Simulating carbon and water fluxes using a coupled process-based terrestrial biosphere model and joint assimilation of leaf area index and surface soil moisture

Sinan Li [1,2], Li Zhang [1,3,*], Jingfeng Xiao [4], Rui Ma [5], Xiangjun Tian [6], Min Yan[1,3]

[1] Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, No. 9 Dengzhuang South Road, Beijing 100094, China.

[2] College of Resources and Environment, University of Chinese Academy of Sciences, No. 19A Yuquan Road, Beijing 100049, China

[3] Key Laboratory of Earth Observation of Hainan Province, Sanya 572029, China

[4] Earth Systems Research Center, Institute for the Study of Earth, Oceans, and Space, University of New Hampshire, Durham, New Hampshire 03824, USA

[5] School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China

[6] International Center for Climate and Environment Sciences (ICCES), Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China

* Correspondence: zhangli@aircas.ac.cn; Tel.: +86-10-8217-8193

Abstract:

21   Reliable modeling of carbon and water fluxes is essential for understanding the terrestrial carbon

22   and water cycles and informing policy strategies aimed at constraining carbon emissions and improving

23   water use efficiency. We used an assimilation framework (LPJ-Vegetation and soil moisture Joint

24   Assimilation, or LPJ-VSJA) to improve gross primary production (GPP) and evapotranspiration (ET)

25   estimates globally. The terrestrial biosphere model that we used is the integrated model - LPJ-PM coupled

26   from the Lund-Potsdam-Jena Dynamic Global Vegetation Model (LPJ-DGVM) and a hydrology module

27   (i.e., the updated Priestley–Taylor Jet Propulsion Laboratory model, PT-JPL$_{SM}$). Satellite-based soil

28   moisture products derived from the Soil Moisture and Ocean Salinity (SMOS) and Soil Moisture Active

29   and Passive (SMAP) and leaf area index (LAI) from the global Land and Ground satellite (GLASS)

30   product were assimilated into LPJ-PM to improve GPP and ET simulations using a Proper Orthogonal

31   Decomposition-based ensemble four-dimensional variational assimilation method (PODEn4DVar). The

32   joint assimilation framework LPJ-VSJA achieved the best model performance (with an R^2 of 0.91and

33   0.81 and an RMSD reduced by 50.4% and 38.4% for GPP and ET, respectively, compared with those of

34   LPJ-DGVM at the monthly scale). The assimilated GPP and ET demonstrated a better performance in the

35   arid and semiarid regions (GPP: R^2=0.73, ubRMSD=1.05 g C m^{-2} d^{-1}; ET: R^2=0.73, ubRMSD= 0.61 mm

36   d^{-1}) than in the humid and sub-dry humid regions (GPP: R^2=0.61, ubRMSD=1.23 g C m^{-2} d^{-1}; ET: R^2=0.66;

37   ubRMSD=0.67 mm d^{-1}). The ET simulated by LPJ-PM that assimilated SMAP or SMOS had a slight

38   difference, and the ET that assimilated SMAP soil moisture data was more improved than that assimilated

39   SMOS data. Our global simulation modeled by LPJ-VSJA was compared with several global GPP and

40   ET products (e.g., GLASS GPP, GOSIF GPP, GLDAS ET, GLEAM ET) using the triple collocation (TC)

41   method. Our products, especially ET, exhibited advantages in the overall error distribution (estimated

42   error (μ): 3.4 mm month^{-1}; estimated standard deviation of μ: 1.91 mm month^{-1}). Our research showed

43   that the assimilation of multiple datasets could reduce model uncertainties, while the model performance

44   differed across regions and plant functional types. Our assimilation framework (LPJ-VSJA) can improve

45   the model simulation performance of daily GPP and ET globally, especially in water-limited regions.

46   **Keywords:** Data Assimilation; SMOS; SMAP; Gross primary production (GPP); evapotranspiration

47   (ET); GLASS

1. Introduction

50   Gross primary production (GPP) and evapotranspiration (ET) are essential components of the carbon

51   and water cycles, respectively. Carbon and water fluxes are inherently coupled on multiple spatial and

52   temporal scales (Law et al. 2002; Sun et al. 2019; Waring and Running 2010). Land surface models

53   (LSMs) are the most sophisticated approach for providing a relatively detailed description of such

54   interdependent relationships regarding water and carbon fluxes and understanding the response of

55   terrestrial ecosystems to changes in atmospheric CO_2 and climate (Kaminski et al. 2017). However, there

56   are still large uncertainties in carbon and water flux estimates at regional to global scales. Both diagnostic

57   and prognostic models show substantial differences in the magnitude and spatiotemporal patterns of GPP

58   and ET. For example, the global annual GPP estimates exhibited a large range (130–169 Pg C yr^{-1}) among

59 16 Earth System Models (Anav et al. 2015). The global ET ranged from 70,000 to 75,000 km^3 yr^{-1}, and

60 the uncertainty of regional or global ET estimates was up to 50% of the annual mean ET value, especially

61 in the semi-arid regions (Miralles et al. 2016). These uncertainties mainly arise from the forcing datasets,

62 simplification of mechanisms or imperfect assumptions in processes, and uncertain parameters in the

63 LSMs and assimilation methods (Xiao et al. 2019).

64   In the last two decades, remote sensing products have been assimilated into LSMs to reduce the

65 uncertainty in modeled carbon and water fluxes (MacBean et al. 2016). Data assimilation (DA) is an

66 effective approach to reduce uncertainties in terrestrial biosphere models by integrating satellite products

67 with models to constrain related parameters or state variables. A DA system contains three main

68 components: a set of observations, an observation operator, and an assimilation method. The assimilation

69 method considers the errors from both LSMs and observations and reduces model uncertainties by

70 minimizing a cost function. The Ensemble Kalman Filter (EnKF) has been widely applied in land surface

71 process models for parameter optimization, and can also significantly improve simulations by periodically

72 updating state variables (e.g., LAI and soil moisture) and without altering the model's structure based on

73 remote sensing observations (Ines et al. 2013; Li et al. 2017; Ma et al. 2013). However, ENKF cannot

74 obtain the dynamic balance of the estimation in the time window like the four-dimensional variational

75 method (4DVar) assimilation method. In particular, the Proper Orthogonal Decomposition (POD)-based

76 ensemble 4DVAR assimilation method (referred to as PODEn4DVar) (Tian and Feng 2015) requires

77 relatively less computation and can simultaneously assimilate the observations at different time intervals.

78 Meanwhile, it maintains the structural information of the four-dimensional space. This method has a

79 satisfactory performance in land DA (Tian et al. 2019; Tian et al. 2010) and can better estimate GPP and

80 ET than ENKF (Ma et al. 2017).

81   Multiple sources of remote sensing data streams have been used to constrain models for assimilation.

82 As a critical biophysical parameter of the land, leaf area index (LAI) is closely related to many land

83 processes, such as photosynthesis, respiration, precipitation interception, ET, and surface energy

84 exchange (Fang et al. 2019). More accurate LAI estimates can improve the simulations of the carbon and

85 water fluxes (Bonan et al. 2014; Liu et al. 2018; Mu et al. 2007). Soil moisture (SM) controls the process

86 of surface infiltration and runoff, determines the amount of water that can be extracted from plant roots,

87 and distributes the input energy into sensible heat flux and latent heat flux (Trugman et al. 2018). More

88 accurate SM data can improve the simulation of hydrologic parameters (SM, Streamflow, etc.) in regions

89 of interest or on a global scale (e.g., (Brocca et al. 2012; Draper et al. 2011; Lee et al. 2011; Li and Rodell

90 2013). Soil moisture is a major driving factor affecting vegetation production in arid ecosystems,

91 especially, in semi-arid areas (Liu et al. 2020). Introducing surface soil moisture (SSM) into the model

92 can significantly improve GPP and ET simulation, particularly in water-limited areas (He et al. 2017; Li

93 et al. 2020).

94   The advancement of earth observation, machine learning, inversion algorithms, and computer

95 technology has improved the accuracy of global LAI products and boosted model-data fusion studies

96 (Fang et al. 2019; Kganyago et al. 2020; Xiao et al. 2017). The Advanced Very High-Resolution

97 Radiometer (AVHRR) generates global LAI products with the longest historic record (since the early

98 1980s). The GLASS LAI product has been verified to have a better accuracy than that of MODIS and

99 CYCLOPES and is more temporally continuous and spatially complete (Xiao et al. 2013). Several recent

100 studies showed that the assimilation of GLASS LAI into terrestrial biosphere models enhanced the

101 performance of the models in simulating carbon cycling (e.g., GPP, net ecosystem exchange or NEE) and

102 hydrological (e.g., ET, SM) processes (Ling et al. 2019; Ma et al. 2017; Yan et al. 2016).

103   Microwave remote sensors are considered effective tools for measuring SM globally (Petropoulos et

104 al. 2015). For example, surface SM products have been derived from the Soil Moisture and Ocean Salinity

105 (SMOS) and Soil Moisture Active and Passive (SMAP) satellites equipped with an L-band microwave

106 instrument. The products from these satellites have been evaluated against in-situ observations and other

107 SM products and overall have high accuracy(Burgin et al. 2017; Cui et al. 2018). Additionally, the SMAP

108 performed better than SMOS and other SM products (e.g., Advanced Scatterometer (ASCAT), Advanced

109 Microwave Scanning Radiometer 2 (AMSR2)) with an overall lower error and a higher correlation based

110 on the verification with in-situ SM data from 231 sites (Cui et al. 2018; Kim et al. 2018). The assimilation

111 of SMAP data can improve the simulation accuracy of carbon and water fluxes (He et al. 2017; Li et al.

112 2020) and hydrological variables (surface soil moisture, root-zoon soil moisture, and streamflow)

113 (Blyverket et al. 2019; Koster et al. 2018; Reichle et al. 2017) on a regional or global scale. For example,

114 one recent study showed that the assimilation of SMAP data performed slightly better than that of SMOS

115 and ESA CCI data (Blyverket et al. 2019).

116 In the nonlinear model or nonlinear observation operator, only simultaneous assimilation makes

117 optimal use of observations (MacBean et al. 2016). Therefore, a joint assimilation of soil moisture and

118 LAI can make use of information on these two variables separately and their interactions. For instance,

119 Ines et al. (2013) found that the joint assimilation of soil moisture and LAI achieved better results than

120 the assimilation of one of the two variables. Bonan et al. (2020) assimilated LAI and SM together into

121 the Interactions between Soil, Biosphere and Atmosphere (ISBA) land model and improved the modeled

122 GPP, ET, and runoff in the Mediterranean region. However, no studies have examined how the

123 assimilation of LAI and SM into terrestrial biosphere models can improve the simulation of carbon and

124 water fluxes at the global scale and how the performance vary across different regions.

125 In this study, we assimilated the SM products from both SMAP and SMOS and a LAI product

126 (GLASS) derived from AVHRR and MODIS into the LPJ-PM to improve the GPP and ET using the

127 PODEn4DVar method. We designed a new assimilation framework, LPJ-VSJA, which consists of three

128 assimilation schemes: (1) assimilation of LAI only, (2) assimilation of SM only, and (3) joint assimilation

129 of LAI and SM. Then, four experiments were performed based on global flux sites to evaluate the

130 assimilation performance. The specific objectives of our study are to: (1) evaluate the assimilation

131 performance of three assimilation schemes for estimating GPP and ET to select the optimal assimilation

132 scheme; (2) compare the assimilation effect of the optimal assimilation scheme in humid and arid areas;

133 (3) compare the assimilation capability in ET assimilation based on two microwave soil moisture products

134 (SMOS and SMAP) (scheme 2); and (4) apply our LPJ-VSJA framework to the globe to evaluate its

135 accuracy and robustness.

136 **2. LPJ-VSJA framework and assimilation strategy**

137 *2.1. Coupled- model (LPJ-PM) for assimilation*

138  In this study, a coupled terrestrial biosphere model, LPJ-PM, was used to simulate daily GPP and

139 ET by assimilating satellite-derived LAI and SM. The LPJ-PM is coupled from LPJ-DGVM and PT-

140 JPL$_{SM}$. The original input data in PT-JPL$_{SM}$ were all inherited from LPJ-DGVM, with the exception of

141 relative humidity (RH) and surface soil moisture (SMOS and SMAP), including the initial LAI calculated

142 by the LPJ-DGVM or assimilated LAI obtained through the LAI assimilation scheme, canopy height, and

143 the fraction of absorbed photosynthetic effective radiation (fAPAR). The detailed processes of the LPJ-

144 PM have been described in Li et al. (2020), and the flow chart for the coupling is shown in Figure 1.

145        **Table 1. Description of the models and outputs in this study**

acronyms	Full name	Description	Output
LPJ-DGVM	Lund-Potsdam-Jena Dynamic Global Vegetation Model	This model is used as a model operator to simulated initial ET	GPP$_{LPJ}$, ET$_{LPJ}$
PT-JPL$_{SM}$	Updated Priestley–Taylor Jet Propulsion Laboratory model	The model is used as a module of the LPJ-PM and establishes a connection between SMAP SM and ET	N/A
LPJ-PM	Lund-Potsdam-Jena coupled model	An integrated model coupled from the PT-JPL$_{SM}$ and LPJ-DGVM	GPP$_{SM}$, ET$_{PM}$
LPJ-VSJA	Lund-Potsdam-Jena Vegetation-Soil	A process-based assimilation framework for assimilating LAI	GPP$_{LAI}$, ET$_{LAI}$; GPP$_{SM}$, ET$_{SM}$; GPP$_{CO}$; ET$_{CO}$

moisture-Joint -   and SSM jointly into LPJ-DGVM

Assimilation system

146

Due to the limitation of soil stratification in LPJ-DGVM, the model was coupled with an updated remote-sensing ET algorithm in the PT-JPL$_{SM}$ that could better simulate ET in water-limited regions than in humid regions (Purdy et al. 2018). The LPJ-DGVM is a process-oriented dynamic model, which considers mutual interaction of carbon and water cycling and is designed to simulate vegetation distribution and carbon, soil and atmosphere fluxes (Sitch et al. 2003). For each plant functional type (PFT), the GPP is calculated by implementing coupled photosynthesis and water balance. In the water cycle module, ET is calculated as the minimum of a plant- and soil-limited supply function (E$_{supply}$) and the atmospheric demand (E$_{demand}$) (Haxeltine and Prentice 1996; Sitch et al. 2003). The soil structure is simplified to a "two-layer bucket" model (the top soil layer at a 0-50 cm depth and the bottom layer at a 50-100 cm depth), and therefore the satellite-derived surface SM cannot be assimilated into LPJ-DGVM directly. The oversimplified soil structure and single soil moisture limitation inevitably lead to sizeable uncertainty in ET simulation. Additionally, the monthly input caused a daily variation of the modeled SM, which was also not transmitted to the calculation of GPP and ET. Thus, the updated PT-JPL model (hereafter referred to as PT-JPL$_{SM}$) was coupled with LPJ-DGVM and the model structure was modified so that surface SM can be directly assimilated into the coupled model at the daily time step.

In both LPJ-DGVM and PT-JPL, three ET components are modelled: soil evaporation (E), vegetation transpiration (T), and leaf evaporation (I). The PT-JPL$_{SM}$ introduced a constraint (0–1) of

164 surface SM for transpiration and soil evaporation, which was used to avoid the implicit soil water control

165 (represented by $f_{SM}=RH^{VPD}$) in the PT-JPL model. The constraint is determined by the amount of soil

166 water available that depends on soil properties, surface SM, atmospheric conditions, potential ET (PET),

167 and canopy height. The SMAP SM was applied to model global ET using PT-JPL$_{SM}$ and the results

168 demonstrated the largest improvements for ET estimates in dry regions (Purdy et al. 2018).

169 The ET$_{PM}$ estimated by the coupled model (LPJ-PM) introducing surface SM is regarded as the "truth

170 value" of the surface SM assimilation scheme (scheme 2). Subsequently, assimilated ET (ET$_{SM}$) was

171 converted to the top soil moisture (surface layer of 0–50 cm) through the soil water conversion function,

172 providing feedback for subsequent hydrologic and carbon cycle processes. Different from other

173 "constant" ET observations, the ET$_{SM}$ ("observation") at each time t were adjusted by absorbing

174 intermediate variables updated after assimilation at time t-1. The ET$_{PM}$ was shown to be better than ET

175 simulated by LPJ-DGVM but was as not as good as that simulated by the model with SMAP SM

176 assimilated (Li et al. 2020).

177 *2.2. Assimilation scheme and experiment procedure*

178 To improve the prediction capability of LPJ-PM, we designed three assimilation schemes:

179 assimilating LAI only(scheme 1, **output: ET$_{LAI}$, GPP$_{LAI}$**), assimilating ET only(scheme 2, **output:**

180 **GPP$_{SM}$, ET$_{SM}$**), and joint assimilation of LAI and ET ((scheme 3, **output: ET$_{CO}$, GPP$_{CO}$**), i.e., LPJ-

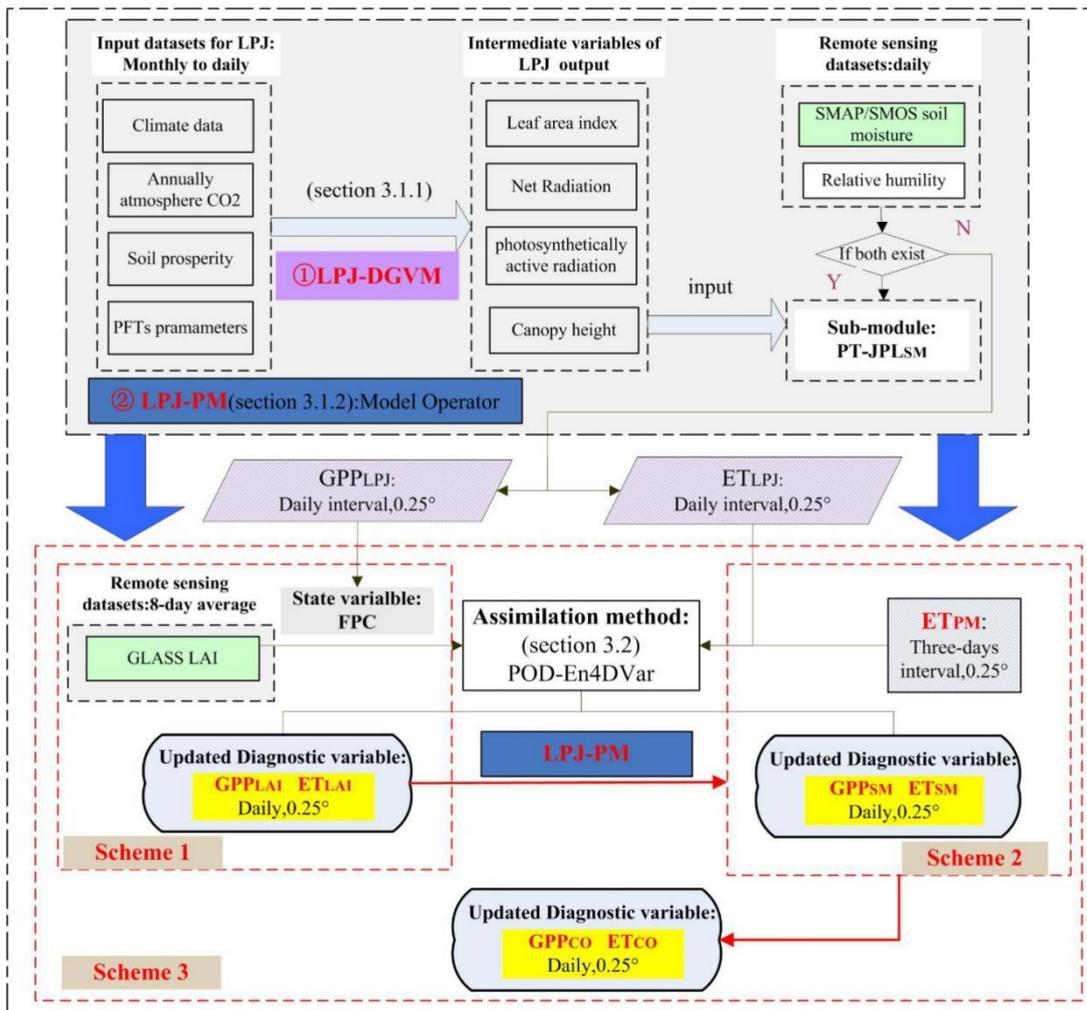
181 VSJA framework) to test the assimilation performance for simulating GPP and ET. The proposed LPJ-

182 VSJA framework consists of four main components: the model operator (the LPJ-PM), the observation

183 operator (to establish the relation between the assimilation variable and the observed variable), the

184 observation series (GLASS LAI and SMOS or SMAP products), and the assimilation algorithm

185 (POD4DVar). With the surface soil moisture constraint in the PT-JPL$_{SM}$, the LPJ-VSJA corrects the

186 output fluxes (GPP and ET in this study).

187

188   **Figure 1. Flowchart of the LPJ-VSJA assimilation scheme: three assimilation schemes and the coupled**

189   **model: LPJ-PM. (adapted from Li et al., 2020)**

190   The experiment consisted of six steps: (1) initialize the LPJ-DGVM and output the reference state

191   variables without assimilation over the experimental period (2010–2018), referred to as the "Control run"

192   scenario; (2) implement schemes 1, 2, and 3, respectively, and the results represent the assimilation

193   integration state (daily GPP and ET assimilation results are referred to as the "GPP$_{LAI}$" and "ET$_{LAI}$" in

194   scheme 1, "GPP$_{SM}$" and "ET$_{SM}$" in scheme 2 and "GPP$_{CO}$" and "ET$_{CO}$" in scheme 3; This scenario used

195   the same input data and model parameter scheme with the "Control run" scenario; (3) evaluate GPP and

196   ET results (schemes 1, 2 and 3) by comparing the parameters, R^2 (correlation coefficient), RMSD (root

197   mean square deviation), BIAS, and ubRMSD (unbiased root mean square deviation), for conditions of

198   without-DA ("Control run" scenario) and with-DA states, and assess the assimilation performance of

199   separate assimilation (schemes 1 and 2) and joint assimilation (scheme 3) to determine the optimal

200   assimilation scheme for GPP and ET, respectively; (4) evaluate the in-situ assimilated GPP and ET results

201   where the sites are located in wet or dry regions by dividing these validation sites into four parts (humid,

202   sub-dry humid, semi-arid, and arid regions), and this step was designed to assess the superiority of the

203   proposed assimilation scheme in water-limited areas; (5) compare the ET assimilation performance by

204   assimilating the SMOS data with that by assimilating the SMAP data; and (6) evaluate the simulated GPP

205   and ET maps based on the optimal assimilation scheme against existing global flux products.

206

2.2.1 DA scheme 1: LAI assimilation

In assimilation scheme 1, the observation operator determines the relationship between LAI and foliage projective cover (FPC) in the process model (equation 2.1), and the assimilated LAI will be propagated by energy transmission and ecosystem processes in the dynamic model to improve GPP and ET simulations. FPC, the vertically projected percentage of the land covered by foliage, regulates the rate of photosynthate conversion and transpiration. In this study, the GLASS LAI with 8-day interval for the period 2010–2018 was selected as the observation dataset for assimilation, and the FPC state variable was updated daily through running the LPJ-PM (GPP$_{LAI}$, ET$_{LAI}$ in this study) as shown below:

$$FPC = 1 - e^{-0.5LAI} \tag{2.1}$$

We set the model and observation errors at a given time as 20% and 10% of the LAI value and the observed LAI value, respectively. By verifying the assimilation performance (R, RMSD, BIAS,) for different scale factors of model simulation and observations in the range of 0.05 to 0.40, taking a step size of 0.05 (a total of 64 combinations), the optimal scale factors (0.2 and 0.1) were determined. The model integration generation method described by Pipunic et al. (2008) was used to determine the minimum number of integration members required to achieve maximum efficiency, and the number of sets was 20.

2.2.2 DA scheme 2: SM assimilation

In this scheme, the surface SM products (SMOS or SMAP) were assimilated to LPJ-PM to obtain more accurate ET (ET$_{SM}$) estimates in water-limited areas. The observation series was the SMOS or SMAP SSM product, and the observation operator was the PT-JPL$_{SM}$ model. The ET$_{PM}$ was directly

226   assimilated as a diagnostic variable. The assimilated ET was applied to compute the top layer SM (50 cm)

227   at the next time step (a nonlinear soil water availability function described by Zhao et al. (2013)). Then,

228   the updated SM values regulated the GPP simulation (output: GPP_{SM}).

229   All assimilation simulations were conducted between January 2010 and December 2018. Between

230   January 2010 and April 2015, SMOS data were used for assimilation, and after May 2015, both SMOS

231   and SMAP data were used for assimilation. An assimilation scheme was conducted when RH and SMOS

232   or SMAP SM data existed simultaneously; otherwise, the original simulation of the LPJ-DGVM was

233   conducted directly without adjustment of assimilation.

234   Similar to the LAI assimilation scheme, the model and observation errors were set as 15% and 5%

235   of ET_{LPJ} and ET_{PM}, respectively (LPJ-PM was adopted before assimilation). The number of set members

236   was set to 50. The ET_{PM} must be rescaled to the ET_{LPJ} distribution via their corresponding cumulative

237   probabilities using the cumulative distribution function (CDF) matching to avoid introducing any bias in

238   the LPJ-VSJA system (Li et al. 2020).

239   2.2.3 DA scheme 3: joint assimilation of LAI and SSM

240   In this scheme, both LAI from GLASS and SM from SMOS or SMAP were the observation datasets.

241   The GLASS LAI was assimilated by scheme 1 to obtain the FPC_{DA} and ET_{LAI}, and then the FPC_{DA} served

242   as input to LPJ-PM to simulate optimized ET_{PM}, and the ET_{LAI} was further assimilated with ET_{PM} to

243   generate ET_{CO}. Then, the SM (referred to as SM_{CO} in Figure S1) updated by ET_{CO} and the FPC_{DA} were

244   used as input to correct GPP. Finally, GPP$_{CO}$ and ET$_{CO}$ were output by joint assimilation based on the

245   POD-En4DVar method.

246   Here, we applied the error regulation in scheme 1 and maintained the error setting of the LAI

247   observation and model simulation. Considering the transmission of integrated model error, we

248   recalculated the model error of LPJ-PM after the LAI assimilation and set model and observation errors

249   of ET$_{LAI}$ and ET$_{PM}$ to be 15 and 10%, respectively. The number of set members was set to 70.

250   *2.3. POD-Based Ensemble 4D Variational Assimilation Method*

251   The PODEn4DVar (Tian et al. 2011) method has the advantage of avoiding the calculation of adjoint

252   patterns as its incremental analysis field, which can be represented linearly by the POD base, and then

253   saving computation. Moreover, the PODEn4DVar can simultaneously assimilate multiple-time

254   observation data and provide flow-dependent error estimates of the background errors. It has shown

255   advantages in terrestrial assimilation, Tan-Tracker joint DA, and Radar assimilation (Tian et al. 2010;

256   Tian et al. 2009; Tian et al. 2014; Zhang and Weng 2015).

257   By minimizing the following initial incremental format of the cost function in the 4DVar algorithm,

258   an analysis field can be obtained:

259   $$J(x') = \frac{1}{2}(x')B^{-1}(x') + \frac{1}{2}[y'(x') - y'_{obs}]^T R^{-1}[y'(x') - y'_{obs}]$$

260   Here, the $x' = x - x_b$, $y'(x') = y(x' + x_b) - y(x_b)$, $y'_{obs} = y_{obs} - y(x_b)$, $y = H[M_{to \to tk}(x)]$.

261   $x'(x'_1, x'_2, \ldots, x'_N)$ is the model perturbation (MP) matrix and $y'(y'_1, y'_2, \ldots, y'_N)$ is the

262   observation perturbation (OP) matrix with N samples. Following Rüdiger et al. (2010), the LAI

263   perturbation was set to a fraction (0.001) of the LAI itself. The perturbation of ET_{PM} and ET_{LPJ} conforms

264   to a Gaussian distribution with a mean of 0 and a specified covariance (10 and 5% of the ET_{PM} and ET_{LPJ}

265   at time t). The subscript b represents the background field, the superscript T represents a transpose, H is

266   the observation operator of scheme 1 as described in section 2.3, and scheme 2 is the PT-JPL$_{SM}$ (described

267   in Purdy et al. (2018)). M is the forecast model (LPJ-PM in this study), B is the background error

268   covariance, R is the observation error covariance, and obs denotes observation.

269   Assuming the approximately linear relationship between OP(y') and MP(x'), POD decomposition

270   and transformation were successively conducted for OP and MP. The transformed OP samples ($\Phi_y =$

271    $y'_1, y'_2, \ldots, y'_n$) are orthogonal and independent, and the transformed MP samples ($\Phi_x =$

272    $x'_1, x'_2, \ldots, x'_n$) are orthogonal to the corresponding OP samples, where n is the number of POD modes.

273   The manifestation of the background error covariance is the same as the Ensemble Kalman filter

274   (EnKF, Evensen (2004)), and the incremental analysis x'_a was expressed by the $\Phi_{x,n}$, and $\widetilde{\Phi}_y (\widetilde{\Phi}_y =$

275    $[(n-1)I_{n \times n} + \Phi_{y,n}^T R^{-1} \Phi_{y,n}]^{-1} \Phi_{y,n}^T R^{-1}$). Finally, the final analysis x_a is calculated as $x_a = x_b +$

276    $\Phi_{x,n} \widetilde{\Phi}_y y'_{obs}$. The detailed derivation process of the algorithm is described by a previous study (Tian et

277   al. 2011).

278   In the ensemble-based method, the number of ensemble members is usually fewer than that of the

279   observation data and the degrees of freedom of the model variables, and spurious long-range correlations

280 occur between observation locations and model variables. A practical method, the localization technique,

281 is applied to address this issue (Mitchell et al. 2002). The final incremental analysis is rewritten as:

282
$$x'_a = \Phi_{x,n}\widetilde{\Phi}_y y'_{obs} C_0(\frac{d_h}{d_{h,0}}) \cdot C_0(\frac{d_v}{d_{v,0}})$$

283 where d_h and d_v are the horizontal and vertical distances between the spatial positions of state and

284 observed variables, respectively; and $d_{h,0}$ and $d_{v,0}$ are the horizontal and vertical covariance localization

285 Schur radii, respectively. The filtering function C_0 is expressed as:

286
$$C_0(r) = \begin{cases} -\frac{1}{4}r^5 + \frac{1}{2}r^4 + \frac{5}{8}r^3 - \frac{5}{3}r^2 + 1, & 0 \le r \le 1, \\ \frac{1}{12}r^5 - \frac{1}{2}r^4 + \frac{5}{8}r^3 + \frac{5}{3}r^2 - 5r + 4 - \frac{2}{3}r^{-1}, & 1 \le r \le 2, \\ 0, & 2 < r \end{cases}$$

287 where r is the radius of the filter.

288   The assimilation algorithm is mainly divided into two steps: (1) prediction: run LPJ-PM in the

289 current assimilation window and generate simulation results and background field vectors; (2) update: the

290 algorithm is used to calculate the optimal assimilation increment x'_a and analysis solution x_a, and the

291 simulation results and the initial conditions of the model in the current window are updated using the

292 analysis solution. The updated initial conditions were applied for model LPJ-PM prediction, and the above

293 process was repeated.

294   *2.4. Validation method for assimilation performance*

295   The R^2 (correlation coefficient), RMSD (root mean square deviation), Bias, and ubRMSD (unbiased

296   root mean square deviation) between simulation and tower-based observations were applied for

297   evaluation. In addition, a Taylor chart was also used to demonstrate the performance of two ET

298   estimations with different SM observations in terms of R, ubRMSD, and SD on 2D plots, to display how

299   closely the datasets matched observations in one diagram (Taylor 2001). In the Taylor diagram, SD

300   represents the radial distance from the origin point and the correlation with the site observations as an

301   angle in the polar plot. The ubRMSD is the distance between the observation and the model and is

302   represented in the figure as a green semi-circular arc with point A as the center of the circle. The closer

303   the model point to the reference point (Point A), the better the performance. This diagram is convenient

304   and visual in evaluating multiple aspects of various models.

305   The error variance of GPP and ET products was estimated using the triple collocation (TC) approach

306   (Stoffelen 1998) to validate the global simulation in this study. The method has been extensively applied

307   in the study of hydrology and oceanography (Caires and Sterl 2003; Khan et al. 2018; O'Carroll et al.

308   2008; Stoffelen 1998), particularly in SM studies (Chan et al. 2016; Kim et al. 2018). The TC provides a

309   reliable platform for comparison of spatial assimilation results and in-situ measurements. In this

310   experiment, no calculation was performed on the non-vegetated areas where the correlation was lower

311   than 0.2 to have independent datasets and avoid correlated errors (crucial assumptions in TC) (Yilmaz

312   and Crow 2014).

313

3. Experiment sites and data

3.1. Description of flux tower sites

We screened over 300 EC flux sites across the globe from the FLUXNET2015 (https://fluxnet.fluxdata.org/data/fluxnet2015-dataset/), AmeriFlux (http://public.ornl.gov/ameriflux), and the HeiHe river basin (Li et al. 2013) for the evaluation of assimilation performance over the period from January 2010 to December 2018. The in-situ half-hourly LE and GPP data from the sites were aggregated into daily data. The daily gap-filled data were excluded if the percentage of gap-filled half-hourly values was more than 20%. Then we corrected the data of energy non-closure by using the Bowen ratio closure method (Twine et al. 2000) to improve the energy closure rate (Huang et al. 2015; Yang et al. 2020). The data were selected to cover the 2010–2018 period with at least one year of reliable data, and the result from the error of assimilation is relative to the LE value and seasonal variation (Purdy et al. 2018; Zou et al. 2017). It is essential to have available data every month during a one-year period, and only days with less than 25% missing data were processed per month (Feng et al. 2015). In addition, for flux tower data, the data were also excluded for the analysis if the SMAP/SMOS SM data were not of good quality.

Finally, we identified a total of 105 sites across the globe encompassing five major biomes: grassland (18 for GPP and 19 for ET), savanna (11), shrubland (4), forest (49 and 53), and cropland (13 and 14). In the comparative analysis of the performance for simulating ET by assimilating SMOS and SMAP SM data separately, we selected 46 AmeriFlux sites (Figure S3) with at least one year of reliable data from

333   2015 to 2018 based on the simultaneous availability of SMAP and SMOS data, including grassland (19),

334   savanna (11), shrubland (5), forest (23), and cropland (7). Figure S2 and S3 illustrate the location and

335   distribution of the 105 and 46 EC flux tower sites, respectively. A more detailed description is

336   summarized in the Supporting Information Table S1.

337   *3.2. Remote sensing datasets: LAI and SM*

338   The GLASS LAI product with an 8-day time step and 5 km resolution was derived from MODIS

339   and CYCLOPES surface reflectance and ground observations using general regression neural networks

340   (GRNNs)   (Liang et al. 2013; Xiao et al. 2016). The verification of the product using the mean values

341   of high-resolution LAI maps showed that the GLASS LAI values were closer to these high-resolution

342   LAI maps (RMSD= 0.78 and R^2= 0.81). Therefore, the GLASS LAI product has satisfactory performance

343   and can be assimilated into terrestrial biosphere models.

344   The SMAP mission (Entekhabi et al. 2010) and SMOS mission (Jacquette et al. 2010), the two

345   dedicated soil moisture satellites currently in orbit equipped with L-band microwave instruments, provide

346   surface SM retrievals. We chose the SMOS-L2 product and the SMAP-L3-Enhanced product, which both

347   provide global coverage every three days for soil depth of 5 cm. Only good-quality SMAP and SMOS

348   data were used. The grid cells with water areas larger than 10% and those with less than 50% good-quality

349   data in one year were masked out, which alleviates the undesirable model simulations caused by the

350   decrease in SMAP retrieval accuracy (Chan et al. 2016; O'Neill et al. 2010). We only adopted the data

351   with an uncertainty below 0.1 m^3 m^{-3}, in the actual range (0.00–0.6 m^3 m^{-3}), and the temperature of the

352   LSM observation layer (the second layer) was higher than 2 °C (Blyverket et al. 2019).

353

354   *3.3. Model-forcing and validation datasets*

355   In this study, the meteorological, soil property, and CO_2 concentration datasets were used to drive

356   the LPJ-PM. The model-forcing datasets were interpolated to 0.25° based on the bilinear method to ensure

357   the consistency of spatial representation. Table 2 provides the spatial and temporal characteristics of the

358   model-forcing datasets in the LPJ-PM (submodule: LPJ-DGVM and PT-JPL$_{SM}$).

359

360   **Table 2. List of the selected forcing and remote-sensing datasets used in this study**

Datasets	Variable	Period	Spatial resolution	References
CRU TS v4.1[a]	Cloud cover, temperature, precipitation, wet day	1901-1930	0.5°× 0.5°	New et al. (2000)
Ice-core measurements and atmospheric	Atmospheric CO_2 concentrations	1901-2018	NA	(Etheridge et al. (1996); Keeling et al.

observations at the Mauna Loa Observatory [a]				(1995))
MERRA-2 [a]	Precipitation, surface temperature, cloud fraction, relative humidity	2010-2018	$0.5° \times 0.625°$	Rienecker et al. (2011)
HWSD (v121)[b]	Soil texture data	NA	1 km×1 km	Wieder et al. (2014)
SPL3SMP_E[b]	Surface soil moisture	2015.4–present	9 km×9 km	Entekhabi et al. (2010)
GLASS LAI [a,b]	Leaf area index	2010-2018	5 km×5 km	Xiao et al. (2016)
SMOS_L3 CATDS[b]	Surface soil moisture	2010-present	25km×25 km	Jacquette et al. (2010)

361   [a]: forcing dataset for LPJ-DGVM

362   [b]: external input dataset for PT-JPL$_{SM}$

363

364   We used four global ET products and three global GPP products (Li et al. 2018; Li and Xiao 2019;

365   Wang et al. 2017) to evaluate the performance of the model with the joint assimilation scheme. Table 3

366   shows the details of these GPP and ET products.

367   **Table 3. Global GPP and ET products for comparison in this study**

Product	Dataset	Temporal resolution	Spatial resolution	Retrieval algorithm	References
MOD17A2	GPP and ET	8-day average	1 km × 1 km	GPP: Based on the light use efficiency (LUE) model ET: Improved Penman formula	Running et al. (2004)
GLASS	GPP and ET	8-day average	5 km × 5 km	GPP: EC-LUE model ET: Combining five Bayesian averages based on process models (BMA)	Yuan et al. (2010)
GOSIF GPP	GPP	8-day average	0.05° × 0.05°	Estimated from solar-induced chlorophyll fluorescence with GPP-SIF relationships	Li and Xiao (2019)
GLDAS ET	ET	daily	0.25° × 0.25°	Processed model assimilation	Fang et al. (2009)

GLEAM v3a ET	ET	daily	0.25°× 0.25°	Processed model assimilation	Martens et al. (2017)

368

4. Results

4.1. Performance of LPJ-PM for simulating GPP and ET with the assimilation of LAI and soil moisture

4.1.1 Accuracy assessment of GPP for separate and joint assimilation

In general, the R^2 between GPP_{LPJ} and GPP_{OBS} was above 0.4 at most of the sites (62 sites) and were relatively weak for some sites ($R^2 < 0.4$). The LAI assimilation improved the simulations at most sites (R^2 value increased at 82 sites), particularly for sites in the U.S. and Europe (Figure 2). The R^2 improvement from the LAI assimilation (scheme 1) was superior to that from the SM assimilation (Figure 2-R^2 (b) and (c)). The performance of the joint assimilation (scheme 3) was similar to that of scheme 1. The GPP assimilation performance in terms of RMSD between scheme 1 and scheme 2 was not significantly different. For sites with no improvement in the separate assimilation (schemes 1 and 2), the RMSD, however, was improved in the joint assimilation (scheme 3) (Figure 2). More sites (Figure 2-BIAS (a)) showed positive bias (GPP_{OBS}-GPP_{LPJ}), which were mainly distributed in the humid and dry-sub humid forest, grassland, and arid cropland regions, showing underestimation for GPP_{OBS}. The assimilation improved the accuracy for overestimated sites, but there was no significant improvement for underestimated sites. The ubRMSD implied that the SM assimilation alone had a

384 better performance than the LAI assimilation alone, especially for sites in arid areas. The analysis of the

385 above four statistical measures (R^2, RMSD, BIAS, and ubRMSD) indicated that the accuracy of joint

386 assimilation was better than that of separate assimilation.

387   At the seasonal scale, all three assimilation schemes corrected the model trajectory and

388 significantly improved the growing season simulations, especially for peak values (IT-Tor, US-NR1,

389 US-NE1). In addition, the linear fitting of GPP_{CO} and GPP_{OBS} on a monthly scale was closer to 1:1 (y=

390 $0.92 + 21.66$ p < 0.001) than that of GPP_{LAI} (y= $0.89 + 28.3$, p < 0.001) and GPP_{SM} (y= $0.86 + 41.70$, p

391 < 0.001) (Figure S5). The results in Table S2 support the above analysis, and the joint assimilation

392 showed advantages in overall accuracy in both arid and humid areas.

393

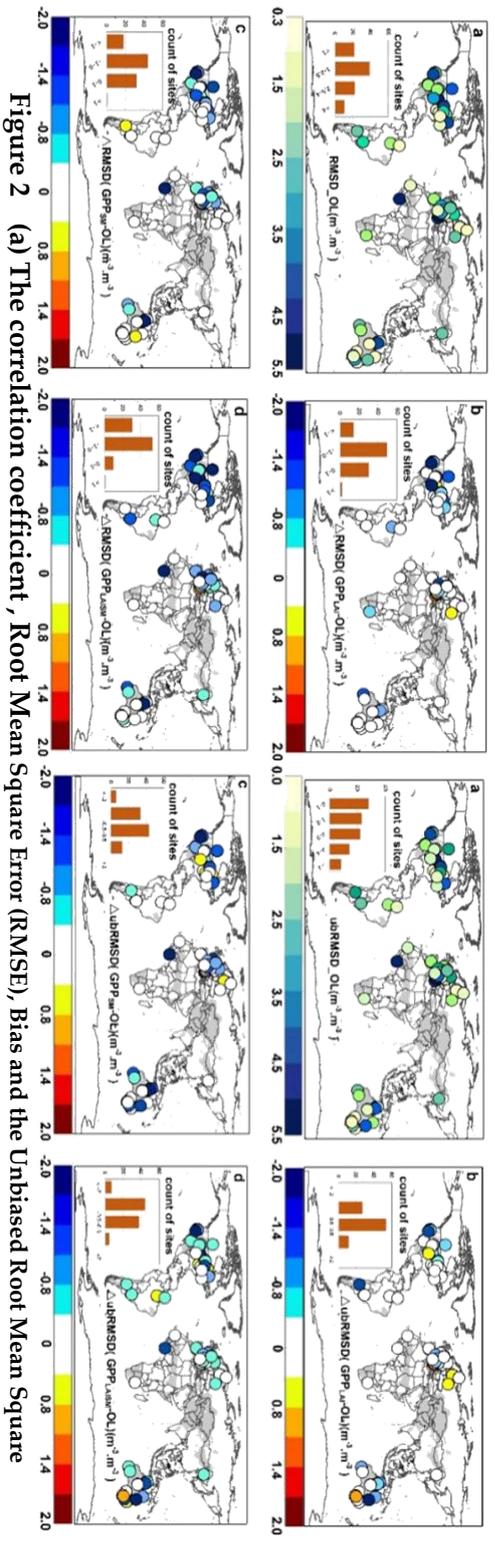
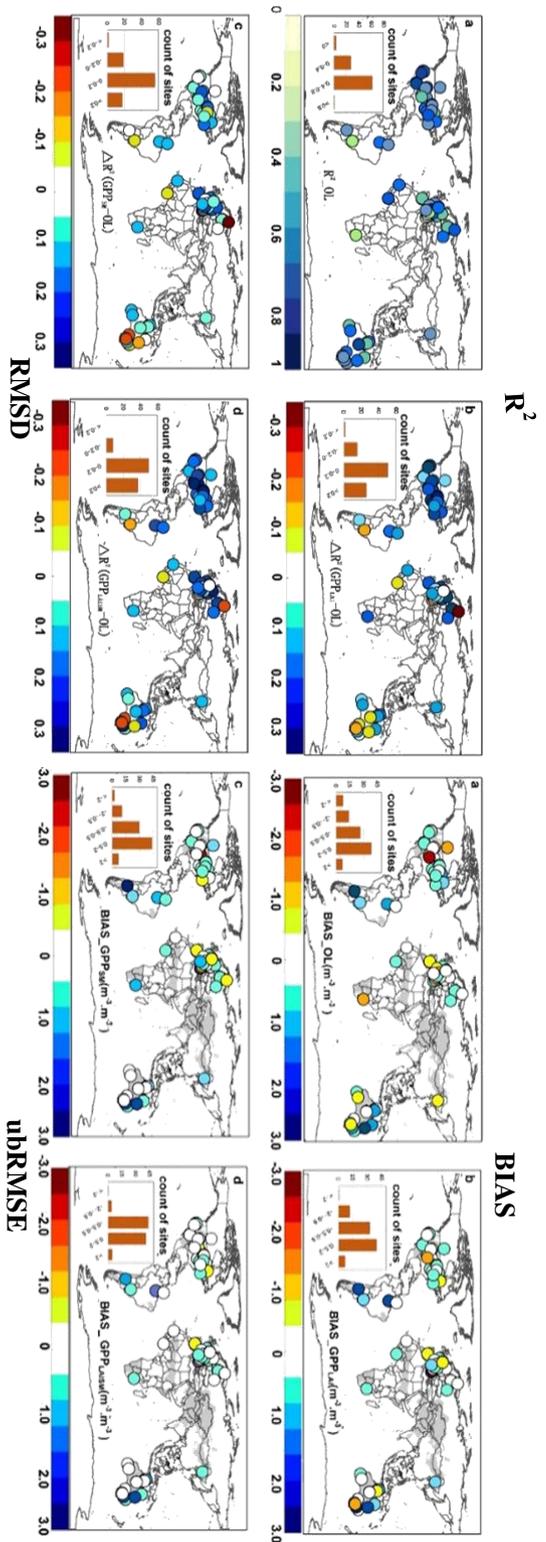



Figure 2 **(a) The correlation coefficient , Root Mean Square Error (RMSE), Bias and the Unbiased Root Mean Square**

Error (ubRMSE) between the GPP(GPP$_{LPJ}$) simulated by the LPJ-DGVM and the site observations, the yellow/blue indicating

low/high correlation ;(b) ΔR(correlation difference between GPP$_{LAI}$ and GPP$_{LPJ}$), ΔRMSD GPP$_{LAI}$-GPP$_{LPJ}$) , BIAS (GPP$_{LAI}$) and

ΔubRMSD(GPP$_{LAI}$- GPP$_{LPJ}$);(c) ΔR(correlation difference between GPP$_{SM}$ and GPP$_{LPJ}$), ΔRMSD(GPP$_{SM}$- GPP$_{LPJ}$),BIAS

(GPP$_{SM}$)and ΔubRMSD(GPP$_{SM}$- GPP$_{LPJ}$);(d) ΔR(The correlation difference between GPP$_{CO}$ and GPP$_{LPJ}$), ΔRMSD

(GPP$_{CO}$- GPP$_{LPJ}$), BIAS(GPP$_{CO}$)and ΔubRMSD(GPP$_{CO}$- GPP$_{LPJ}$), blue/red represent positive/negative values.

395

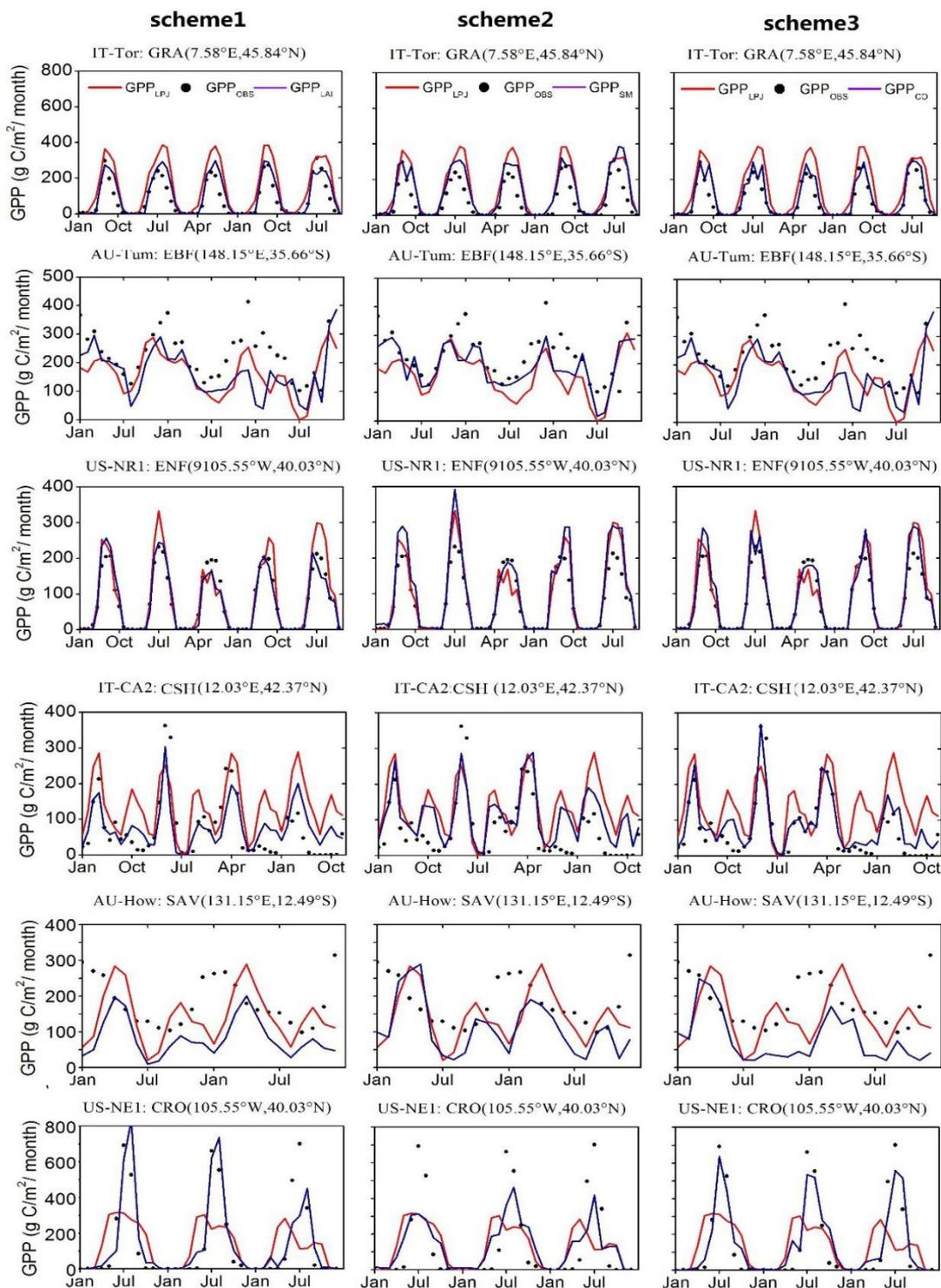



Figure 3. Seasonal cycles of tower GPP and simulated gross primary productivity (GPP) from Lund-Potsdam-Jena (LPJ), GLASS LAI assimilation (scheme 1), SMOS assimilation (scheme 2) and joint assimilation (scheme 3) for six sites representing six PFTs.

The residual analysis indicated that the three assimilation schemes for GPP (Figure S7 (left)) were different. For the assimilation results, most of the errors were distributed around $-70 \sim 60$ g C m^{-2} month^{-1}. The high GPP$_{OBS}$ values were considerably underestimated. The maximum negative error reached 100 g C m^{-2} month^{-1}. The error distribution of GPP$_{SM}$ was more dispersed than that of GPP$_{LAI}$ and GPP$_{CO}$. Among the residuals of these three schemes, GPP$_{SM}$ significantly overestimated the GPP$_{OBS}$, mainly distributed in the 0–200 g C m^{-2} month^{-1} range. GPP$_{LAI}$ showed significant improvement in the overestimation of GPP$_{OBS}$ compared with GPP$_{CO}$. In general, the GPP$_{CO}$ with the most concentrated error distribution had significant improvement.

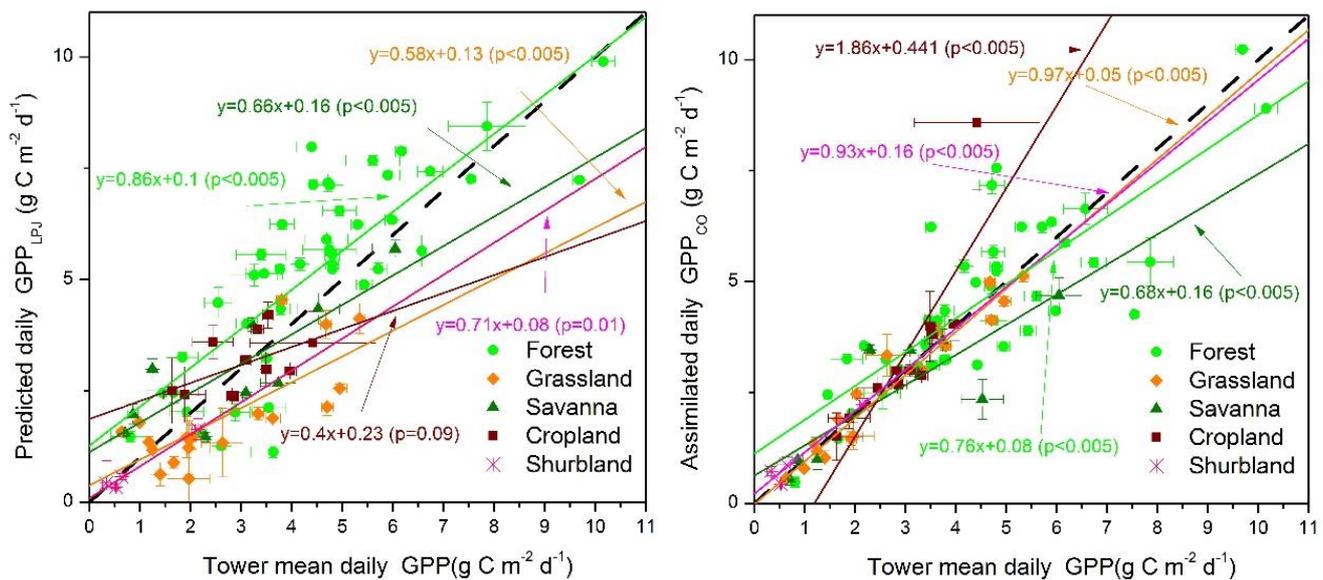



Figure 4. Scatterplots of daily GPP$_{LPJ}$ (left) and GPP$_{CO}$ (right) versus tower GPP for different PFTs.

After determining the optimal assimilation scheme (scheme 3), we evaluated the GPP$_{LPJ}$ and GPP$_{CO}$ at the site level (Fig.4). The results showed that GPP$_{CO}$ performed better (R^2= 0.83, RMSD= 1.05 g C m^{-2} d^{-1}) than GPP$_{LPJ}$ (R^2= 0.69, RMSD= 2.15 g C m^{-2} d^{-1}). The noticeable underestimation in all PFTs and overestimation at most forest sites for GPP$_{LPJ}$ were corrected by joint assimilation (GPP$_{CO}$). Our joint assimilation methods had better performance in forests, shrublands, and grasslands than in croplands and savannas. Except for the cropland, the linear fitting results of other types were all below the 1:1 line, showing the overall underestimation. Superior performance in both original simulation and assimilation occurred at shrubland (R^2= 0.93, RMSD= 0.27 g C m^{-2} d^{-1}) and grassland (R^2= 0.97, RMSD= 0.38 g C m^{-2} d^{-1}) sites. However, the standard deviation of GPP$_{CO}$ and GPP$_{OBS}$ at savanna sites was relatively large, and the assimilated GPP at several savanna sites was significantly underestimated.

4.1.2 Accuracy assessment of ET for separate and conjunct assimilation

In general, the coefficient of determination (R^2) between ET$_{LPJ}$ and ET$_{OBS}$ was generally over 0.4 (the simulations were superior to GPP$_{LPJ}$) (Figure 5). ET$_{LAI}$ showed slightly higher R^2, while some sites showed reduced values (41 sites). The ET$_{SM}$ and ET$_{CO}$ were significantly improved compared with the ET$_{LAI}$. The R^2 increased considerably in Australia but declined at some sites in the United States after assimilation. For RMSD and ubRMSD, ET$_{CO}$ performed better than ET$_{SM}$ and ET$_{LAI}$. The SM assimilation improved more in humid regions, while the RMSD of ET$_{LAI}$ only slightly improved in Australia and the ubRMSD of ET$_{SM}$ was slightly higher in South America. According to the bar chart

428 (Fig.5-RMSD(b)), the LAI assimilation increased the ET simulation error. In the original LPJ-DGVM

429 simulation, the sites with a negative bias were mostly located in the humid and dry-sub humid regions,

430 while most of the sites in arid and semi-arid regions had underestimation (Fig. 5-BIAS(a), Table S3).

431 The assimilation improved ET at some of the overestimated sites, but the underestimation over these

432 sites showed little improvement.

Figure 5 (a) The correlation coefficient, the Root Mean Square Error(RMSE), BIAS and the Unbiased Root Mean Square Error (ubRMSE) between the ET(GPP$_{LPJ}$) simulated by the LPJ-DGVM and the site observations, with yellow/blue indicating low/high correlation or ubRMSE;(b) ΔR(correlation difference between ET$_{LAI}$ and ET$_{LPJ}$), ΔRMSD (ET$_{LAI}$- ET$_{LPJ}$) , BIAS (ET$_{LAI}$) and Δ ubRMSD (ET$_{LAI}$- ET$_{LPJ}$) ;; (c) ΔR (correlation difference between ET$_{SM}$ and ET$_{LPJ}$) , ΔRMSD (ET$_{SM}$- ET$_{LPJ}$) , BIAS (ET$_{SM}$) and Δ ubRMSD (ET$_{SM}$- ET$_{LPJ}$) ;(d) ΔR (The correlation difference between ET$_{CO}$ and ET$_{LPJ}$) , ΔRMSD (ET$_{CO}$- ET$_{LPJ}$) , BIAS (ET$_{CO}$) and ΔubRMSD (ET$_{CO}$- ET$_{LPJ}$) , blue/red represent positive/negative value.

434  At the seasonal scale, the model simulations were able to capture the temporal trend of ET_{OBS}, and

435  joint assimilation significantly improved the simulation in the growing season (US-NR1, US-NE1);

436  overall underestimation was observed for ET_{OBS}, especially in winter (Fig.6). Overall, the linear fitting of

437  monthly ET_{CO} and ET_{OBS} was closer to 1:1 than that of ET_{LAI} and ET_{SM} (Figure S6). The simulation

438  accuracy of joint assimilation was better than that of separate assimilation, and the performance of the

439  SM assimilation was better than that of the LAI assimilation.

440  The ET residual analysis (Figure S7 (right)) indicated that the three assimilation scheme errors

441  showed underestimation for ET_{OBS}. In general, the error distribution of separate assimilations was more

442  dispersed than that of the joint assimilation. Similar to the assimilation performance of GPP, ET_{CO} and

443  ET_{SM} significantly improved the overestimation of ET_{OBS}, but did not significantly improve the

444  underestimation. For the ET_{CO}, most of the errors were distributed around -30–18 mm month^{-1}. The region

445  with high ET_{OBS} was considerably underestimated, and the maximum negative error reached -57 mm

446  month^{-1}.

447  We also evaluated the ET assimilation results at the PFT scale (Figure 7). The results showed that

448  our assimilated ET performed better at the site level (R^2= 0.77, RMSD= 0.28 mm d^{-1}) than that of ET_{LPJ}

449  (R^2= 0.67, RMSD=0.98 mm d^{-1}). Joint assimilation significantly reduced the errors of those shrubland

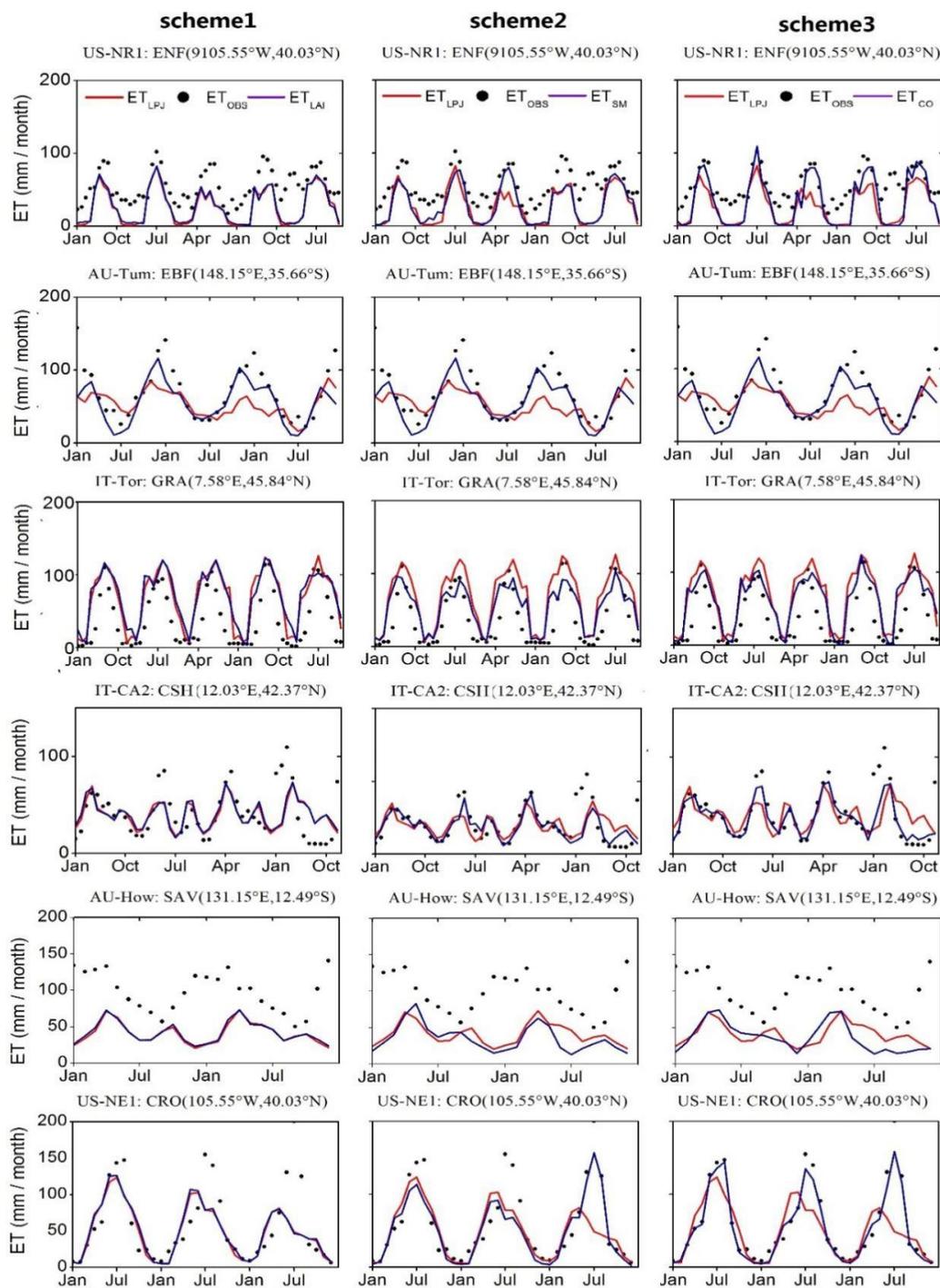
450  sites with overestimation for ET_{OBS}, and the site distribution was closer to the 1:1 line. Our assimilation

451  methods had better performance in forest, savanna, and grassland ecosystems than in cropland and

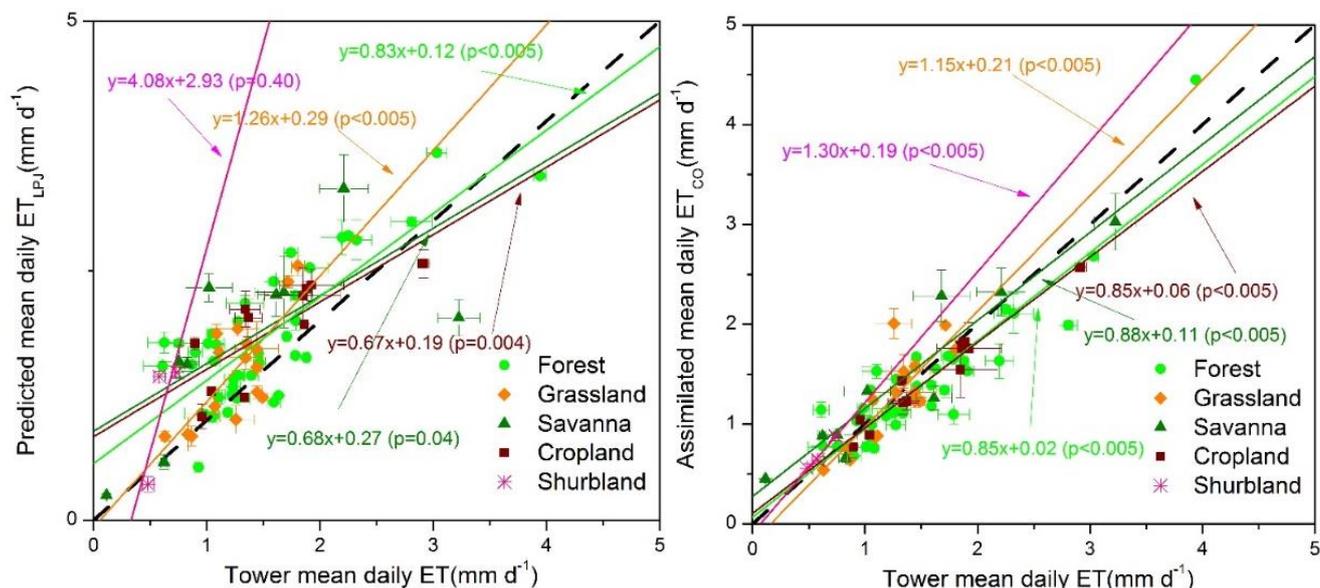
452  shrubland (Table S3). The linear fitting results of grassland and shrubland were all above the 1:1 line,

453  showing overall overestimation. Although the original simulation and assimilation performance were

454  superior at savanna sites (R^2= 0.95, RMSD= 0.33 mm d^{-1}), the standard deviations of ET_{CO} and ET_{OBS} at

455   savanna sites were relatively large, which was similar to the GPP results at savanna sites.

457 **Figure 6. Seasonal cycles of tower-based and simulated ET from Lund-Potsdam-Jena (LPJ), GLASS LAI**

458 **assimilation (scheme 1), SMOS assimilation (scheme 2) and joint assimilation (scheme 3) for the six sites**

459 **representing six PFTs during the study period.**

461 **Figure 7. Scatter plots of daily ET_{CO} versus tower ET under different PFTs.**

462 *4.2. Comparison of assimilation performance in semi-arid and arid regions with that in humid and dry-*

463 *sub humid regions*

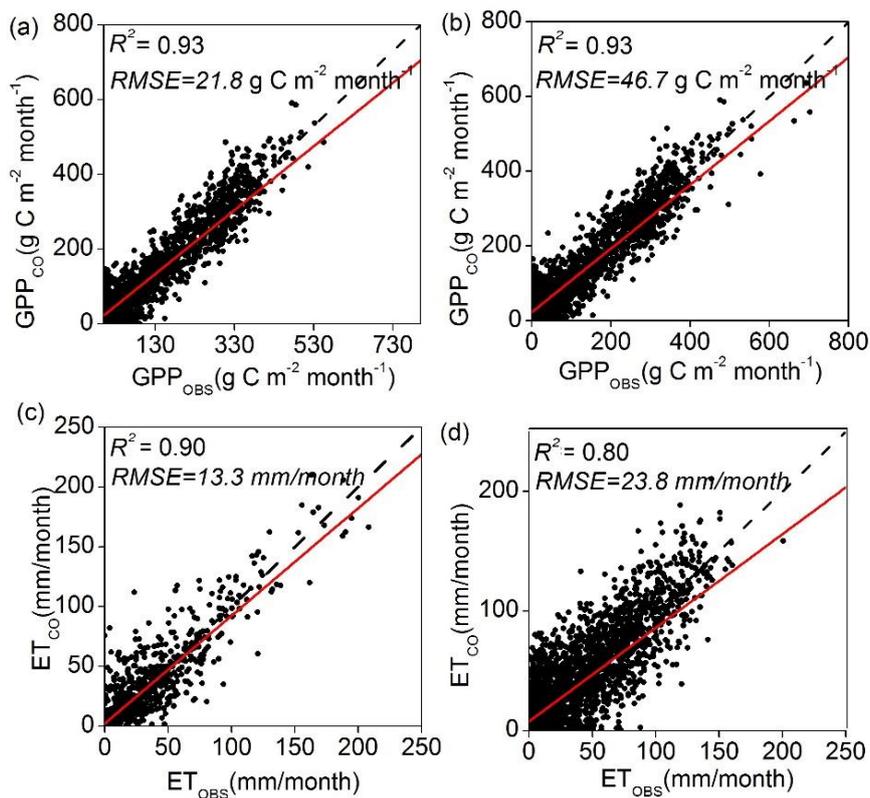



Figure 8. Scatter plots of daily tower GPP and ET versus GPP$_{CO}$ and ET$_{CO}$ under different arid and humid sites: (a) and (c) are the fitting results of GPP and ET in arid and semi-arid regions, respectively; (b) and (d) are the fitting results of GPP and ET in humid and dry sub-humid zone, respectively.

During the period 2010–2014, monthly GPP$_{CO}$ and ET$_{CO}$ performed differently in humid and sub-dry humid regions and semi-arid and arid regions (Figure 8, Table S2,3). Overall, the GPP and ET simulations had good consistency with the tower data in the two regions. For GPP$_{CO}$, there was no significant difference in the correlation and fitting coefficients between the two regions. Due to the higher standard deviation and GPP$_{OBS}$ values in the humid and sub-dry humid regions than in the semi-

473   arid and arid regions, the higher RMSD of the humid and sub-dry humid regions did not reflect higher

474   accuracy in semi-arid and arid regions. As for ET_{CO}, the fitting results and R^2 values in the semi-arid

475   and arid regions performed better than those in the humid and sub-dry humid regions, which also

476   suggested the importance of surface SM for ET estimation in water-limited areas.

477   On the daily scale, the original GPP simulations (GPP_{LPJ}) performed better in the semi-arid and

478   arid regions than in the humid and sub-dry humid regions with higher R^2 and lower ubRMSD (Table S2).

479   For GPP_{CO}, the shrubland in the semi-arid and arid regions had the lowest R^2 values and the second lowest

480   ubRMSD. The forest in the semi-arid and arid regions had the largest improvement after assimilation. In

481   the humid and sub-dry humid regions, the GPP_{CO} of the savanna and cropland showed the largest

482   improvement (R^2 increased by 64.7% and 71.1%, respectively; ubRMSD decreased by 47.0% and 31.8%,

483   respectively). The grassland in the semi-arid and arid regions had the highest R^2, and the savanna by

484   combining all indicators had the best assimilation results compared to other types in both regions.

485   Similar to ET_{CO}, the ET_{LPJ} in the semi-arid and arid *regions* was better than that in humid and sub-

486   dry humid regions in terms of four evaluation indicators (RMSD decreased by 32.7%, ubRMSD decreased

487   by 34.4% in semi-arid and arid regions and RMSD decreased by 26.4%, and the ubRMSD decreased by

488   30.9% in humid and sub-dry humid regions compared with ET_{LPJ}). The performance of the original

489   simulation and assimilation of grassland sites in the semi-arid and arid regions was the best among all

490   five PFTs.

491

492

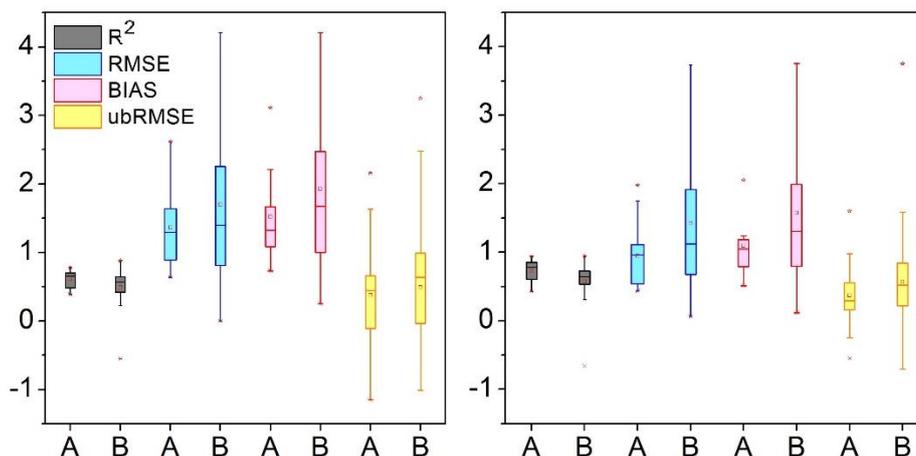



Figure 9. Boxplots of R^2, ubRMSD, RMSD and BIAS for GPP$_{SM}$ (left) and ET$_{PM}$ (right). A represents the sites in arid and semi-arid areas, and B represents the sites in humid and dry sub-humid areas.

To investigate the reasons for better assimilation performance in water-limited regions, we evaluated the GPP and ET simulated by the LPJ-PM according to R^2, ubRMSD, RMSD, and BIAS (Figure 7). Compared with the semi-arid and arid regions, the humid and sub-dry humid region had smaller R^2 mean, larger BIAS and RMSD mean, and no significant difference in mean ubRMSD for GPP$_{SM}$. In general, the evaluation results of joint assimilation for ET$_{PM}$ were generally consistent with those for GPP$_{SM}$ and GPP$_{SM}$. ET$_{PM}$ showed underestimation, which was consistent with the underestimation in SM assimilation. These results indicated that, both GPP and ET modeled by LPJ-PM with joint assimilation were less stable and had a lower performance in the humid and sub-dry regions than in the semi-arid and arid regions.

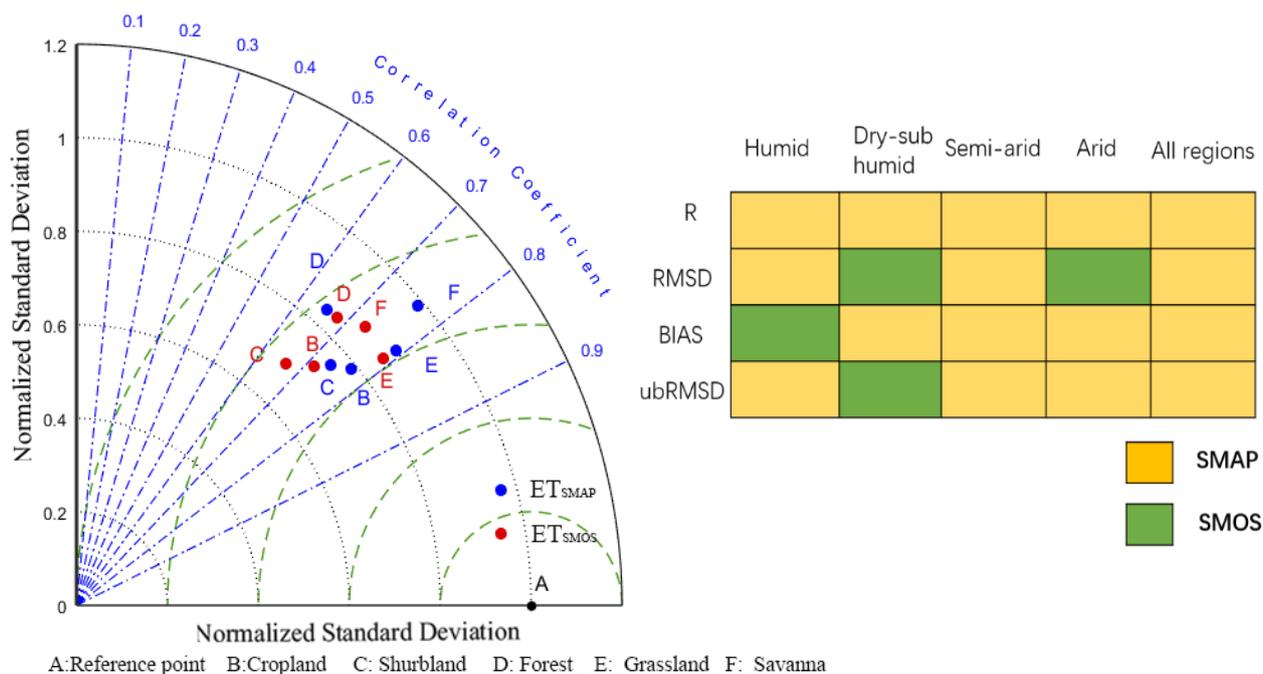
4.3. Comparison of assimilation performance in assimilating SMOS and SMAP soil moisture data

504

Figure 10. Taylor diagram (left) comparing ET simulations with observations at all 46 AmeriFlux sites at

the daily time step between April 2015 and December 2018. Blue dots represent results based on

assimilation with SMAP SM only and red dots represent results based on assimilation with SMOS SM

only. Reference points A and B-F correspond to the vegetation functional types (PFTs). The grid diagram

(right) compares the evaluation indices of ET simulations with those of the observed values at all 46

AmeriFlux sites with different wet and dry zones at the daily time step; the yellow cells indicate that

ET$_{SMAP}$ performs better in the metric, and green cells indicate that ET$_{SMOS}$ performs better in the metric.

512

513   The Taylor chart was used to compare the assimilation performance of ET$_{SMAP}$ and ET$_{SMOS}$ at 46

514   AmeriFux sites (Figure 10-left). The results showed that ET_{SMAP} performed better than ET_{SMOS} for all

515   PFTs. Both ET_{SMAP} and ET_{SMOS} performed well for grassland (closer to point A), and there was little

516   difference between R^2 and standardized RMSD. The NSD of ET_{SMAP} in grassland was 0.88, which was

517   closer to 1 than that of ET_{SMOS}. The assimilation of ET in the forest had a lower R and higher standardized

518   RMSD (0.7-0.8) than those of other PFTs, and the NSD of cropland and shrubland was lower than that of

519   other PFTs (0.6-0.8), indicating that the assimilation for cropland and shrubland could not reproduce the

520   variations in ET effectively. However, ET_{SMAP} showed significant improvement in R^2 compared with

521   ET_{SMOS} for shrubland and cropland. The assimilation performance of ET_{SMAP} and ET_{SMOS} for savanna

522   showed the greatest difference. In general, the ET_{SMAP} and ET_{SMOS} were slightly different, and the ET_{SMAP}

523   was more improved than ET_{SMOS}.

524   Figure 10 (right) shows the assimilation accuracy of ET_{SMOS} and ET_{SMAP} in different humid and arid

525   regions. The ET_{SMAP} had significant advantages for the four indicators. The R of ET_{SMAP} was higher than

526   that of ET_{SMOS} in all the areas. However, ET_{SMOS} in some evaluation indicators showed a better

527   performance than ET_{SMAP} (BIAS in the humid region; RMSD and ubRMSD in the sub-dry humid region).

528   This may be due to the overall more humid nature of SMOS SM than the SMAP SM. Moreover, the

529   sensitivity of deep soil moisture contributed more to the ET in humid areas than in the water-limited areas.

530   *4.4. Global simulations of GPP and ET with joint assimilation of LAI and soil moisture data*

531   To assess the spatial scalability of the LPJ-VSJA assimilation scheme, we simulated the global daily

532   GPP and ET for 2010–2018 with a spatial resolution of 0.25°. The original results simulated by the LPJ-

533   DGVM and LPJ-VSJA were referred to as LPJ-DGVM GPP(ET) and LPJ-VSJA GPP(ET), respectively.

534   We compared the annual spatial GPP and ET values and the error standard deviation of the LPJ-VSJA

535   with several existing flux products.

536   Figures 11 and 12 depict the spatial distribution of the annual mean and the differences between our

537   simulation results and the global independent satellite-based products. The developed LPJ-VSJA GPP

538   was the closest to GOSIF GPP (Li and Xiao 2019) in most regions with the lowest spatial mean deviation

539   (LPJ-VSJA-GOSIF) (27.9 g C m^{-2} yr^{-1}), followed by GLASS GPP (51.2 g C m^{-2} yr^{-1}) (Yuan et al. 2010),

540   LPJ-DGVM (-73.4 g C m^{-2} yr^{-1}), and MODIS GPP (93.1 g C m^{-2} yr^{-1}). LPJ-VSJA had higher GPP values

541   than GOSIF GPP in tropical regions, such as Amazonia, Central Africa, and Southeast Asia. In general,

542   the annual mean and differences between MODIS, GOSIF GPP, LPJ-DGVM, and our LPJ-VSJA were

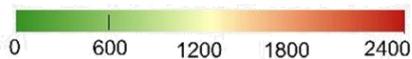
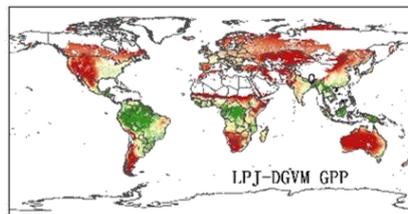
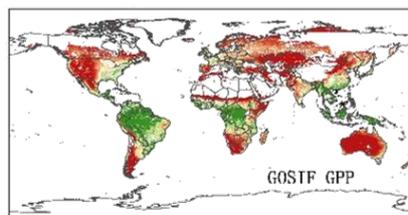
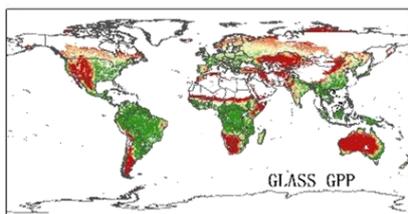
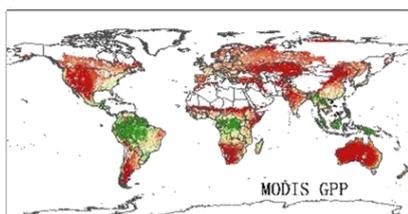
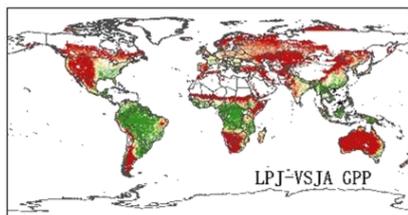
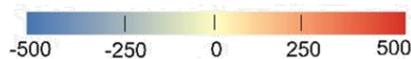
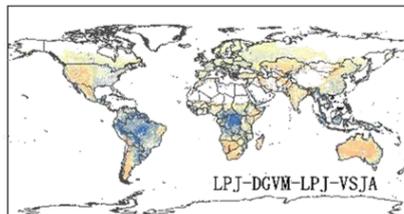
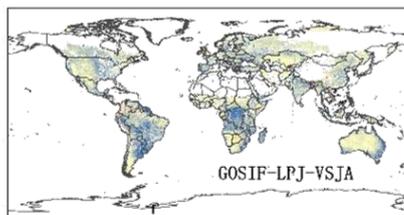
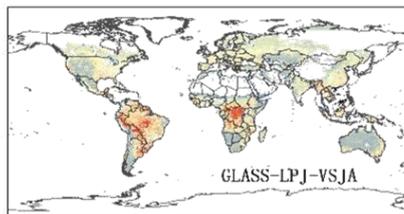
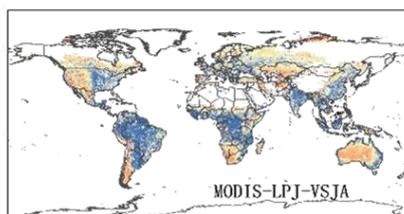
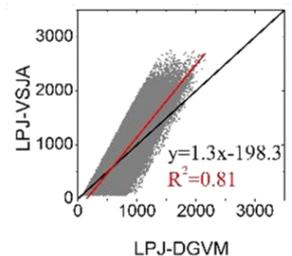
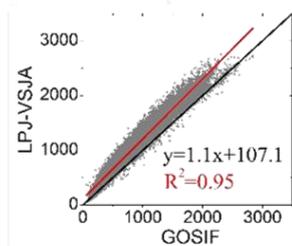
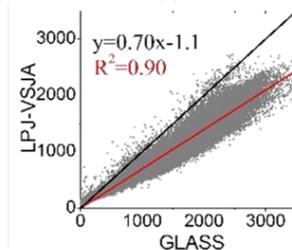
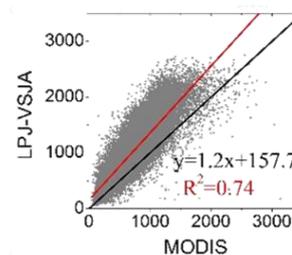
543   in broad agreement (with higher R^2 ranging from 0.74 to 0.95).

544   LPJ-VSJA ET was the closest to GLEAM ET on the spatial average with the least spatial average

545   deviation (-13.9 mm yr^{-1}) and highest R^2 (0.88), followed by GLASS ET (-23.1 mm yr^{-1}and 0.82), GLDAS

546   ET (-34.7 mm yr^{-1} and 0.73), LPJ-DGVM (-48.7 and 0.66 mm yr^{-1}), and MODIS ET (-122.1and 0.54 mm

547   yr^{-1}).

548

549 **Figure 11. Column A: Spatial distribution of annual LPJ-VSJA GPP and other independent satellite-based**

550 **datasets (a: MODIS GPP; b: GLASS GPP; c: GOSIF GPP; e: LPJ-DGVM). Column B: Spatial**

551 **distribution of the difference between annual LPJ-VSJA GPP and other independent satellite-based**

552 **datasets. Column C: Scatter plots between these products. Black lines show the 1:1-line, red lines show the**

553 **regression fit.**

554

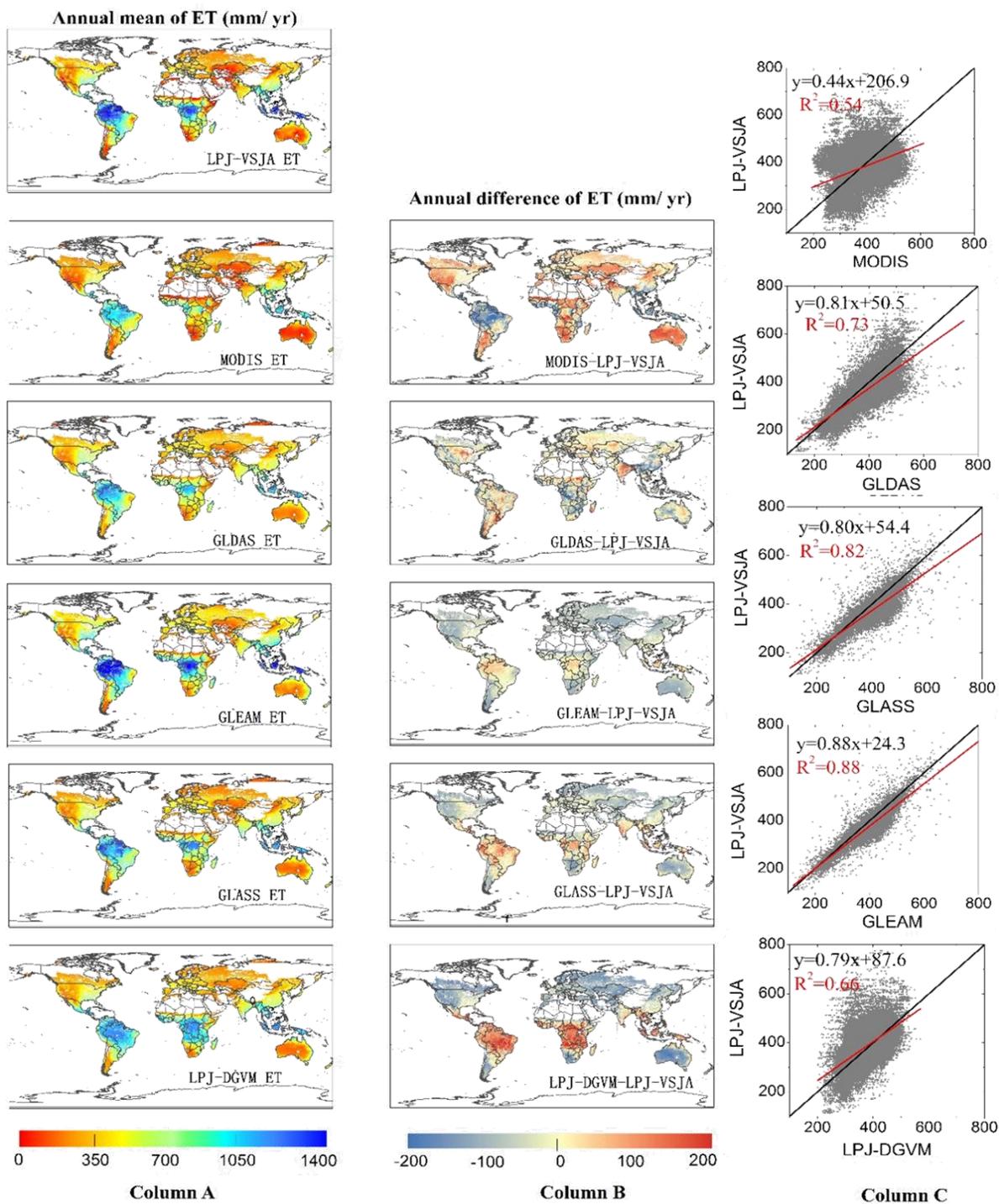



Figure 12. Column A: Spatial distribution of annual LPJ-VSJA ET and other independent satellite-based datasets (a: MODIS GPP; b: GLDAS ET; c: GLEAM ET; d: GLASS ET; e: LPJ-DGVM ET). Column B: Spatial distribution of the difference between annual LPJ-VSJA ET and other independent satellite-based datasets. Column C: Scatter plots between these products are provided on the right of the difference maps. Black lines show the 1:1-line, red lines show the regression fit.

Figure 13 (a)–(e) represent the spatial error standard deviation (σ) distribution of MODIS, GLASS, GOSIF, and LPJ-VSJA GPP, respectively. The graphs on the right side depict the corresponding histograms. The σ of the MODIS GPP was evenly distributed between 30 and 60 g C/m^2/month, while the average σ of other products was concentrated in 0–20 g C m^{-2} month^{-1} (90%). The high errors of all products were concentrated in the high temperature and humid areas of southern North America, eastern South America, humid and dry sub-humid areas of South Asia, and the savannas of Africa and Australia. The error histogram of GOSIF GPP and LPJ-DGVM GPP were in line with the normal distribution, with an average value of 8.3 g C m^{-2} month^{-1} and 22.4 g C m^{-2} month^{-1}. The GLASS GPP product had the lowest mean value (3.6 g C m^{-2} month^{-1}), followed by LPJ-VSJA (4.7 g C m^{-2} month^{-1}), but the error variance of the LPJ-VSJA product was the lowest, indicating a stability of the regional error (Table S4). Compared to the LPJ-DGVM, the joint assimilation results showed improvement in all regions (the average error reduced by 17.7 g C m^{-2} month^{-1}), especially in the humid regions of South Asia, Australia, and the United States. Our LPJ-VSJA GPP was generally proven to have high accuracy and stability for spatial analysis and could provide a reference for other model products.

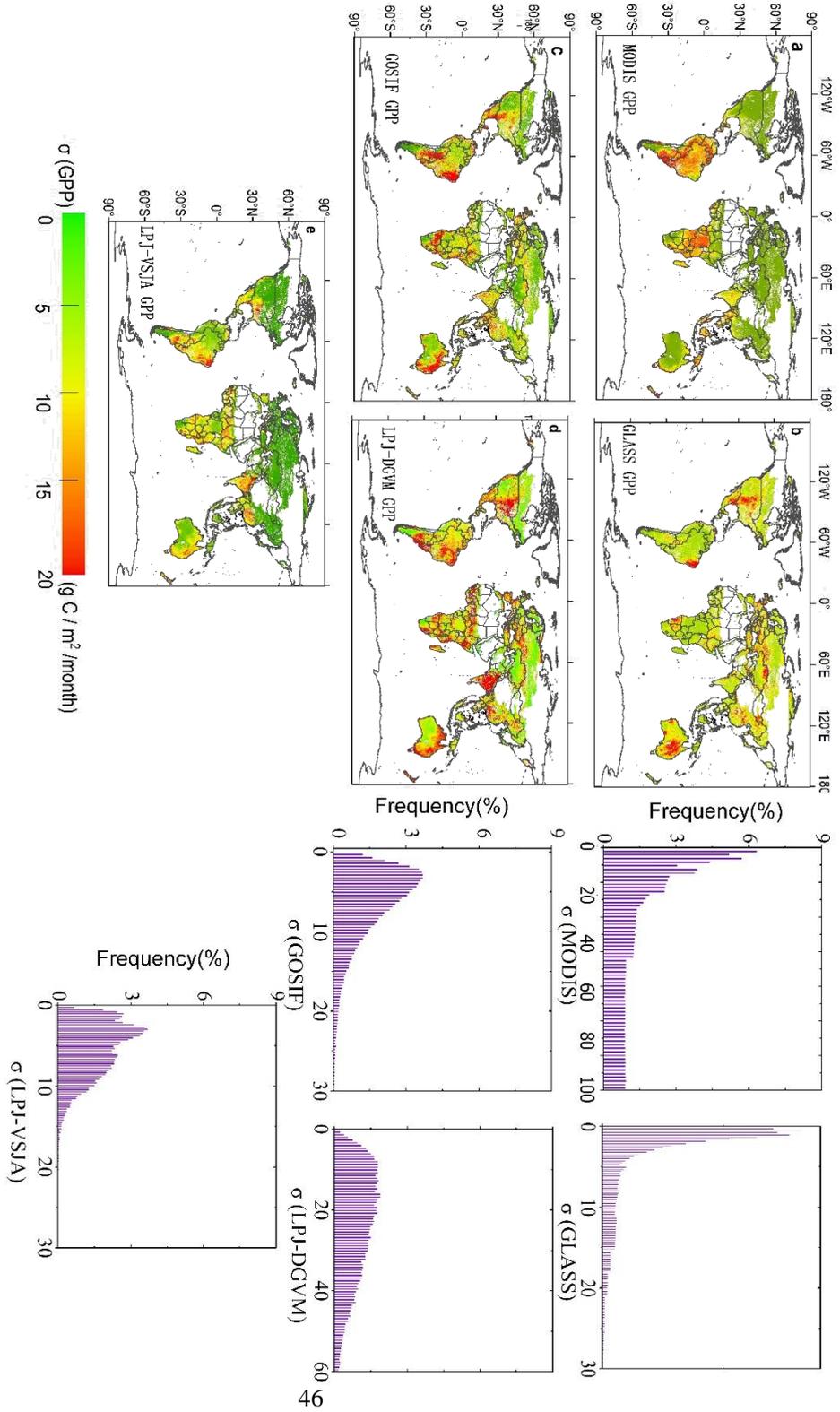
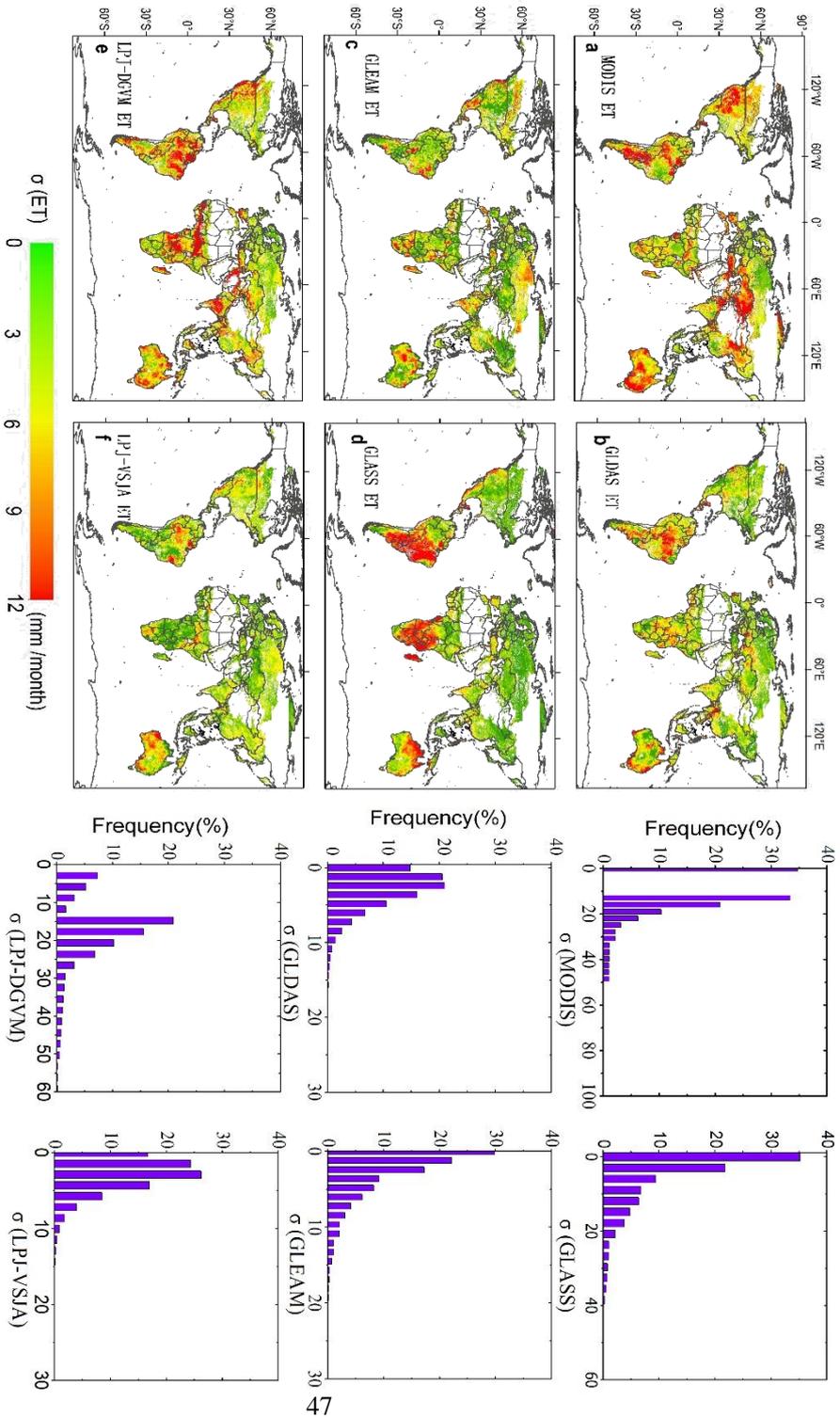



Figure 13. Spatial distribution and histograms of error standard deviation (σ) for global GPP products:

MODIS (a), GOSIF (b), GLASS (c), LPJ-DGVM (d), and LPJ-VSJA (e).

Figure14. Spatial distribution and histograms of error standard deviation (σ) for global ET products:

MODIS (a),GLDAS (b),GLEAM (c), GLASS (d), LPJ-DGVM (e), and LPJ-VSJA (f).

577   Figures 14 (a)–(f) show the σ of MODIS, GLDAS, GLEAM, GLASS, and LPJ-VSJA ET (the units

578   are mm/month), and the right graphs are the corresponding histograms. The σ values of GLDAS and LPJ-

579   VSJA represented a normal distribution trend. Except for MODIS, GLASS, and LPJ-DGVM (0–60 mm

580   month^{-1}), the σ of other products was generally between 0-20 mm month^{-1}. The simulation error was

581   relatively smaller in the Northern Hemisphere than in the Southern Hemisphere, especially for GLASS

582   ET and GLDAS ET. Significant improvements in joint assimilation were observed in the northern

583   hemisphere (especially in the semi-arid areas of the western United States and savanna and cropland areas

584   of central India) and African savanna areas, and the average error was reduced by 15.1 mm month^{-1}. In

585   general, the error mean and variance of LPJ-VSJA and GLEAM products were relatively low (Table S4),

586   and there was no apparent extremely high value region in the error distribution. Among the five products,

587   LPJ-VSJA had the lowest error mean and variance and the highest accuracy.

5. Discussion

5.1 Advantage of joint assimilation for GPP and ET

The benefit of employing multiple data flows in an assimilation system is the complementarity of the data, which enables constraints on different components of the underlying process-based terrestrial biosphere model. Due to the interaction and feedback between the internal components of the model, the assimilation of multiple observations has a synergistic effect, and the integrated constraints are greater than the individual constraint (Kato et al. (2013). The advantage of our joint assimilation is that it can improve the simulation accuracy of both GPP and ET, especially ET, in arid and semi-arid regions.

In the GPP assimilation experiment, the performance of the LAI assimilation was better than that of the SSM assimilation possibly for two reasons: (1) the LPJ-VSJA is more controlled by LAI data because the ratio of assimilated LAI (daily input) to SM observations (3-day interval input) is approximately 3:1, which makes the likelihood function biased to LAI data; (2) the SM directly influences the simulation of ET, and the corresponding time function (computes the top layer SM (50 cm)) used here (section 2.4) will result in the error of the updated top SM and propagating the error to the GPP_{SM}. Moreover, GPP is not only directly affected by LAI but also by other vegetation and environmental conditions. These multiple dependencies (direct and indirect) make GPP less likely to respond to transient changes in LAI (Zobitz et al. 2014). In addition, the LAI dynamic variability varies depending on the ecosystem type. For instance, the coniferous forest has less dynamic variability in LAI than in cropland and the deciduous forest (Turner et al. 2006; Turner et al. 2005). Therefore, the LAI assimilation with different time scales (e.g., 8-day interval, monthly, annual) for different ecosystems may achieve better results.

608   Current studies on terrestrial water and carbon flux assimilation mostly focus on the assimilation

609   between a single model framework and observation results, lacking the fusion and comparison between

610   multiple models. The processed models used in DA are simplifications and approximations of reality, and

611   different models focus on different ecological processes. In this study, the updated ET module was

612   integrated to compensate for the simplification of soil stratification and the lack of SM information in the

613   hydrological module of the LPJ-DGVM. Therefore, the integration of multiple types of models and multi-

614   source observation data (remotely sensed data, ecological inventory data (National Ecological

615   Observatory Network, NEON (Keller et al. 2008)), and other measurements (Desai et al. 2011; Hayes et

616   al. 2012) is expected to more objectively and effectively simulate the real state of ecosystems.

617   *5.2 Comparison of joint assimilation (LPJ-VSJA) and other models for GPP and ET across regions and*

618   *vegetation types*

619   Global GPP and ET for different products were calculated by multiplying the global mean GPP

620   density flux with the global vegetation area (122.4 million km^2) originated from the MODIS land cover

621   product (Friedl et al. 2010). The mean global GPP of the LPJ-VSJA (130.2 Pg C yr^{-1}) was

622   approximately 12% lower than that of PML-V2 (145.8 Pg C yr^{-1}) and 18% higher than that of GLASS

623   and MODIS, respectively (Table S6). The GPP values of LPJ-VSJA and GOSIF were the most similar.

624   The GOSIF GPP was developed from gridded SIF using simple linear relationships between SIF and

625   GPP. Our global LPJ-VSJA GPP estimates were within the currently most plausible 110–150 Pg C/yr

626   range.

627   As for ET, our results were similar to those of GLEAM ET and lower than those of PML-V2,

628   GLDAS-2, and GLASS ET (~72000 mm yr^{-1}). Joint assimilation improved the overestimation of LPJ-

629   DGVM ET. At the daily scale, the estimation accuracy of PML-V2 and GLDAS-2 products, calibrated

630   with flux tower data, was better than that of our estimates, which suggests an underestimation of LPJ-

631   VSJA ET in wet regions. It is likely because the SM of SMAP or SMOS was underestimated in the wet

632   region or the influence of deep SM was under-represented. According to Seneviratne et al. (2010),

633   satellite-based ET estimation approaches often overestimate ET in areas of arid and semi-arid climatic

634   regimes in the magnitude of 0.50 to 3.00 mm d^{-1}. The poor performance of these models can largely be

635   attributed to the lack of constraints of SM and more accurate vegetation parameters (Gokmen et al. 2012;

636   Pardo et al. 2014). For instance, the monthly estimated ET modeled by the Penman-Monteith-Leuning

637   (PML) model agreed with flux tower data well (R^2 = 0.77; bias = − 9.7%, approximately 0.2 mm d^{-1}).

638   Our annual ET simulations were lower than other products and slightly underestimated tower ET with a

639   bias of 0.19 mm d^{-1} (ET$_{OBS}$-ET$_{CO}$).

640   In general, GPP and ET had better assimilation performance in arid and semi-arid regions than in

641   humid and semi-humid regions likely because of the following reasons. First, the incorporation of surface

642   SM is more important for vegetation growth in water-limited areas. The module PT-JPL$_{SM}$ has been

643   proven to have better performance in semi-arid and arid regions (Purdy et al. 2018). Our integrated model

644   LPJ-PM also performed better in semi-arid and arid regions by assimilating SMAP soil moisture (Li et al.

645   2020). Second, the input performance, including SMOS and SMAP SM products, is better in arid and

646   temperate regions than in cold and humid regions (Zhang et al. 2019). Third, the vegetation types in humid

647   regions are more complex and relatively less accurately simulated by the LPJ-DGVM within a single grid

648   cell. For comparison, Zhang et al. (2020) used a data-driven upscaling approach to estimate GPP and ET

649   in global semi-arid regions. This data-driven approach ($R^2 = 0.79$, RMSD = 1.13 g C m^{-2} d^{-1}) had slightly

650   higher performance in estimating GPP than our LPJ-VSJA ($R^2 = 0.73$ and RMSD= 1.14 g C m^{-2} d^{-1}); the

651   data-driven method ($R^2 = 0.72$ and RMSD = 0.72mm d^{-1}) had identical performance for estimating ET

652   with our LPJ-VSJA($R^2 = 0.73$ and RMSD= 0.72 mm d^{-1}).

653   Our assimilation performance varied with PFT. The GPP and ET assimilation results of savanna sites

654   performed well in both dry and wet regions, and those of shrubland sites showed the most remarkable

655   improvement for simulations of LPJ-DGVM. The original simulation and assimilation performance of

656   grassland sites in the semi-arid and arid regions were the best for all five PFTs. Consistent with our

657   research, previous studies also showed better GPP or ET simulations for grassland, savannas, and

658   shrublands biomes. For instance, Feng et al. (2015) validated five satellite-based ET algorithms for semi-

659   arid ecosystems and concluded that all the models produced acceptable and relatively better results for

660   most grassland, savanna, and shrubland sites. Yang et al. (2017) demonstrated that he GLEAM ET had a

661   superior performance for the grassland sites. The GOSIF GPP demonstrated better simulation for

662   grassland and woody savannas sites at 8-day time steps with higher R^2 (0.77 and 0.83, respectively) and

663   lower RMSD (1.48 g C m^{-2} d^{-1} and 1.1 g C m^{-2} d^{-1}) (Li and Xiao 2019). In contrast, our LPJ-VSJA GPP

664   showed an R^2 of 0.87 for grassland and 0.75 for savannas and an RMSD of 1.11 g C m^{-2} d^{-1} and 1.1 g C

665   m^{-2} d^{-1}, respectively, in semi-arid and arid regions.

5.3 Uncertainty analysis of joint assimilation

Our validation results at both site and regional scales indicated that uncertainty existed in LPJ-VSJA daily GPP and ET estimates. The errors from the tower EC observations, model-driven data, model structure, error of satellite-based observations (e.g., LAI and SM), and the spatial scale mismatch between the ground observed footprint size and satellite-derived footprint size were the vital factors affecting assimilation performance. First, recent studies have revealed errors in the GLASS LAI and SMOS or SMAP SM compared with ground measurements. By computing the RMSD and R^2 of each product, the GLASS LAI accuracy was clearly superior to that of MODIS and Four-Scale Geometric Optical Model based LAI (FSGOM) in forests and GLASS and FSGOM led to in much higher annual GPP and ET estimates compared to MCD15(Liu et al. 2018). The vegetation type (or land cover) misclassification caused 15–50% differences in LAI retrieval (Fang and Liang 2005; Gonsamo and Chen 2011). Yan et al. (2016) calculated a RMSD of 0.18 for the GLASS LAI over a range of HeiHe drainage basin sites and used the error to improve the simulation of LAI and fluxes by assimilating GLASS LAI data. Previous studies reported an improvement in the performance of the SMOS and SMAP products (Lievens et al. 2015; Miernecki et al. 2014), which both provide an accuracy of 0.04 m^3 m^{-3} (Zhang et al. 2019). However, the actual observation error of these two products typically depends on the spatial location and time of the year (RMSD varying between 0.035 and 0.056 m^3 m^{-3} for several retrieval configurations) (Brocca et al. 2012). According to Purdy et al. (2018), the ET simulated by PT-JPL$_{SM}$ using the 9 km SM_L3_P_E data showed an inferior agreement (R^2= 0.47) but a relatively low RMSD (0.77 mm d^{-1}), due to the SMAP errors in the grid cell with soil heterogeneity and the climatological differences between model SM

686   forecasts and SMAP SM (Reichle and Koster 2004). We rescaled the ET_{PM} to the probability distribution

687   of the ET_{LPJ} through a cumulative distribution function (CDF) to correct the potential seasonal biases of

688   ET_{PM} before assimilation.

689   Second, there is large uncertainty in the influence of root zone SM as the source of water available

690   to plants (Albergel et al. 2008; Bonan et al. 2020). Our GPP results of cropland sites were largely

691   influenced by US-Ne1, an irrigate site. This site maintained high annual GPP in 2012 despite the drought

692   (Figure S4). However, the SMOS SM in 2012 had a lower surface SM annual mean than the site

693   observations likely because the detected soil layer (0-50 cm) of the site observation is deeper than that of

694   the satellite retrieval and the cumulative deep soil moisture due to the regular irrigation was higher than

695   the surface SM that could easily be vaporized during the drought period (Figure S4). Therefore, the

696   influence of deep SM of some cropland sites during the drought years induced large simulation errors and

697   unsatisfactory assimilation performance. Moreover, some deep-rooted forests maintain a high LAI during

698   drought by absorbing deep SM (>2 m) and groundwater (Zhang et al. 2016). Thus, joint assimilation of

699   the LAI and SM may eliminate a portion of the underestimation of GPP of such vegetation in drought

700   periods. Therefore, further research is needed on how to optimally utilize satellite SM data for improving

701   GPP and ET simulations.

702   Third, the problem of mixed pixels and mismatches in the observation footprints may also have an

703   influence on the accuracy of estimated GPP and ET. The 5 km spatial resolution of the GLASS LAI ,9

704   km of SMAP, and 25 km of SMOS products cannot capture the sub-grid-scale condition, especially in

705 grid cells for complex land surfaces or strong soil heterogeneity. To ensure the consistency of the grid-

706 cell representativeness for the LAI and SM, the interpolation result in errors that propagate through the

707 modeling and assimilation, causing the accumulation of output errors (Nijssen and Lettenmaier 2004).

708 Moreover, the shrubland in the LPJ-DGVM was most likely simulated as C4 grassland in the

709 hydrothermal condition of semi-arid and arid regions. In contrast, the shrubland tended to be hybrid

710 vegetation types (grassland mixed with other types of forest vegetation) in the hydrothermal condition of

711 humid and sub-dry humid regions, and the simulated canopy height is closer to the real condition of

712 shrubland. This might also be the reason for the superior performance of ET_{LPJ} and assimilation results

713 of shrubland sites in humid and sub-dry humid regions.

6. Conclusions

715 We developed an assimilation system LPJ-VSJA that integrates GLASS LAI, SMOS SM, and

716 SMAP SM data to improve GPP and ET estimates globally. The system was designed to assimilate two

717 SM products (SMOS and SMAP) into the integrated model - LPJ-PM for both dry and humid regions

718 through separate and joint assimilation. The results show that the joint constraints provided by vegetation

719 and soil variable strategies improve model simulations. Both the original and joint assimilation results for

720 GPP and ET in semi-arid and arid regions performed better than those in humid and dry-sub humid regions,

721 and the LPJ-PM that emphasized the SM information is more suitable for the water-limited regions. For

722 ET assimilation, the different SM products influence assimilation performance, and SMAP SM possesses

723 a slight advantage in most vegetation types and in both dry and humid regions. Our global LPJ-VSJA

GPP and ET products have relatively higher accuracy than other products, especially in water-limited regions with lower ET values.

Data availability

The LPJ-DGVM v4.1 version code (LPJ-ML) and example configurations are public available via the project homepage (https://github.com/PIK-LPJmL/LPJmL). We used the 3.01 version of LPJ-DGVM, which removed the agricultural management module. The access of all the input and validation dataset of assimilation system have been described in article. The assimilation method code configured by Fortran platform could be provided by contacting the X.T co-author. The modified code of LPJ-PM model and the underlying and global LPJ-VSJA GPP and ET data can be obtained by contacting the lead author of this manuscript.

Author contributions

S.L. and L.Z. designed the experiment and wrote the paper with support from all coauthors. S.L. and R.M. implemented the codes necessary for the experiments. J.X. contributed to the structure of the article and comparison of assimilation performance between the SMOS and SMAP experiments. X.T provided the POD-En4DVAR method and the code. M.Y contributed to the validation and analysis of the results. All the authors contributed to the synthesis of results and key conclusions.

Competing interests

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

Financial support

S.L., L.Z., R.M., and M.Y. were funded by the National Natural Science Foundation of China (Grant No. 41771392; PI Li Zhang) and (Grant No. 41901364; PI Min Yan).

References

Albergel, C., Rüdiger, C., Pellarin, T., Calvet, J.-C., Fritz, N., Froissard, F., Suquia, D., Petitpa, A., Piguet, B., & Martin, E. (2008). From near-surface to root-zone soil moisture using an exponential filter: an assessment of the method based on in-situ observations and model simulations. *Hydrology and Earth System Sciences, 12*, 1323-1337

Anav, A., Friedlingstein, P., Beer, C., Ciais, P., Harper, A., Jones, C., Murray‐Tortarolo, G., Papale, D., Parazoo, N.C., & Peylin, P. (2015). Spatiotemporal patterns of terrestrial gross primary production: A review. *Reviews of Geophysics, 53*, 785-818

Blyverket, J., Hamer, P.D., Bertino, L., Albergel, C., Fairbairn, D., & Lahoz, W.A. (2019). An Evaluation of the EnKF vs. EnOI and the Assimilation of SMAP, SMOS and ESA CCI Soil Moisture Data over the Contiguous US. *Remote Sensing, 11*, 478

Bonan, B., Albergel, C., Zheng, Y., Barbu, A.L., Fairbairn, D., Munier, S., & Calvet, J.-C. (2020). An ensemble square root filter for the joint assimilation of surface soil moisture and leaf area index within the Land Data Assimilation System LDAS-Monde: application over the Euro-Mediterranean region. *Hydrology and Earth System Sciences, 24*, 325-347

Bonan, G., Williams, M., Fisher, R., & Oleson, K. (2014). Modeling stomatal conductance in the earth system: linking leaf water-use efficiency and water transport along the soil–plant–atmosphere continuum. *Geoscientific Model Development, 7*, 2193-2222

Brocca, L., Tullo, T., Melone, F., Moramarco, T., & Morbidelli, R. (2012). Catchment scale soil moisture spatial–temporal variability. *Journal of hydrology, 422*, 63-75

Burgin, M.S., Colliander, A., Njoku, E.G., Chan, S.K., Cabot, F., Kerr, Y.H., Bindlish, R., Jackson, T.J., Entekhabi, D., & Yueh, S.H. (2017). A comparative study of the SMAP passive soil moisture product with existing satellite-based soil moisture products. *IEEE Transactions on Geoscience and Remote Sensing, 55*, 2959-2971

769 Caires, S., & Sterl, A. (2003). Validation of ocean wind and wave data using triple collocation. *Journal of geophysical*
770 *research: oceans, 108*

771 Chan, S.K., Bindlish, R., O'Neill, P.E., Njoku, E., Jackson, T., Colliander, A., Chen, F., Burgin, M., Dunbar, S., &
772 Piepmeier, J. (2016). Assessment of the SMAP passive soil moisture product. *IEEE Transactions on Geoscience and*
773 *Remote Sensing, 54*, 4994-5007

774 Cui, C., Xu, J., Zeng, J., Chen, K.-S., Bai, X., Lu, H., Chen, Q., & Zhao, T. (2018). Soil moisture mapping from satellites:
775 An intercomparison of SMAP, SMOS, FY3B, AMSR2, and ESA CCI over two dense network regions at different spatial
776 scales. *Remote Sensing, 10*, 33

777 Desai, A.R., Moore, D.J., Ahue, W.K., Wilkes, P.T., De Wekker, S.F., Brooks, B.G., Campos, T.L., Stephens, B.B.,
778 Monson, R.K., & Burns, S.P. (2011). Seasonal pattern of regional carbon balance in the central Rocky Mountains from
779 surface and airborne measurements. *Journal of Geophysical Research: Biogeosciences, 116*

780 Draper, C., Mahfouf, J.-F., Calvet, J.-C., Martin, E., & Wagner, W. (2011). Assimilation of ASCAT near-surface soil
781 moisture into the SIM hydrological model over France. *Hydrology and Earth System Sciences, 15*, 3829-3841

782 Entekhabi, D., Njoku, E.G., O'Neill, P.E., Kellogg, K.H., Crow, W.T., Edelstein, W.N., Entin, J.K., Goodman, S.D.,
783 Jackson, T.J., & Johnson, J. (2010). The soil moisture active passive (SMAP) mission. *Proceedings of the IEEE, 98*,
784 704-716

785 Etheridge, D.M., Steele, L., Langenfelds, R.L., Francey, R.J., Barnola, J.M., & Morgan, V. (1996). Natural and
786 anthropogenic changes in atmospheric CO2 over the last 1000 years from air in Antarctic ice and firn. *Journal of*
787 *Geophysical Research: Atmospheres, 101*, 4115-4128

788 Evensen, G. (2004). Sampling strategies and square root analysis schemes for the EnKF. *Ocean dynamics, 54*, 539-560

789 Fang, H., Baret, F., Plummer, S., & Schaepman‐Strub, G. (2019). An overview of global leaf area index (LAI): Methods,
790 products, validation, and applications. *Reviews of Geophysics, 57*, 739-799

791 Fang, H., Beaudoing, H.K., Rodell, M., Teng, W.L., & Vollmer, B.E. (2009). Global Land data assimilation system
792 (GLDAS) products, services and application from NASA hydrology data and information services center (HDISC). In,
793 *ASPRS 2009 Annual Conference, Baltimore, Maryland* (pp. 8-13)

794 Fang, H., & Liang, S. (2005). A hybrid inversion method for mapping leaf area index from MODIS data: Experiments
795 and application to broadleaf and needleleaf canopies. *Remote Sensing of Environment, 94*, 405-424

796 Feng, F., Chen, J., Li, X., Yao, Y., Liang, S., Liu, M., Zhang, N., Guo, Y., Yu, J., & Sun, M. (2015). Validity of five
797 satellite-based latent heat flux algorithms for semi-arid ecosystems. *Remote Sensing, 7*, 16733-16755

798 Friedl, M.A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., & Huang, X. (2010). MODIS
799 Collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote Sensing of*
800 *Environment, 114*, 168-182

801 Gokmen, M., Vekerdy, Z., Verhoef, A., Verhoef, W., Batelaan, O., & Van der Tol, C. (2012). Integration of soil moisture
802 in SEBS for improving evapotranspiration estimation under water stress conditions. *Remote Sensing of Environment,*
803 *121*, 261-274

804 Gonsamo, A., & Chen, J.M. (2011). Evaluation of the GLC2000 and NALC2005 land cover products for LAI retrieval
805 over Canada. *Canadian Journal of Remote Sensing, 37*, 302-313

Haxeltine, A., & Prentice, I.C. (1996). BIOME3: An equilibrium terrestrial biosphere model based on ecophysiological constraints, resource availability, and competition among plant functional types. *Global biogeochemical cycles, 10*, 693-709

Hayes, D.J., Turner, D.P., Stinson, G., McGuire, A.D., Wei, Y., West, T.O., Heath, L.S., De Jong, B., McConkey, B.G., & Birdsey, R.A. (2012). Reconciling estimates of the contemporary North American carbon balance among terrestrial biosphere models, atmospheric inversions, and a new approach for estimating net ecosystem exchange from inventory‐based data. *Global Change Biology, 18*, 1282-1299

He, L., Chen, J.M., Liu, J., Bélair, S., & Luo, X. (2017). Assessment of SMAP soil moisture for global simulation of gross primary production. *Journal of Geophysical Research: Biogeosciences, 122*, 1549-1563

Huang, C., Li, Y., Gu, J., Lu, L., & Li, X. (2015). Improving estimation of evapotranspiration under water-limited conditions based on SEBS and MODIS data in arid regions. *Remote Sensing, 7*, 16795-16814

Ines, A.V., Das, N.N., Hansen, J.W., & Njoku, E.G. (2013). Assimilation of remotely sensed soil moisture and vegetation with a crop simulation model for maize yield prediction. *Remote Sensing of Environment, 138*, 149-164

Jacquette, E., Al Bitar, A., Mialon, A., Kerr, Y., Quesney, A., Cabot, F., & Richaume, P. (2010). SMOS CATDS level 3 global products over land. In, *Remote Sensing for Agriculture, Ecosystems, and Hydrology XII* (p. 78240K): International Society for Optics and Photonics

Kaminski, T., Scholze, M., Vossbeck, M., Knorr, W., Buchwitz, M., & Reuter, M. (2017). Constraining a terrestrial biosphere model with remotely sensed atmospheric carbon dioxide. *Remote Sensing of Environment, 203*, 109-124

Kato, T., Knorr, W., Scholze, M., Veenendaal, E., Kaminski, T., Kattge, J., & Gobron, N. (2013). Simultaneous assimilation of satellite and eddy covariance data for improving terrestrial water and carbon simulations at a semi-arid woodland site in Botswana. *Biogeosciences, 10*, 789-802

Keeling, C.D., Whorf, T.P., Wahlen, M., & Van der Plichtt, J. (1995). Interannual extremes in the rate of rise of atmospheric carbon dioxide since 1980. *Nature, 375*, 666-670

Keller, M., Schimel, D.S., Hargrove, W.W., & Hoffman, F.M. (2008). A continental strategy for the National Ecological Observatory Network. *Frontiers in Ecology and the Environment, 6*, 282-284

Kganyago, M., Mhangara, P., Alexandridis, T., Laneve, G., Ovakoglou, G., & Mashiyi, N. (2020). Validation of sentinel-2 leaf area index (LAI) product derived from SNAP toolbox and its comparison with global LAI products in an African semi-arid agricultural landscape. *Remote Sensing Letters, 11*, 883-892

Khan, M.S., Liaqat, U.W., Baik, J., & Choi, M. (2018). Stand-alone uncertainty characterization of GLEAM, GLDAS and MOD16 evapotranspiration products using an extended triple collocation approach. *Agricultural and Forest Meteorology, 252*, 256-268

Kim, H., Parinussa, R., Konings, A.G., Wagner, W., Cosh, M.H., Lakshmi, V., Zohaib, M., & Choi, M. (2018). Global-scale assessment and combination of SMAP with ASCAT (active) and AMSR2 (passive) soil moisture products. *Remote Sensing of Environment, 204*, 260-275

Koster, R.D., Crow, W.T., Reichle, R.H., & Mahanama, S.P. (2018). Estimating basin‐scale water budgets with SMAP soil moisture data. *Water resources research, 54*, 4228-4244

842 Law, B., Falge, E., Gu, L.v., Baldocchi, D., Bakwin, P., Berbigier, P., Davis, K., Dolman, A., Falk, M., & Fuentes, J.
843 (2002). Environmental controls over carbon dioxide and water vapor exchange of terrestrial vegetation. *Agricultural and*
844 *Forest Meteorology, 113*, 97-120

845 Lee, H., Seo, D.-J., & Koren, V. (2011). Assimilation of streamflow and in situ soil moisture data into operational
846 distributed hydrologic models: Effects of uncertainties in the data and initial model soil moisture states. *Advances in*
847 *water resources, 34*, 1597-1615

848 Li, B., & Rodell, M. (2013). Spatial variability and its scale dependency of observed and modeled soil moisture over
849 different climate regions. *Hydrology and Earth System Sciences, 17*, 1177-1188

850 Li, S., Wang, G., Sun, S., Chen, H., Bai, P., Zhou, S., Huang, Y., Wang, J., & Deng, P. (2018). Assessment of multi-
851 source evapotranspiration products over china using eddy covariance observations. *Remote Sensing, 10*, 1692

852 Li, S., Zhang, L., Ma, R., Yan, M., & Tian, X. (2020). Improved ET assimilation through incorporating SMAP soil
853 moisture observations using a coupled process model: A study of US arid and semiarid regions. *Journal of hydrology,*
854 *590*, 125402

855 Li, X., Cheng, G., Liu, S., Xiao, Q., Ma, M., Jin, R., Che, T., Liu, Q., Wang, W., & Qi, Y. (2013). Heihe watershed allied
856 telemetry experimental research (HiWATER): Scientific objectives and experimental design. *Bulletin of the American*
857 *Meteorological Society, 94*, 1145-1160

858 Li, X., Mao, F., Du, H., Zhou, G., Xu, X., Han, N., Sun, S., Gao, G., & Chen, L. (2017). Assimilating leaf area index of
859 three typical types of subtropical forest in China from MODIS time series data based on the integrated ensemble Kalman
860 filter and PROSAIL model. *ISPRS Journal of Photogrammetry and Remote Sensing, 126*, 68-78

861 Li, X., & Xiao, J. (2019). A global, 0.05-degree product of solar-induced chlorophyll fluorescence derived from OCO-
862 2, MODIS, and reanalysis data. *Remote Sensing, 11*, 517

863 Liang, S., Zhao, X., Liu, S., Yuan, W., Cheng, X., Xiao, Z., Zhang, X., Liu, Q., Cheng, J., & Tang, H. (2013). A long-
864 term Global LAnd Surface Satellite (GLASS) data-set for environmental studies. *International Journal of Digital Earth,*
865 *6*, 5-33

866 Lievens, H., Tomer, S.K., Al Bitar, A., De Lannoy, G.J., Drusch, M., Dumedah, G., Franssen, H.-J.H., Kerr, Y.H.,
867 Martens, B., & Pan, M. (2015). SMOS soil moisture assimilation for improved hydrologic simulation in the Murray
868 Darling Basin, Australia. *Remote Sensing of Environment, 168*, 146-162

869 Ling, X.-L., Fu, C.-B., Yang, Z.-L., & Guo, W.-D. (2019). Comparison of different sequential assimilation algorithms
870 for satellite-derived leaf area index using the Data Assimilation Research Testbed (version Lanai). *Geoscientific Model*
871 *Development, 12*, 3119-3133

872 Liu, L., Gudmundsson, L., Hauser, M., Qin, D., Li, S., & Seneviratne, S.I. (2020). Soil moisture dominates dryness stress
873 on ecosystem production globally. *Nature communications, 11*, 1-9

874 Liu, Y., Xiao, J., Ju, W., Zhu, G., Wu, X., Fan, W., Li, D., & Zhou, Y. (2018). Satellite-derived LAI products exhibit
875 large discrepancies and can lead to substantial uncertainty in simulated carbon and water fluxes. *Remote Sensing of*
876 *Environment, 206*, 174-188

877 Ma, H., Huang, J., Zhu, D., Liu, J., Su, W., Zhang, C., & Fan, J. (2013). Estimating regional winter wheat yield by
878 assimilation of time series of HJ-1 CCD NDVI into WOFOST–ACRM model with Ensemble Kalman Filter.
879 *Mathematical and Computer Modelling, 58*, 759-770

880 Ma, R., Zhang, L., Tian, X., Zhang, J., Yuan, W., Zheng, Y., Zhao, X., & Kato, T. (2017). Assimilation of remotely-
881 sensed leaf area index into a dynamic vegetation model for gross primary productivity estimation. *Remote Sensing, 9*,
882 188

883 MacBean, N., Peylin, P., Chevallier, F., Scholze, M., & Schürmann, G. (2016). Consistent assimilation of multiple data
884 streams in a carbon cycle data assimilation system. *Geoscientific Model Development, 9*, 3569-3588

885 Martens, B., Miralles, D.G., Lievens, H., Schalie, R.v.d., De Jeu, R.A., Fernández-Prieto, D., Beck, H.E., Dorigo, W.A.,
886 & Verhoest, N.E. (2017). GLEAM v3: Satellite-based land evaporation and root-zone soil moisture. *Geoscientific Model
887 Development, 10*, 1903-1925

888 Miernecki, M., Wigneron, J.-P., Lopez-Baeza, E., Kerr, Y., De Jeu, R., De Lannoy, G.J., Jackson, T.J., O'Neill, P.E.,
889 Schwank, M., & Moran, R.F. (2014). Comparison of SMOS and SMAP soil moisture retrieval approaches using tower-
890 based radiometer data over a vineyard field. *Remote Sensing of Environment, 154*, 89-101

891 Miralles, D.G., Jiménez, C., Jung, M., Michel, D., Ershadi, A., McCabe, M., Hirschi, M., Martens, B., Dolman, A.J., &
892 Fisher, J.B. (2016). The WACMOS-ET project–Part 2: Evaluation of global terrestrial evaporation data sets. *Hydrology
893 and Earth System Sciences, 20*, 823-842

894 Mitchell, H.L., Houtekamer, P.L., & Pellerin, G. (2002). Ensemble size, balance, and model-error representation in an
895 ensemble Kalman filter. *Monthly weather review, 130*, 2791-2808

896 Mu, Q., Zhao, M., Heinsch, F.A., Liu, M., Tian, H., & Running, S.W. (2007). Evaluating water stress controls on primary
897 production in biogeochemical and remote sensing based models. *Journal of Geophysical Research: Biogeosciences, 112*

898 New, M., Hulme, M., & Jones, P. (2000). Representing twentieth-century space–time climate variability. Part II:
899 Development of 1901–96 monthly grids of terrestrial surface climate. *Journal of climate, 13*, 2217-2238

900 Nijssen, B., & Lettenmaier, D.P. (2004). Effect of precipitation sampling error on simulated hydrological fluxes and
901 states: Anticipating the Global Precipitation Measurement satellites. *Journal of Geophysical Research: Atmospheres,
902 109*

903 O'Neill, P., Entekhabi, D., Njoku, E., & Kellogg, K. (2010). The NASA soil moisture active passive (SMAP) mission:
904 Overview. In, *2010 IEEE International Geoscience and Remote Sensing Symposium* (pp. 3236-3239): IEEE

905 O'Carroll, A.G., Eyre, J.R., & Saunders, R.W. (2008). Three-way error analysis between AATSR, AMSR-E, and in situ
906 sea surface temperature observations. *Journal of atmospheric and oceanic technology, 25*, 1197-1207

907 Pardo, N., Sánchez, M.L., Timmermans, J., Su, Z., Pérez, I.A., & García, M.A. (2014). SEBS validation in a Spanish
908 rotating crop. *Agricultural and Forest Meteorology, 195*, 132-142

909 Petropoulos, G.P., Ireland, G., & Barrett, B. (2015). Surface soil moisture retrievals from remote sensing: Current status,
910 products & future trends. *Physics and Chemistry of the Earth, Parts A/B/C, 83*, 36-56

911 Pipunic, R., Walker, J., & Western, A. (2008). Assimilation of remotely sensed data for improved latent and sensible
912 heat flux prediction: A comparative synthetic study. *Remote Sensing of Environment, 112*, 1295-1305

913 Purdy, A.J., Fisher, J.B., Goulden, M.L., Colliander, A., Halverson, G., Tu, K., & Famiglietti, J.S. (2018). SMAP soil
914 moisture improves global evapotranspiration. *Remote Sensing of Environment, 219*, 1-14

915 Rüdiger, C., Albergel, C., Mahfouf, J.F., Calvet, J.C., & Walker, J.P. (2010). Evaluation of the observation operator
916 Jacobian for leaf area index data assimilation with an extended Kalman filter. *Journal of Geophysical Research:
917 Atmospheres, 115*

918 Reichle, R.H., De Lannoy, G.J., Liu, Q., Koster, R.D., Kimball, J.S., Crow, W.T., Ardizzone, J.V., Chakraborty, P.,
919 Collins, D.W., & Conaty, A.L. (2017). Global assessment of the SMAP level-4 surface and root-zone soil moisture
920 product using assimilation diagnostics. *Journal of Hydrometeorology, 18*, 3217-3237

921 Reichle, R.H., & Koster, R.D. (2004). Bias reduction in short records of satellite soil moisture. *Geophysical Research
922 Letters, 31*

923 Rienecker, M.M., Suarez, M.J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M.G., Schubert, S.D., Takacs,
924 L., & Kim, G.-K. (2011). MERRA: NASA's modern-era retrospective analysis for research and applications. *Journal of
925 climate, 24*, 3624-3648

926 Running, S.W., Nemani, R.R., Heinsch, F.A., Zhao, M., Reeves, M., & Hashimoto, H. (2004). A continuous satellite-
927 derived measure of global terrestrial primary production. *Bioscience, 54*, 547-560

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

930 Sitch, S., Smith, B., Prentice, I.C., Arneth, A., Bondeau, A., Cramer, W., Kaplan, J.O., Levis, S., Lucht, W., & Sykes,
931 M.T. (2003). Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic
932 global vegetation model. *Global Change Biology, 9*, 161-185

933 Stoffelen, A. (1998). Toward the true near‐surface wind speed: Error modeling and calibration using triple collocation.
934 *Journal of geophysical research: oceans, 103*, 7755-7766

935 Sun, P., Wu, Y., Xiao, J., Hui, J., Hu, J., Zhao, F., Qiu, L., & Liu, S. (2019). Remote sensing and modeling fusion for
936 investigating the ecosystem water-carbon coupling processes. *Science of the total environment, 697*, 134064

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

939 Tian, S., Renzullo, L.J., Van Dijk, A.I., Tregoning, P., & Walker, J.P. (2019). Global joint assimilation of GRACE and
940 SMOS for improved estimation of root-zone soil moisture and vegetation response. *Hydrology and Earth System
941 Sciences, 23*, 1067-1081

942 Tian, X., & Feng, X. (2015). A non-linear least squares enhanced POD-4DVar algorithm for data assimilation. *Tellus A:
943 Dynamic Meteorology and Oceanography, 67*, 25340

944 Tian, X., Xie, Z., Dai, A., Jia, B., & Shi, C. (2010). A microwave land data assimilation system: Scheme and preliminary
945 evaluation over China. *Journal of Geophysical Research: Atmospheres, 115*

946 Tian, X., Xie, Z., Dai, A., Shi, C., Jia, B., Chen, F., & Yang, K. (2009). A dual‐pass variational data assimilation
947 framework for estimating soil moisture profiles from AMSR‐E microwave brightness temperature. *Journal of
948 Geophysical Research: Atmospheres, 114*

949 Tian, X., Xie, Z., Liu, Y., Cai, Z., Fu, Y., Zhang, H., & Feng, L. (2014). A joint data assimilation system (Tan-Tracker)
950 to simultaneously estimate surface CO_2 fluxes and 3-D atmospheric CO_2 concentrations from observations.
951 *Atmospheric Chemistry and Physics, 14*, 13281-13293

952 Tian, X., Xie, Z., & Sun, Q. (2011). A POD-based ensemble four-dimensional variational assimilation method. *Tellus
953 A: Dynamic Meteorology and Oceanography, 63*, 805-816

Trugman, A., Medvigy, D., Mankin, J., & Anderegg, W. (2018). Soil moisture stress as a major driver of carbon cycle uncertainty. *Geophysical Research Letters, 45*, 6495-6503

Turner, D.P., Ritts, W.D., Cohen, W.B., Gower, S.T., Running, S.W., Zhao, M., Costa, M.H., Kirschbaum, A.A., Ham, J.M., & Saleska, S.R. (2006). Evaluation of MODIS NPP and GPP products across multiple biomes. *Remote Sensing of Environment, 102*, 282-292

Turner, D.P., Ritts, W.D., Cohen, W.B., Maeirsperger, T.K., Gower, S.T., Kirschbaum, A.A., Running, S.W., Zhao, M., Wofsy, S.C., & Dunn, A.L. (2005). Site‐level evaluation of satellite‐based global terrestrial gross primary production and net primary production monitoring. *Global Change Biology, 11*, 666-684

Twine, T.E., Kustas, W., Norman, J., Cook, D., Houser, P., Meyers, T., Prueger, J., Starks, P., & Wesely, M. (2000). Correcting eddy-covariance flux underestimates over a grassland. *Agricultural and Forest Meteorology, 103*, 279-300

Wang, L., Zhu, H., Lin, A., Zou, L., Qin, W., & Du, Q. (2017). Evaluation of the latest MODIS GPP products across multiple biomes using global eddy covariance flux data. *Remote Sensing, 9*, 418

Waring, R.H., & Running, S.W. (2010). *Forest ecosystems: analysis at multiple scales*. Elsevier

Wieder, W., Boehnert, J., Bonan, G., & Langseth, M. (2014). Regridded harmonized world soil database v1. 2. *ORNL DAAC*

Xiao, J., Chevallier, F., Gomez, C., Guanter, L., Hicke, J.A., Huete, A.R., Ichii, K., Ni, W., Pang, Y., & Rahman, A.F. (2019). Remote sensing of the terrestrial carbon cycle: A review of advances over 50 years. *Remote Sensing of Environment, 233*, 111383

Xiao, Z., Liang, S., & Jiang, B. (2017). Evaluation of four long time-series global leaf area index products. *Agricultural and Forest Meteorology, 246*, 218-230

Xiao, Z., Liang, S., Wang, J., Chen, P., Yin, X., Zhang, L., & Song, J. (2013). Use of general regression neural networks for generating the GLASS leaf area index product from time-series MODIS surface reflectance. *IEEE Transactions on Geoscience and Remote Sensing, 52*, 209-223

Xiao, Z., Liang, S., Wang, J., Xiang, Y., Zhao, X., & Song, J. (2016). Long-time-series global land surface satellite leaf area index product derived from MODIS and AVHRR surface reflectance. *IEEE Transactions on Geoscience and Remote Sensing, 54*, 5301-5318

Yan, M., Tian, X., Li, Z., Chen, E., Wang, X., Han, Z., & Sun, H. (2016). Simulation of forest carbon fluxes using model incorporation and data assimilation. *Remote Sensing, 8*, 567

Yang, W., Wang, Y., Liu, X., Zhao, H., Shao, R., & Wang, G. (2020). Evaluation of the rescaled complementary principle in the estimation of evaporation on the Tibetan Plateau. *Science of the total environment, 699*, 134367

Yang, X., Yong, B., Ren, L., Zhang, Y., & Long, D. (2017). Multi-scale validation of GLEAM evapotranspiration products over China via ChinaFLUX ET measurements. *International Journal of Remote Sensing, 38*, 5688-5709

Yilmaz, M.T., & Crow, W.T. (2014). Evaluation of assumptions in soil moisture triple collocation analysis. *Journal of Hydrometeorology, 15*, 1293-1302

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

991 Zhang, D.-H., Li, X.-R., Zhang, F., Zhang, Z.-S., & Chen, Y.-L. (2016). Effects of rainfall intensity and intermittency
992 on woody vegetation cover and deep soil moisture in dryland ecosystems. *Journal of hydrology, 543*, 270-282

993 Zhang, F., & Weng, Y. (2015). Predicting hurricane intensity and associated hazards: A five-year real-time forecast
994 experiment with assimilation of airborne Doppler radar observations. *Bulletin of the American Meteorological Society,*
995 *96*, 25-33

996 Zhang, L., Xiao, J., Zheng, Y., Li, S., & Zhou, Y. (2020). Increased carbon uptake and water use efficiency in global
997 semi-arid ecosystems. *Environmental Research Letters, 15*, 034022

998 Zhang, R., Kim, S., & Sharma, A. (2019). A comprehensive validation of the SMAP Enhanced Level-3 Soil Moisture
999 product using ground measurements over varied climates and landscapes. *Remote Sensing of Environment, 223*, 82-94

1000 Zhao, L., Xia, J., Xu, C.-y., Wang, Z., Sobkowiak, L., & Long, C. (2013). Evapotranspiration estimation methods in
1001 hydrological models. *Journal of Geographical Sciences, 23*, 359-369

1002 Zobitz, J., Moore, D.J., Quaife, T., Braswell, B.H., Bergeson, A., Anthony, J.A., & Monson, R.K. (2014). Joint data
1003 assimilation of satellite reflectance and net ecosystem exchange data constrains ecosystem carbon fluxes at a high-
1004 elevation subalpine forest. *Agricultural and Forest Meteorology, 195*, 73-88

1005 Zou, L., Zhan, C., Xia, J., Wang, T., & Gippel, C.J. (2017). Implementation of evapotranspiration data assimilation with
1006 catchment scale distributed hydrological model via an ensemble Kalman filter. *Journal of hydrology, 549*, 685-702

1007