agricultural and Forest Meteorology, 335, 109451, https://doi.org/10.1016/j.agrformet.2023.109451, 2023c.

Lian, X., Piao, S., Huntingford, C., Li, Y., Zeng, Z., Wang, X., Ciais, P., McVicar, T. R., Peng, S., Ottlé, C., Yang, H., Yang, Y., Zhang, Y., and Wang, T.: Partitioning global land evapotranspiration using CMIP5 models constrained by observations, Nature Clim Change, 8, 640–646, https://doi.org/10.1038/s41558-018-0207-9, 2018.

Lin, C., Gentine, P., Huang, Y., Guan, K., Kimm, H., and Zhou, S.: Diel ecosystem conductance response to vapor pressure deficit is suboptimal and independent of soil moisture, Agricultural and Forest Meteorology, 250, 24–34, 2018.

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.

Lu, J., Wang, G., Chen, T., Li, S., Hagan, D. F. T., Kattel, G., Peng, J., Jiang, T., and Su, B.: A harmonized global land evaporation dataset from model-based products covering 1980–2017, Earth System Science Data, 13, 5879–5898, https://doi.org/10.5194/essd-13-5879-2021, 2021.

Ma, N., Szilagyi, J., and Jozsa, J.: Benchmarking large-scale evapotranspiration estimates: A perspective from a calibration-free complementary relationship approach and FLUXCOM, Journal of Hydrology, 590, https://doi.org/10.1016/j.jhydrol.2020.125221, 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, https://doi.org/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, https://doi.org/10.5194/gmd-10-1903-2017, 2017.

McColl, K. A., Vogelzang, J., Konings, A. G., Entekhabi, D., Piles, M., and Stoffelen, A.: Extended triple collocation: Estimating errors and correlation coefficients with respect to an unknown target, Geophysical Research Letters, 41, 6229–6236, https://doi.org/10.1002/2014gl061322, 2014.

Ming, W., Ji, X., Zhang, M., Li, Y., Liu, C., Wang, Y., and Li, J.: A Hybrid Triple Collocation-Deep Learning Approach for Improving Soil Moisture Estimation from Satellite and Model-Based Data, Remote Sensing, 14, https://doi.org/10.3390/rs14071744, 2022.

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

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, Ann N Y Acad Sci, 1436, 19–35, https://doi.org/10.1111/nyas.13912, 2019.

Mu, Q., Zhao, M., and Running, S. W.: Improvements to a MODIS global terrestrial evapotranspiration algorithm, Remote sensing of environment, 115, 1781–1800, 2011.

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles, D. G., Piles, M., Rodríguez-Fernández, N. J., Zsoter, E., Buontempo, C., and Thépaut, J.-N.: ERA5-Land: a state-of-the-art global reanalysis dataset for land applications, Earth Syst. Sci. Data, 13, 4349–4383, https://doi.org/10.5194/essd-13-4349-2021, 2021.

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, https://doi.org/10.5194/hess-24-1485-2020, 2020.

Park, J., Baik, J., and Choi, M.: Triple collocation-based multi-source evaporation and transpiration merging, Agricultural and Forest Meteorology, 331, 109353, 2023.

Pastorello, G., Trotta, C., Canfora, E., Chu, H., Christianson, D., Cheah, Y. W., Poindexter, C., Chen, J., Elbashandy, A., Humphrey, M., Isaac, P., Polidori, D., Reichstein, M., Ribeca, A., van Ingen, C., Vuichard, N., Zhang, L., Amiro, B.,

Ammann, C., Arain, M. A., Ardo, J., Arkebauer, T., Arndt, S. K., Arriga, N., Aubinet, M., Aurela, M., Baldocchi, D., Barr, A., Beamesderfer, E., Marchesini, L. B., Bergeron, O., Beringer, J., Bernhofer, C., Berveiller, D., Billesbach, D., Black, T. A., Blanken, P. D., Bohrer, G., Boike, J., Bolstad, P. V., Bonal, D., Bonnefond, J. M., Bowling, D. R., Bracho, R., Brodeur, J., Brummer, C., Buchmann, N., Burban, B., Burns, S. P., Buysse, P., Cale, P., Cavagna, M., Cellier, P., Chen, S., Chini, I., Christensen, T. R., Cleverly, J., Collalti, A., Consalvo, C., Cook, B. D., Cook, D., Coursolle, C., Cremonese, E., Curtis, P. S., D'Andrea, E., da Rocha, H., Dai, X., Davis, K. J., Cinti, B., Grandcourt, A., Ligne, A., De Oliveira, R. C., Delpierre, N., Desai, A. R., Di Bella, C. M., Tommasi, P. D., Dolman, H., Domingo, F., Dong, G., Dore, S., Duce, P., Dufrene, E., Dunn, A., Dusek, J., Eamus, D., Eichelmann, U., ElKhidir, H. A. M., Eugster, W., Ewenz, C. M., Ewers, B., Famulari, D., Fares, S., Feigenwinter, I., Feitz, A., Fensholt, R., Filippa, G., Fischer, M., Frank, J., Galvagno, M., et al.: The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data, Sci Data, 7, 225, https://doi.org/10.1038/s41597-020-0534-3, 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.

Qi, W., Liu, J., and Chen, D.: Evaluations and Improvements of GLDAS2.0 and GLDAS2.1 Forcing Data's Applicability for Basin Scale Hydrological Simulations in the Tibetan Plateau, JGR Atmospheres, 123, https://doi.org/10.1029/2018JD029116, 2018.

Qi, W., Liu, J., Yang, H., Zhu, X., Tian, Y., Jiang, X., Huang, X., and Feng, L.: Large Uncertainties in Runoff Estimations of GLDAS Versions 2.0 and 2.1 in China, Earth and Space Science, 7, e2019EA000829, https://doi.org/10.1029/2019EA000829, 2020.

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, https://doi.org/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, https://doi.org/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.

Sheffield, J., Goteti, G., and Wood, E. F.: Development of a 50-Year High-Resolution Global Dataset of Meteorological Forcings for Land Surface Modeling, Journal of Climate, 19, 3088–3111, https://doi.org/10.1175/JCLI3790.1, 2006.

Stoffelen, A.: Toward the true near-surface wind speed: Error modeling and calibration using triple collocation, Journal of Geophysical Research: Oceans, 103,

7755–7766, https://doi.org/10.1029/97jc03180, 1998.

Su, C.-H. and Ryu, D.: Multi-scale analysis of bias correction of soil moisture, Hydrology and Earth System Sciences, 19, 17–31, 2015.

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, https://doi.org/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, https://doi.org/10.1016/j.rse.2020.112146, 2021.

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, https://doi.org/10.5194/hess-23-3057-2019, 2019.

Tsamalis, C.: Clarifications on the equations and the sample number in triple collocation analysis using SST observations, Remote Sensing of Environment, 272, https://doi.org/10.1016/j.rse.2022.112936, 2022.

Twine, T. E., Kustas, W., Norman, J., Cook, D., Houser, Pr., 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., Stoffelen, A., and Verhoef, A.: The Effect of Error Non-Orthogonality on Triple Collocation Analyses, Remote Sensing, 14, 4268, 2022.

Wu, K., Ryu, D., Nie, L., and Shu, H.: Time-variant error characterization of SMAP and ASCAT soil moisture using Triple Collocation Analysis, Remote Sensing of Environment, 256, https://doi.org/10.1016/j.rse.2021.112324, 2021.

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

Yilmaz, M. T. and Crow, W. T.: The optimality of potential rescaling approaches in land data assimilation, Journal of Hydrometeorology, 14, 650–660, 2013.

Yilmaz, M. T. and Crow, W. T.: Evaluation of Assumptions in Soil Moisture Triple Collocation Analysis, Journal of Hydrometeorology, 15, 1293–1302, https://doi.org/10.1175/JHM-D-13-0158.1, 2014.

Yilmaz, M. T., Crow, W. T., Anderson, M. C., and Hain, C.: An objective methodology for merging satellite- and model-based soil moisture products, Water Resources Research, 48, n/a-n/a, https://doi.org/10.1029/2011wr011682, 2012.

Yin, G. and Park, J.: The use of triple collocation approach to merge satellite- and model-based terrestrial water storage for flood potential analysis, Journal of Hydrology, 603, https://doi.org/10.1016/j.jhydrol.2021.127197, 2021.

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

1465 Hydrology, 600, https://doi.org/10.1016/j.jhydrol.2021.126538, 2021.

1466 Zhang, Y., Leuning, R., Hutley, L. B., Beringer, J., McHugh, I., and Walker, J. P.:
1467 Using long-term water balances to parameterize surface conductances and calculate
1468 evaporation at 0.05 spatial resolution, Water Resources Research, 46, 2010.

1469 Zhang, Y., Kong, D., Gan, R., Chiew, F. H. S., McVicar, T. R., Zhang, Q., and Yang,
1470 Y.: Coupled estimation of 500 m and 8-day resolution global evapotranspiration and
1471 gross primary production in 2002–2017, Remote Sensing of Environment, 222, 165–
1472 182, https://doi.org/10.1016/j.rse.2018.12.031, 2019.

1473 Zhao, M., Liu, Y., and Konings, A. G.: Evapotranspiration frequently increases during
1474 droughts, Nature Climate Change, 1–7, 2022.

1475 Zhu, G., Li, X., Zhang, K., Ding, Z., Han, T., Ma, J., Huang, C., He, J., and Ma, T.:
1476 Multi-model ensemble prediction of terrestrial evapotranspiration across north China
1477 using Bayesian model averaging, Hydrological Processes, 30, 2861–2879,
1478 https://doi.org/10.1002/hyp.10832, 2016.

1479 Zwieback, S., Su, C.-H., Gruber, A., Dorigo, W. A., and Wagner, W.: The impact of
1480 quadratic nonlinear relations between soil moisture products on uncertainty estimates
1481 from triple collocation analysis and two quadratic extensions, Journal of
1482 Hydrometeorology, 17, 1725–1743, 2016.

1483