Evaluation of remote sensing-based evapotranspiration products at low-latitude eddy covariance sites

Diego Salazar-Martínez[a], Friso Holwerda[b,*], Thomas R.H. Holmes[c], Enrico A. Yépez[d], Christopher R. Hain[e], Susana Alvarado-Barrientos[f], Gregorio Ángeles-Pérez[g], Tulio Arredondo-Moreno[h], Josué Delgado-Balbuena[i], Bernardo Figueroa-Espinoza[j], Jaime Garatuza-Payán[d], Eugenia González del Castillo[b], Julio C. Rodríguez[k], Nidia E. Rojas-Robles[d], Jorge M. Uuh-Sonda[j], Enrique R. Vivoni[l]

[a] Posgrado en Ciencias de la Tierra, Universidad Nacional Autónoma de México, Ciudad de México, Mexico
[b] Instituto de Ciencias de la Atmósfera y Cambio Climático, Universidad Nacional Autónoma de México, Ciudad de México, Mexico
[c] Hydrological Sciences Lab, NASA Goddard Space Flight Center, Greenbelt, MD, USA
[d] Departamento de Ciencias del Agua y Medio Ambiente, Instituto Tecnológico de Sonora, Ciudad Obregón, Sonora, Mexico
[e] Earth Science Office, NASA Marshall Space Flight Center, Huntsville, AL, USA
[f] Red de Ecología Funcional, Instituto de Ecología, A.C., Xalapa, Veracruz, Mexico
[g] Posgrado en Ciencias Forestales, Campus Montecillo, Colegio de Postgraduados, Texcoco, Estado de México, Mexico
[h] División de Ciencias Ambientales, Instituto Potosino de Investigación Científica y Tecnológica, A.C., San Luís Potosí, Mexico
[i] Centro Nacional de Investigación Disciplinaria Agricultura Familiar, Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias, Ojuelos de Jalisco, Jalisco, Mexico
[j] Laboratorio de Ingeniería y Procesos Costeros, Instituto de Ingeniería, Universidad Nacional Autónoma de México, Sisal, Mexico
[k] Departamento de Agricultura y Ganadería, Universidad de Sonora, Hermosillo, Sonora, Mexico
[l] School of Earth and Space Exploration & School of Sustainable Engineering and the Built Environment, Arizona State University, Tempe, AZ, USA

* Corresponding author.

E-mail address: friso.holwerda@gmail.com (F. Holwerda).

Abstract

Remote sensing-based evapotranspiration (ET) products have been evaluated primarily using data from northern middle latitudes; therefore, little is known about their performance at low latitudes. To address this bias, an evaluation dataset was compiled using eddy covariance data from 40 sites between latitudes 30° S and 30° N. The flux data were obtained from the emerging network in Mexico (MexFlux) and from openly available databases of FLUXNET, AsiaFlux, and OzFlux. This unique reference dataset was then used to evaluate remote sensing-based ET products in environments that have been underrepresented in earlier studies. The evaluated products were: MODIS ET (MOD16, both the discontinued collection 5 (C5) and the latest collection (C6)), Global Land Evaporation Amsterdam Model (GLEAM) ET, and Atmosphere-Land Exchange Inverse (ALEXI) ET. Products were compared with unadjusted fluxes (ET_{orig}) and with fluxes corrected for the lack of energy balance closure (ET_{ebc}). Three common statistical metrics were used: coefficient of determination (R^2), root mean square error (RMSE), and percent bias (PBIAS). The effect of a vegetation mismatch between pixel and site on product evaluation results was investigated by examining the relationship between the statistical metrics and product-specific vegetation match indexes. Evaluation results of this study and those published in the literature were used to examine the performance of the products across latitudes. Differences between the MOD16 collection 5 and 6 datasets were generally smaller than differences with the other products. Performance and ranking of the evaluated products depended on whether ET_{orig} or ET_{ebc} was used. When using ET_{orig}, GLEAM generally had the highest R^2, smallest PBIAS, and best RMSE values across the studied land cover types and climate zones. Neither MOD16 nor ALEXI performed consistently better than the other. When using ET_{ebc}, none of the products stood out in terms of both low bias and strong correlations. The use of ET_{ebc}

instead of ET_{orig} affected the biases more than the correlations. The product evaluation results showed no significant relationship with the degree of match between the vegetation at the pixel and site scale. The latitudinal comparison showed tendencies of lower R^2 (all products) but better PBIAS and normalized RMSE values (MOD16 and GLEAM) for forests at low latitudes than for forests at northern middle latitudes. For non-forest vegetation, the products showed no clear latitudinal differences in performance.

Keywords: MOD16; GLEAM; ALEXI; tropics; subtropics

1. Introduction

The low latitudes (30° S–30° N) are characterized by large contrasts in terrestrial evapotranspiration (ET). They are home to tropical rainforests and other ecosystems with abundant rainfall where energy (radiation) is the main constraint to ET (see, for example, Fisher et al., 2009; Bruijnzeel et al., 2011). They are also home to tropical and subtropical arid and semi-arid ecosystems where ET is limited by water supply (see, for example, Eamus et al., 2013; Delgado-Balbuena et al., 2019). Estimates of ET for this region are needed for local and regional applications such as water resource management and drought monitoring and for global applications such as climate change studies (Wang and Dickinson, 2012; Fisher et al., 2017). The spatial and temporal scale of these applications require other methods than those used to study ET at the plot to ecosystem scale (i.e., lysimeter, sap flow, and micrometeorological methods). Recently, the potential of remote sensing-based ET estimates for these purposes has been recognized (Dolman et al., 2014; Fisher et al., 2017; Sheffield et al., 2018).

Since the 1990s, numerous remote sensing-based ET models have been developed (see Ke Zhang et al., 2016 for an overview). These models can be broadly divided into three categories (in no specific order): models based on the (1) Penman-Monteith (Monteith, 1965) or (2) Priestley-Taylor (Priestley and Taylor, 1972) equations and (3) models that determine the sensible heat flux (H) and calculate ET (or latent heat flux, LE) as the residual of the surface energy balance (i.e., so-called SEB models; Wang and Dickinson, 2012; Ke Zhang et al., 2016; Chen and Liu, 2020). Of interest to the user community is the development of global ET products from these models that are readily available to the public and regularly updated to include the latest data. Two such datasets have been produced since the early 2010s: 1) the MODIS ET product based on the MOD16 algorithm (Penman-Monteith type model; Mu et al., 2007, 2011; Running et al., 2019); and 2) the ET product from the Global Land Evaporation Amsterdam Model (GLEAM;

Priestley-Taylor type model; Miralles et al., 2011; Martens et al., 2017). More recently, global ET datasets based on the SEB model of Senay et al. (2013, 2020) and the Penman-Monteith-Leuning (PML) model of Yongqiang Zhang et al. (2016, 2019) have become available. Efforts are also underway to develop a global ET product based on the Atmosphere-Land Exchange Inverse (ALEXI) model (another SEB-based approach; Anderson et al., 2011; Hain and Anderson, 2017; Holmes et al., 2018).

Measurements of ET from eddy covariance flux towers have been used as the standard reference data against which remote sensing-based ET products are evaluated (Miralles et al., 2011; Mu et al., 2011; Holmes et al., 2018; Yongqiang Zhang et al., 2019; Senay et al., 2020), despite the problems related to the lack of energy balance closure observed at eddy covariance sites and the scale difference between the flux footprint and the model pixels (see below). Broadly speaking, two different types of evaluation studies can be distinguished: 1) those that evaluate the published ET datasets (hereafter referred to as product evaluation studies); and 2) those that evaluate the performance of the underlying models (model evaluation studies). In the latter type of study, all models are run with the same input data to isolate the effect of different modeling approaches from differences in forcing data (Vinukollu et al., 2011a, b; McCabe et al., 2016; Michel et al., 2016; Melo et al., 2021). Because remote sensing ET models are sensitive to changes in input data (Vinukollu et al., 2011b; Badgley et al., 2015), the results of model evaluation studies do not necessarily apply to the actual products.

The performance of remote sensing-based ET products at low latitudes is largely unknown because most evaluation studies have focused on the northern middle latitudes (30° N–60° N; Miralles et al., 2011; Mu et al., 2011; Kim et al., 2012; Hu et al., 2015; Velpuri et al., 2013; Tang et al., 2015; Reitz et al., 2017; Holmes et al., 2018; Khan et al., 2018). The few studies that evaluated ET products at low latitudes did this at a small number (two to five) of eddy covariance sites (Ruhoff et al., 2013; Ramoelo et al., 2014; Aguilar et al., 2018; Souza et al., 2019). The bias toward the northern middle latitudes can be explained by geographic differences in the availability of eddy covariance data (Schimel et al., 2015; Villareal and Vargas, 2021). Because of the lack of evaluation results from the low latitudes, it is unknown whether global remote sensing-based ET products perform equally well at all latitudes. One can think of several reasons why this might not be the case. For example, the MOD16 ET algorithm was calibrated using eddy covariance data from sites located primarily in the US and Canada (Mu et al., 2011). Hence, it is possible that the model is less accurate in other regions of the world, including the low latitudes (Kun Zhang et al., 2019). Similarly, GLEAM uses constant values for the Priestley-Taylor coefficient (α; Miralles et al., 2011), while α varies with climate (Shuttleworth, 1993) and forest type (Komatsu, 2005). Because the distribution of climate and forest types is related to latitude, the use of constant values for α may result in (apparent) latitude-dependent biases in ET. Latitudinal differences in product performance can also be caused by regional differences in input data quality (Vinukollu et al., 2011b) or cloud cover (Running et al., 2019).

While eddy covariance observations of ET are probably the best option to evaluate remote sensing datasets, there are two problems to consider: 1) the energy balance observed at eddy covariance sites is usually not closed; and 2) the footprint of the eddy covariance observations and the pixels of the ET products have different spatial scales. The degree of energy balance closure is quantified by the energy balance ratio (EBR), which is the ratio of turbulent energy fluxes (H + LE) to available energy, A (Wilson et al., 2002). Available energy is the difference between net radiation (R_n) and changes in energy storage. The average EBR observed at eddy covariance sites is about 0.8 (Wilson et al., 2002; Stoy et al., 2013). While the cause of the energy imbalance is still being investigated, there are several plausible explanations, including the systematic underestimation of the eddy covariance fluxes (Frank et al., 2016; Gao et al., 2017; Mauder et al., 2020). As a practical solution to the closure problem, the energy surplus is added to H and LE. Because it is unknown in what proportion the energy should be divided between the fluxes (Mauder et al., 2020), the surplus is usually distributed in proportion to the magnitude of H and LE, which preserves the Bowen ratio (Twine et al., 2000). Comparisons with independent estimates of ET have yielded contrasting results, with some studies finding better agreement for energy balance closure-corrected ET (Barr et al., 2012; Mauder et al., 2018) and others for unadjusted ET (Denager et al., 2020). Although the energy balance closure problem has been recognized for many years (Wilson et al., 2002; Foken et al., 2011), its effect on the evaluation results of remote sensing-based ET products or models has rarely been examined (Michel et al., 2016; Melton et al., 2021).

The evaluation results can also be affected by the scale difference between the footprint of the eddy covariance observations and the pixels of the ET products. The flux footprint is typically smaller than 1 km^2 (Chu et al., 2021), while the pixel sizes of ET products are as small as 0.25 km^2 (MOD16) and as large as 750 km^2 (GLEAM). The scale difference can result in a mismatch in vegetation between pixel and site (Hobeichi et al., 2018; Jiménez et al., 2018). Such a mismatch may also result from errors in the vegetation input data used by the models (due to, for example, incorrect classification). Because most models calculate ET using land cover-specific parameters (Anderson et al., 2007; Miralles et al., 2011; Mu et al., 2011), a mismatch between the actual vegetation of the observation site and that detected in the model pixel could potentially affect the evaluation results (Hu et al., 2015). However, the few studies that have examined this issue found no clear effect (Hobeichi et al., 2018; Jiménez et al., 2018).

The objectives of this study were to: 1) evaluate the performance of the MOD16 and GLEAM global ET products as well as of ET based on the ALEXI model at 40 eddy covariance sites in the low latitudes; 2) examine the effect of the energy balance closure problem on product evaluation results; 3) examine the dependence of product evaluation results on the vegetation-match between pixel and site; and 4) investigate potential latitudinal dependence of product performance. The MOD16 and GLEAM products were chosen because they are the longest regularly produced remote sensing-based ET datasets. From MOD16, both the discontinued collection 5 (C5) and the latest collection (C6) were evaluated (Mu et al., 2011; Running et al., 2019). In the case of GLEAM, the v3.3a dataset was evaluated (Martens et al., 2017). While most applications of ALEXI have focused on the continental US, recent efforts have paved the way for routine global implementation of ALEXI (Hain and Anderson, 2017). The reference dataset compiled in this study provides an excellent opportunity to evaluate the performance of ALEXI at low latitudes. The products were evaluated using a reference dataset of eddy covariance observations, including data from the emerging flux network in Mexico (MexFlux; Vargas et al., 2013; Delgado-Balbuena et al., 2018) and from openly available databases of FLUXNET (Pastorello et al., 2020), AsiaFlux, and OzFlux (Beringer et al., 2016).

2. Methods

2.1. Data

The remote sensing-based ET products evaluated in this study have different spatial and temporal resolutions (Table 1). The comparisons with the eddy covariance ET observations were made at the original spatial resolution of each product, except in the case of MOD16 C6 for which the 500-m data were resampled to a 1-km resolution to match MOD16 C5. Using the original spatial resolution is the common practice when evaluating these products against eddy covariance data (see references in Table S4). An exception was made for MOD16 C6 to allow for a more direct comparison with the previous C5 version. The effect of the scale mismatch between product pixel and flux footprint on the evaluation results was examined using the vegetation match index (Section 2.3). For each product, ET data were obtained from the pixels matching the location of the flux towers (Velpuri et al., 2013; Hu et al., 2015). To evaluate all products at the same temporal resolution (some performance statistics depend on the temporal resolution of the data), the daily GLEAM and ALEXI data were averaged over the 8-day MODIS interval. This was the highest common temporal resolution possible among the evaluated datasets. Likewise, the eddy covariance data were averaged to yield mean daily ET for each MODIS interval (Section 2.1.4).

The remote sensing ET products were evaluated by grouping the data by land cover type and climate zone (Section 2.2). The eddy covariance data from the various sites were collected during different periods between 2000 and 2019, with the length of the data records ranging from 1 to 11 years (Table 2). Hence, the flux datasets for a given land cover type or climate zone may not coincide in time. In addition, data availability varied among the evaluated products. MOD16 C5 was discontinued in 2015 and GLEAM data for 2019 were not available at the time of download (Table 1). For GLEAM and ALEXI, seven and four sites, respectively, were omitted from the analysis because the fraction of open water in the corresponding pixels was too high due to proximity to the coast (Sections 2.1.2 and 2.1.3). This problem did not affect MOD16 because of the smaller pixel size. As a result, the amount of data available for each of the comparisons by land cover type and climate zone often varied from product to product (Table 3). Ideally, one would compare the products using a common reference dataset (i.e., same sites and same MODIS intervals). However, this would reduce the amount of available data by about one-third (12 fewer sites and about 36% fewer MODIS intervals). Therefore, it was decided to perform the regression analysis of observations versus product estimates (Section 2.2.1) and the comparison of the performance statistics by land cover type and climate zone (Sections 2.2.2 and 2.2.3) using the complete dataset. The extent to which the two approaches (all data or a common reference dataset) may have influenced the results was examined through a sensitivity analysis (Section 2.2.4). The seasonal trend analysis (Section 2.2.5) was performed using the common reference dataset.

The MOD16 and GLEAM ET data were extracted from the published global ET datasets. Because detailed information about the models and datasets used to generate these products can be found in the references listed in Table 1, only a brief explanation is provided below. The ALEXI ET data were calculated specifically for this study. The methodology is described in Anderson et al. (2011) and Hain and Anderson (2017). For completeness, the main features of the model and the specific input datasets used are briefly described below.

<Table 1>

2.1.1. MODIS ET data

The MOD16 ET product is derived using a three-source Penman-Monteith model, which estimates ET as the sum of evaporation from the dry canopy (transpiration), wet canopy (interception loss), and soil (Mu et al., 2007, 2011). Separate calculations are performed for the day and night. The model uses MODIS retrievals of: albedo (for the calculation of R_n); fraction of absorbed photosynthetically active radiation, FPAR (to partition R_n between canopy and soil); land cover type (to assign the physiological parameters needed to calculate the leaf stomatal and aerodynamic resistances); and leaf area index (to calculate the bulk canopy resistances). The land cover-specific parameters in the MOD16 algorithm were obtained by comparison with eddy covariance flux data from 46 sites (located primarily in the US and Canada). MOD16 C5 used C4 MOD12Q1 Land Cover Type 2 data, while MOD16 C6 uses the MCDLCHKM product (Running et al., 2019). The meteorological data (incoming shortwave radiation and air temperature and humidity) are obtained from reanalysis products (Modern-Era Retrospective analysis for Research and Applications version 2 (MERRA-2) for C5 and Goddard Earth Observing System Model Version 5 (GEOS-5) for C6). Transpiration and soil evaporation are constrained by vapor pressure deficit, VPD (i.e., there is no soil moisture control). The C5 data were downloaded from the University of Montana's Numerical Terradynamic Simulation Group (NTSG) website (https://www.ntsg.umt.edu/project/modis/mod16.php). The C6 data were obtained from NASA's Land Processes Distributed Active Archive Center (LP DAAC) (https://lpdaac.usgs.gov/). The gap-filled version of the C6 dataset was used. The gap-filling method is the same as that used for MOD16 C5 (Running et al., 2019). Apart from the difference in resolution, the C5 and C6 datasets were produced using different reanalysis datasets and different MODIS vegetation and albedo products.

2.1.2. GLEAM ET data

In GLEAM, ET is defined as the sum of the following processes: transpiration from short and tall vegetation, bare soil evaporation, rainfall interception loss from tall vegetation, open water evaporation, and snow sublimation (Miralles et al., 2011; Martens et al., 2017). The rainfall interception loss module is based on the Gash (1979) analytical rainfall interception model (Miralles et al., 2010). GLEAM v3.3a used the MEaSUREs VCF5KYRv001 product (Hansen et al., 2018) to determine the fractions of bare soil, short vegetation, and tall vegetation. The model first calculates potential ET with the Priestley-Taylor equation using R_n and air temperature from reanalysis data (ERA-Interim). For bare soil and short vegetation, the typical value of 1.26 is used for the Priestley-Taylor α coefficient, while for tall vegetation $\alpha = 0.97$ (Martens et al., 2017). Actual ET is calculated by multiplying potential evaporation with land cover-dependent stress functions. The stress functions simulate soil water constraints on transpiration and soil evaporation. Soil water content is estimated using a multilayer running water balance model that uses a merged precipitation product, ET from the previous time step, and microwave surface soil moisture as the main inputs. The soil is divided in three layers: shallow (0–10 cm); intermediate (10–100 cm); and deep (100–250 cm). Tall vegetation can extract water from all three layers, short vegetation can extract water from the shallow and intermediate layers, and for bare soil evaporation only water from the shallow layer is available. The stress functions for vegetation also simulate the effect of phenology using microwave vegetation optical depth. The data were accessed through the GLEAM website (https://www.gleam.eu). GLEAM pixels containing more than 20% open water were excluded (this concerned a total of seven sites; Table S2). The open water fraction (OWF) was obtained from the MOD44B product (Section 2.3). This filtering was performed only for GLEAM. In the case of ALEXI, sites affected by the presence of open water were filtered out during production of the dataset (Section 2.1.3), while in the case of MOD16, no sites were affected because of the smaller pixel size.

2.1.3. ALEXI ET data

The ALEXI algorithm consists of a two-source SEB model coupled with an atmospheric boundary layer model (Anderson et al., 1997, 2007). The latent heat flux is calculated separately for the canopy and soil. An initial estimate of the canopy LE is obtained using the Priestley-Taylor equation with $\alpha = 1.26$ (assuming potential transpiration). Next, the soil LE is calculated as the residual of the energy balance. If the resulting soil LE is negative, the actual canopy LE must be less than the potential value (which may indicate an effect of soil water limitation on transpiration). The α coefficient is then reduced until the residual soil LE is non-negative. The calculated LE represents the instantaneous flux at approximately one hour before local noon. This time corresponds to the end of the time span over which H is calculated (see below). The instantaneous latent heat fluxes are extrapolated to daily ET values by multiplying by the ratio of daily total to instantaneous shortwave radiation and dividing by the latent heat of vaporization. The ALEXI algorithm calculates H from the morning rise in the radiometric surface temperature (Hain and Anderson, 2017). By using the temporal change in surface temperature, the effect of bias in the temperature retrievals on H is minimized. This ALEXI implementation uses the MODIS land surface temperature product (MYD11C1), retrieved using a generalized split-window atmospheric compensation technique (Wan, 2004). The composite values of surface temperature are partitioned between canopy and soil using estimates of vegetation cover fraction from leaf area index. The leaf area index data were obtained from the 8-day MODIS MOD15A3 product (Myneni et al., 2002). Instead of using absolute values of air temperature, ALEXI uses the slope of the vertical temperature profile (lapse rate) in the boundary layer. The lapse rate profile, as well as the surface longwave radiation flux and wind speed were obtained from the NCEP Climate Forecast System Reanalysis product (CFS-R, CFSRv2; Saha et al., 2010). Incoming shortwave radiation fluxes were obtained from the CERES SYN1deg product (Doelling, 2012). Soil heat flux is calculated as a diurnal varying function of net radiation (Santanello and Friedl, 2003). The ALEXI model uses land cover data to assign canopy parameters such as canopy height (to calculate the aerodynamic resistances to H) and leaf absorptivity (to estimate R_n for the canopy and soil). The land cover data were obtained from the MODIS MCD12C1 product (Land Cover Type 2). Since the thermal infrared based surface temperature observations are only available during clear sky conditions, ALEXI employs a gap-filling technique to generate estimates of weekly totals. The clear-sky fraction of actual ET to incoming radiation is interpolated to a daily record and then multiplied by the daily incoming radiation to generate a complete record. Along the coast the coarse-scale meteorological inputs result in limited retrievals; this is why four coastal sites (Table S2) are not included in the

ALEXI dataset.

2.1.4. Eddy covariance ET data

Data from four different flux networks (MexFlux, FLUXNET, AsiaFlux, OzFlux) were used to evaluate the ET products. The data from MexFlux were obtained directly from the site PIs (12 sites) because they were not available through a repository. The data from the other networks were obtained through the respective web-based portals. FLUXNET data available under the open data policy (tier 1) of the FLUXNET2015 dataset were used (Pastorello et al., 2020). This dataset includes a total of 28 sites between latitudes 30° S and 30 °N. From OzFlux and AsiaFlux, openly available data from sites not included in FLUXNET2015 were considered (three and nine sites, respectively). Prior to the more extensive data quality control (see below), sites for which the data record was too short (< 1 year), latent heat flux data were not available, or the degree of energy balance closure was too low (EBR < 0.5) were excluded (one site from MexFlux, four sites from FLUXNET, and seven sites from AsiaFlux). This left a total of 40 sites for further analysis (Figure 1, Table 2). Information needed for the correction of the soil heat flux (G) data or for the calculation of the sensible and latent heat storage terms, S (see below) was obtained from the metadata accompanying the datasets, from articles or other publications, or directly from the site PIs.

The remote sensing ET products were evaluated using the mean daily eddy covariance ET (mm day^{-1}) calculated for each MODIS interval. The comparisons were made using the unadjusted eddy covariance fluxes (ET_{orig}) and those corrected for the lack of energy balance closure (ET_{ebc}). FLUXNET2015 includes corrected fluxes (Pastorello et al., 2020) but the datasets from the other networks do not. For consistency, the fluxes were corrected using the same method for all datasets (including FLUXNET2015). After filling the missing half-hourly or hourly values (see below), a correction factor was calculated for each MODIS interval as $A/(H + \text{LE})$, where each term is the average daytime flux in W m^{-2} (see above for definition of terms). Daytime was defined as having solar radiation > 10 W m^{-2}. This method is based on the assumption that H and LE were underestimated by the same percentage (Twine et al., 2000). The available energy was calculated as $R_\text{n} - G - S$. The correction was only applied to the daytime data because, in absolute terms, the missing energy is small during the night (Stoy et al., 2013; Mauder et al., 2020) so that the correction will have little effect on total daily ET. In addition, this eliminated the need to ensure the completeness and consistency of the energy balance data for the nighttime period. The daytime and nighttime LE as well as the other energy balance terms (only daytime data) were converted from energy units (W m^{-2}) to millimetres (mm) using a constant value for the latent heat of vaporization (2.45 MJ kg^{-1}). The unadjusted nighttime fluxes were added to daytime ET_orig and ET_ebc to give daily ET_orig and ET_ebc.

Data on G were available for 24 of the 40 sites. At all these sites G was measured using the soil heat flux plate method (Sauer, 2002). For six sites, the measurements were not corrected for heat storage above the plates (Mayocchi and Bristow, 1995). This correction was applied retrospectively using the method of Wang and Bou-Zeid (2012). This method calculates G at the soil surface (which is required in the energy balance calculations) from the time series of G at any depth. It requires the thermal diffusivity of the soil, which was calculated as the ratio of soil thermal conductivity to soil volumetric heat capacity. The thermal conductivity was calculated following Lu et al. (2014) using site-specific soil physical data. The volumetric heat capacity was calculated from soil bulk density and soil moisture. For sites without data on G but with data on soil temperature (seven sites), G was estimated using the method of Hsieh et al. (2009). Estimates of G derived from temperature measurements at depths > 2 cm were corrected for heat storage using the method of Wang and Bou-Zeid (2012). For the remaining nine sites, G was estimated using the method of Mu et al. (2011), using in situ air temperature and R_n, and vegetation cover estimated from the MODIS FPAR product (MCD15A2H; see Section 2.1.5 for more details about this dataset).

The sensible and latent heat storage terms are generally not included in the flux datasets (Stoy et al., 2013; Pastorello et al., 2020). In this study, S was estimated from the half-hourly changes in air temperature and humidity measured at the reference level (Brutsaert, 1982). This estimate did not include heat storage in the vegetation biomass.

The data from the 40 sites were carefully screened for inconsistencies. These checks were in addition to those performed by the site PIs/teams and by some of the networks (FLUXNET, Pastorello et al., 2014, 2020; OzFlux, Isaac et al., 2017). For the daytime period, all data needed for the energy balance calculations were checked. For the nighttime period, only the LE data were screened. The quality checks were similar to those performed by Pastorello et al. (2014) for FLUXNET2015. Where possible, errors in the radiation, air temperature, and relative humidity data were corrected with the help of the site PIs, using calculated clear-sky radiation (in the case of the radiation data), or using data from another sensor or from a nearby station (Allen, 2008; Pastorello et al., 2014). No attempts were made to correct questionable eddy covariance flux data (H, LE) or soil data (G, temperature, moisture).

Gap-filling of the data was carried out in two steps. In the first step, gaps in the half-hourly or hourly data were filled on a daily basis. For the daytime period, missing values of any energy balance term, x (where x = LE, H, G, S) on a particular day were filled using $\overline{(x/R_n)}R_{n,i}$, where $\overline{(x/R_n)}$ is the average daytime ratio of x to R_n and $R_{n,i}$ the net radiation during time step i with missing data. For any x, the maximum allowed percentage of missing values was 30%. For the nighttime period, missing values of LE on a particular day were replaced by the mean nighttime LE for that day (also using an upper threshold of 30% for the percentage of missing data). For consistency, the same method was used for all datasets (i.e., the gap-filled data in FLUXNET2015 were not used). In the second step, missing daily values of ET_{orig} and ET_{ebc} were replaced by the mean of the available observations for individual MODIS intervals. The maximum allowed percentage of missing values was 25% (i.e., two days for an 8-day MODIS interval) (Hu et al., 2015).

Energy balance closure was analyzed for each site individually by summing the 8-day mean daytime totals of $H + ET_{orig}$ and A and calculating the energy balance ratio as: EBR = $\sum(H + ET_{orig})/\sum A$, with all terms in mm (Wilson et al., 2002). Energy balance closure was also analyzed by grouping the data according to land cover type and climate zone (Section 2.2). For the pooled data in each group, the 8-day mean daytime totals of $H + ET_{orig}$ were plotted against A and the corresponding linear regression line and EBR were calculated.

Geographic coordinates and land cover type data for each site were obtained from the metadata accompanying the datasets or from the literature (Table 2). The classification scheme of the International Geosphere-Biosphere Programme (IGBP) was followed. This classification system is adopted by most flux networks. It is also used in most evaluation studies of remote sensing ET models (see, e.g., Velpuri et al., 2013; McCabe et al., 2016; Michel et al., 2016).

For each site, the evaporative fraction (EF) was calculated as (Shuttleworth et al., 1989): EF = $\sum ET_{orig}/\sum(H + ET_{orig})$, where ET_{orig} and H are the mean daytime latent and sensible heat fluxes for each MODIS interval, not corrected for the lack of energy balance closure. The obtained values are listed in Table 2.

<Figure 1>

<Table 2>

2.1.5. Other datasets

The Köppen-Geiger climate class of each site was obtained using the 1-km resolution global map of Beck et al. (2018). The map was downloaded from www.gloh2o.org/koppen/. Each site was assigned the climate class of the pixel where the flux tower was located. The 40 sites represented a total of 10 different climate classes (Table 2). For the evaluation of the remote sensing ET products, these were grouped into four main climate zones (Section 2.2). For each of these climate zones, the average EF was calculated using the site-specific values listed in Table 2.

To investigate the match between the actual vegetation type at the flux tower site and the vegetation class or category used in the remote sensing ET models (Section 2.3), the yearly MODIS land cover (MCD12Q1; 500 m resolution) and vegetation cover (MOD44B; 250 m resolution) products were used. The data were downloaded from the NASA LP DAAC website. From MCD12Q1, the Land Cover Type 2 data were used. From MOD44B, the data layers containing percent tree cover and percent nontree vegetation were used. For each site, the following three subsets were generated for the years with eddy covariance data: Subset 1) four pixels of MCD12Q1 data corresponding to the 1-km MOD16 pixel; Subset 2) all pixels of MOD44B data falling within the 0.25° GLEAM pixel; and Subset 3) all pixels of MCD12Q1 data corresponding to the 5-km ALEXI pixel. These subsets were used in the analysis described in Section 2.3.

Finally, FPAR data from the MCD15A2H product were used to calculate *G* with the method of Mu et al. (2011) (Section 2.1.4). This product is an 8-day composite dataset with a spatial resolution of 500 m. The data were again obtained from NASA's LP DAAC. The pixels matching the location of the flux towers were used. Data with a cloud flag or retrieved by the backup algorithm were replaced by interpolated values (Zhao et al., 2005).

2.2. Evaluation of product performance

The remote sensing ET products were evaluated by grouping the data by IGBP land cover type and Köppen-Geiger climate zone (Velpuri et al., 2013; McCabe et al., 2016). To avoid groups with only one site, the woody savanna site was included in the group with the savanna sites and the closed shrubland site was left out of the evaluations by land cover type (but included in the evaluations by climate zone). This resulted in the following five groups of vegetation cover types: evergreen broadleaf forest (EBF); deciduous broadleaf forest (DBF); evergreen needleleaf forest (ENF); savanna (SAV); and grassland (GRA).

Likewise, the sites were grouped into the following four main climate zones: i) Af, Am: tropical fully humid and tropical monsoon, respectively (from now on referred to as tropical wet); ii) Aw: tropical savanna; iii) B: dry; and iv) C: mild temperate. Sites assigned the mild temperate (C) climate were either located on tropical or subtropical mountains (five sites) or in lowland areas in the subtropics (three sites) (see also Richter, 2016). Table 3 shows the number of sites and the number of site years available in the complete dataset for each product by land cover type and climate zone.

<Table 3>

2.2.1. Scatter plots and regression analysis

Scatter plots allow visual evaluation of the match between the remote sensing-based and the observed ET data (Velpuri et al., 2013; McCabe et al., 2016; see also Chang and Hanna, 2004). In addition, the slope, intercept, and coefficient of determination (R^2) of the fitted linear regression line provide a quantitative way to evaluate product performance (Willmott, 1982;

Velpuri et al., 2013; McCabe et al., 2016). Most studies evaluating remote sensing ET products perform the regression analysis with the product estimates on the *y*-axis and the observations on the *x*-axis (see, e.g., Mu et al., 2011; Velpuri et al., 2013; McCabe et al., 2016). However, Piñeiro et al. (2008) showed that this can lead to erroneous estimates of the regression coefficients. Therefore, in this study the observations were used as the *y* variable and the product estimates as the *x* variable (Piñeiro et al., 2008). For each land cover type and climate zone in Table 3, the eddy covariance observations were plotted against the ET estimates of each product and the corresponding linear regression lines were calculated, using the pooled data from the different sites in each group. This analysis was performed using both ET_{orig} and ET_{ebc}.

2.2.2. Statistical performance metrics

In addition to visual inspection of the scatter plots and examination of the regression results, three commonly used statistics in evaluation studies of remote sensing ET products were calculated: root mean square error (RMSE), percent bias (PBIAS), and the coefficient of determination (R^2) (see references in Table S4). The use of these common statistics allowed for comparison with evaluation results from other latitudes (Section 2.4). The selected metrics provide complementary information about product performance. The RMSE is a measure of total error (i.e., both random and systematic errors) and is defined by:

$$\mathrm{RMSE} = \sqrt{N^{-1} \sum [\mathrm{ET(Prod)} - \mathrm{ET(Obs)}]^2} \quad (1)$$

where ET(Prod) is the product ET, ET(Obs) the eddy covariance ET, and *N* the total number of data points (i.e., the number of MODIS intervals).

The PBIAS is the systematic (bias) error in percent of the average of the observations:

$$\mathrm{PBIAS} = \frac{N^{-1}\sum[\mathrm{ET(Prod)} - \mathrm{ET(Obs)}]}{N^{-1}\sum \mathrm{ET(Obs)}} \times 100 \quad (2)$$

Third, in addition to the R^2 calculated from the pooled data (Section 2.2.1), the R^2 of the linear regression between product ET and observed ET was calculated for each site separately. Besides being a measure of correlation, R^2 indicates how much of the variation in observed ET is explained by the product ET.

The three metrics were calculated using both ET_{orig} and ET_{ebc}. Averages of both sets of RMSE, PBIAS, and R^2 values were calculated for each land cover type and climate zone in Table 3. The average metrics by land cover type and climate zone were displayed graphically in plots for each product (McCabe et al., 2016) for ET_{orig} and ET_{ebc}. The results for the individual sites can be found in Tables S2 and S3.

2.2.3. Combining the different performance metrics into a single score

To facilitate comparison of the overall performance of the different ET products, the individual metrics (R^2, RMSE, PBIAS) were combined into the Ideal Point Error (IPE) score (Elshorbagy et al., 2010; Dawson et al., 2012). The IPE score takes values between 0 and 1, with 0 indicating perfect performance (i.e., all metrics are at their optimum values) and 1 being assigned to the worst performing product. In practice, no product (and no observation) is without error. Therefore, the best performing product will usually have an IPE greater than 0. The IPE values were calculated for each of the comparisons by land cover type and climate zone. The calculation of IPE consists of two steps. In the first step, each performance metric is standardized to the worst score for that metric. Dawson et al. (2012) provides expressions for this standardization step for different categories of performance measures (denoted by S1–S5; their Table 1). PBIAS is not listed in this table. However, as mentioned by Dawson et al. (2012), the flexibility of this method allows other metrics to be included. PBIAS classifies as an S4 category metric and was standardized using the corresponding expression. In the second step, the IPE is calculated from the standardized metrics using Equation (2) in Dawson et al. (2012). The results were plotted as heatmaps for ET_{orig} and ET_{ebc}.

2.2.4. Sensitivity to the choice of reference dataset

The statistical metrics (R^2, RMSE, PBIAS) and the IPE scores were calculated as explained above but now using the common reference dataset. This direct comparison approach included 12 fewer sites and about 36% fewer MODIS intervals than when using all data (see Table S1 for the number of sites and site years by land cover type and climate zone). The sensitivity analysis was performed for ET_{orig} only. Differences in the ranking of products for each of the comparisons by land cover type and climate zone were determined by comparing the IPE scores from both approaches. Changes in ranking were indicated by adding an asterisk to the IPE scores in the heatmap for ET_{orig}.

2.2.5. Evaluation of seasonal trends in ET from products

The ability of the products to capture seasonal changes in ET was examined by plotting the average monthly ET for each product together with the average monthly ET_{orig} and ET_{ebc}. This was again done for each land cover type and climate zone in Table 3. To account for the different timing of the rainy seasons, separate plots were made for sites located in the northern and southern hemispheres.

2.3. Vegetation match index (VMI) and open water fraction (OWF)

The effect of a mismatch between the vegetation at the flux tower site and that detected in the model pixel on the product evaluation results was examined by calculating a vegetation match index (VMI). The models underlying the investigated ET products differ in the level of detail with which they distinguish between different vegetation types. Both MOD16 and ALEXI assign land cover-specific parameters to a wide range of cover types, while GLEAM only considers two vegetation categories (i.e., tall and short vegetation). However, also for MOD16 and ALEXI the largest differences between the land cover-specific parameters occur between tall and short (or forest and non-forest) vegetation types (Anderson et al., 2007; Mu et al., 2011). Therefore, for all three products, VMI was calculated based on these two vegetation categories.

The datasets used to calculate the VMIs are described in Section 2.1.5. As explained in Section 1, a mismatch in vegetation can be caused by scale differences or inaccuracies in the vegetation input data. To account for the latter, vegetation data were selected that were as similar as possible to those used to generate the products (Sections 2.1.1–2.1.3). For MOD16 and ALEXI, MCD12Q1 Land Cover Type 2 data were used (Subsets 1 and 3, respectively). The data from Subsets 1 and 3 were aggregated into forest and non-forest categories. For sites with a forest land cover (EBF, DBF, ENF; Table 2), VMI_{MOD16} or VMI_{ALEXI} was calculated as the proportion of forest vegetation present in the 1-km MOD16 or 5-km ALEXI pixel. For sites with a non-forest land cover (SAV, GRA), the VMIs were calculated as the proportion of non-forest vegetation. For GLEAM, MOD44B vegetation cover data were used (Subset 2). These data were assumed to be similar to those of the VCF5KYR product (used as input to GLEAM v3.3a; Section 2.1.2). The VCF5KYR product is based on AVHRR observations calibrated with MODIS data (Hansen et al., 2018). In each data layer of the MOD44B product, pixels with water are masked out with a fill value of 200. Hence, VMI_{GLEAM} was calculated as either the average percent tree cover (for sites with forest vegetation) or the average percent nontree vegetation (for sites with non-forest vegetation) multiplied by the fraction of land pixels. In addition, the open water fraction (OWF) was calculated. This index was used to filter out sites for which the pixel contained more than 20% water (Section 2.1.2).

The dependence of product performance on the vegetation-match between pixel and site was examined by plotting the performance metrics (R^2, RMSE, PBIAS) against VMI. Individual site values for the metrics were bin-averaged into four evenly spaced intervals of 0.25 VMI units wide in the case of GLEAM and ALEXI or for each of the five discrete VMI values in the case of MOD16. For each metric-VMI combination, the linear regression line was calculated. In addition to visual inspection of the scatter plots, the *p*-values of the calculated regression slopes were used to evaluate whether there was a relationship between VMI and product performance. For this analysis, performance statistics obtained for ET_{orig} were used.

2.4. Latitudinal comparison of product performance

To investigate latitudinal dependence of the performance of the ET products examined here, a literature search was conducted to find studies that evaluated these products. To allow for direct comparison, only studies that evaluated the products with eddy covariance-based ET were considered. Furthermore, a study needed to report at least one of the three performance metrics used in this study (R^2, RMSE, PBIAS) or provide the data from which these metrics could be calculated. The performance statistics depend on the averaging time used. Hence, ideally, only studies using the same time average as used here (8-day) should be considered. This would, however, drastically reduce the number of evaluation results available. Therefore, studies using daily or monthly time averages were also included in the initial search.

In the end, a total of 15 studies were found (Table S4). As will be shown below, the evaluation results were different for ET_{orig} and ET_{ebc}. Of the 15 studies found in the literature, 13 used ET_{orig}

and only two used ET_{ebc}. No studies were found that used both. For the final analysis, only studies using ET_{orig} were considered. The study of Miralles et al. (2011) was also excluded because: i) GLEAM ET was calculated using in situ measured R_n; ii) comparisons were made using modelled ET for the vegetation type (i.e., tall or short vegetation) matching that at the tower site; and iii) days with rainfall were excluded. Likewise, the study of Mu et al. (2011) was excluded because their evaluation results are in fact calibration results. This yielded a total of 12 studies, including the current one. Most studies used MOD16 C5 because MOD16 C6 was only recently released. Therefore, the results obtained here for C5 were used. To account for latitudinal differences in ET, RMSE was normalized by mean ET_{orig} (NRMSE). Not all studies reported ET_{orig} (Table S4). The studies evaluated product performance at a minimum of two eddy covariance sites to a maximum of 119 sites. Most studies reported evaluation results for individual sites but some reported averages for land cover classes (e.g., Velpuri et al., 2013; Reitz et al., 2017). The latter were treated as if they were results for individual sites. Performance results were grouped into results for forest and non-forest vegetation; there were not enough performance data available to create more specific subgroups. The results were further grouped into three latitudinal bands: southern low latitudes (30° S–0°); northern low latitudes (0°–30° N); and northern middle latitudes (30° N–60° N). For latitudes outside these regions, there were not enough data available (Table S4). Table 4 summarizes the number of evaluation results (N_{ER}) available, broken down into results for R^2, NRMSE, and PBIAS, and grouped by latitudinal zone, product, and vegetation category. Averages of each performance metric for each product-vegetation category combination were plotted as a function of latitude.

<Table 4>

3. Results

3.1. Energy balance closure of eddy covariance data

Table 2 shows the daytime energy balance ratio (EBR) for each of the 40 individual sites. The average daytime EBR for the 40 sites was 0.83, with a standard deviation (SD) of 0.10, and with values ranging from 0.63 to 1.03. Figures 2 and 3 show scatter plots between the sum of the daytime turbulent heat fluxes and available energy grouped by land cover type and climate zone, respectively. The daytime EBR values calculated from the pooled data were similar across land cover types (ranging from 0.79 to 0.87) and climate zones (ranging from 0.77 to 0.85). The slopes and intercepts of the regression lines ranged from 0.67 to 0.80 and −0.01 to 0.67 mm day^{-1}, respectively, across land cover types and from 0.72 to 0.81 and 0.00 to 0.69 mm day^{-1}, respectively, across climate zones. The coefficient of determination (R^2) ranged from 0.64 and 0.82 across land cover types and from 0.59 and 0.82 across climate zones.

<Figure 2>

<Figure 3>

3.2. Evaluation of ET products by land cover type

Figures 4 and S1 show scatter plots comparing eddy covariance-based and remote sensing-based ET by land cover type for each of the evaluated products for ET_{orig} and ET_{ebc}, respectively. First, the results for ET_{orig} will be examined. Although the scatter plots and the regression results for MOD16 C5 and MOD16 C6 show some differences, these were generally smaller than the differences with the other products (see also below). Hence, from now on the two collections will be referred to as MOD16. When necessary, a distinction will be made between the two.

Overall, GLEAM ET showed the best agreement with ET_{orig}. This follows from the results of the regression analysis (i.e., slope closer to 1, intercept closer to 0, higher R^2) and can be observed visually as a narrower distribution of data points around the 1:1 line. For DBF and SAV, the correlations between GLEAM ET and ET_{orig} were strong (R^2 of 0.81 and 0.73, respectively). A weak correlation was observed for EBF ($R^2 = 0.32$). The agreement with ET_{orig} was generally poorer for MOD16 and ALEXI. Neither of these products consistently outperformed the other. The scatter plots show a clear overestimation of ET_{orig} by MOD16 for ENF and a clear underestimation for SAV. Although both products showed weaker correlations with ET_{orig} than GLEAM, this was most pronounced for ALEXI. Also MOD16 and ALEXI had the strongest correlations for DBF and SAV and the weakest for EBF. When evaluating the products with ET_{ebc}, the regression slopes and intercepts increased, while changes in R^2 were generally small (Figure S1).

<Figure 4>

Figure 5 shows the mean R^2, RMSE, and PBIAS by land cover type for each of the evaluated products for ET_{orig} and ET_{ebc}. Again, the results for ET_{orig} will be examined first. As already observed in Figure 4, the mean performance statistics show that the differences between MOD16 C5 and MOD16 C6 are generally smaller than the differences with the other products. One exception is ENF; this group, however, included a relatively small number of sites and site years (Table 3), which may have affected the comparisons. Figure 5 confirms the superior performance of GLEAM. Overall, GLEAM had the strongest correlations, the lowest RMSEs, and the smallest PBIAS values. In agreement with the graphical analysis, neither MOD16 nor ALEXI was second best over all land covers. Both GLEAM and MOD16 tended to overestimate ET_{orig}

for forest vegetation and underestimate ET_{orig} for non-forest vegetation; however, biases were smaller for GLEAM. ALEXI tended to overestimate ET_{orig} for all land cover types. The variation in PBIAS across land cover types was smaller for GLEAM and ALEXI than for MOD16. As seen in the scatter plots, ALEXI had the weakest correlations with ET_{orig}. All ET products had the strongest correlations for DBF and SAV and the weakest for EBF.

<Figure 5>

As expected, PBIAS shifted to more negative values when the products were evaluated with ET_{ebc} (Figure 5). Depending on whether PBIAS decreased or increased, the corresponding RMSE became smaller or larger (although not for all products; see below). The use of ET_{ebc} generally had little effect on the correlations (as also seen in the scatter plots). For GLEAM, PBIAS values were negative for all land cover types when using ET_{ebc} and were generally greater in absolute terms than when using ET_{orig}. As a result, the RMSEs were larger (and closer to those of the other products) when using ET_{ebc} than when using ET_{orig}. For MOD16, PBIAS values were also negative for most land cover types when using ET_{ebc}. The corresponding RMSEs were either somewhat larger (e.g., SAV, GRA) or smaller (e.g., EBF, ENF) than when using ET_{orig}. In the case of ALEXI, PBIAS values decreased for all land cover types except SAV. However, only in the case of EBF this was accompanied by a decrease in RMSE. For ENF, DBF and GRA, the RMSE actually increased. A partial explanation for this is the tendency of ALEXI to overestimate low ET_{ebc} and underestimate high ET_{ebc} (DBF, GRA; Figure S1).

Figure 6 shows the IPE scores for the different ET products by land cover type as obtained using ET_{orig} or ET_{ebc} for evaluation. The IPE scores confirm that GLEAM ET best matched ET_{orig} across all land cover types. The IPE scores also support the earlier observations that: i) the differences in performance between the two MOD16 collections were generally smaller than the differences with the other products; and ii) neither MOD16 nor ALEXI consistently outperformed the other. Figure 6 shows that the IPE values of the products converged when using ET_{ebc} for evaluation. This largely reflects the changes in PBIAS and RMSE mentioned above. When using ET_{ebc} there is no product that stands out in terms of both small PBIAS and high R^2 across all land cover types.

<Figure 6>

Figure 7 compares the seasonal trends in ET from the products with those from the observations by land cover type by hemisphere. Note that these curves were calculated using the common reference dataset. Only curves calculated with data from at least two sites are shown. Clear differences in the seasonality and timing of rainfall can be observed. In both hemispheres, ET of EBF was characterized by weak seasonality, with constant high values throughout the year. Yet, MOD16 C5 seemed to capture the small variations in ET quite well. This was also the case for GLEAM, except during the wet season in the southern hemisphere when it showed a strong positive bias. A closer look at the data showed that this involved the two Brazilian rainforests (Table 2; Figure 1). Similarly, ALEXI had a strong positive bias at the end of the dry season in the southern hemisphere. This could be traced mainly to EBF in northeastern Australia (Table 2; Figure 1). For ENF, all products seemed to represent the observed seasonal trend in ET fairly well. For DBF, GLEAM closely followed the observed seasonal trend in ET. MOD16 had a negative bias during the dry season. Conversely, ALEXI had a positive bias during the transition from the wet to dry season. For SAV and GRA, both GLEAM and MOD16 had a strong negative bias during the dry season. Conversely, ALEXI seemed to have a positive bias during the dry period in these cover types.

<Figure 7>

3.3. Evaluation of ET products by climate zone

The performance of the ET products was also examined across four main climate zones (Table 3). For each climate zone, an average evaporative fraction (EF) was calculated from the site-specific values in Table 2, yielding (ranked from wet to dry): 0.73 ± 0.04 (SD) for Af, Am (tropical wet); 0.60 ± 0.10 for C (mild temperate); 0.50 ± 0.11 for Aw (tropical savanna); and 0.35 ± 0.11 for B (dry). The tropical wet climate zone included mainly EBF sites (seven in total; Table 3). The mild temperate climate zone included all ENF sites and for the rest mainly EBF sites. The savanna and dry climate zones included mostly SAV, DBF, and GRA sites. The results of this analysis were presented in the same way as in the previous section, i.e., scatter plots (Figures 8 and S2), average performance statistics (R^2, RMSE, PBIAS; Figure 9), heatmaps of IPE scores (Figure 6), and average seasonal trends in ET (Figure 10).

The comparisons by climate zone confirmed many of the findings in the previous section. Again, the differences in performance between the two MOD16 ET collections were usually smaller than the differences with the other products (Figures 8 and 9). Furthermore, the performance and ranking of the products depended on whether ET_{orig} or ET_{ebc} was used for evaluation. When using ET_{orig}, GLEAM again showed the strongest correlations and best agreement (i.e., closeness to observations) (Figures 8 and 9). As a result, GLEAM had the smallest RMSEs and best IPE scores across all climate zones (Figures 9 and 6, respectively). Again, neither MOD16 nor ALEXI performed consistently better than the other. That is, MOD16 showed better agreement with ET_{orig} for the Aw climate zone, whereas ALEXI gave better results for the B and C climate zones (as summarized by the IPE scores in Figure 6). Finally, there was no clear ranking among the products when ET_{ebc} was used for evaluation (Figure 6). This mainly reflected the underestimation of GLEAM ET with respect to ET_{ebc}, leading to higher (i.e., more negative) PBIAS values and larger RMSEs than when using ET_{orig} (Figure 9). For MOD16, PBIAS and RMSE values both decreased (e.g., C climate zone) and increased (Aw climate zone). For ALEXI, PBIAS decreased to values close to zero (C, Aw, B); however, instead of decreasing, the corresponding RMSEs increased. The use of ET_{ebc} generally had little effect on the correlations (Figures 8, S2, and 9).

<Figure 8>

<Figure 9>

All products had the weakest correlations in the wet tropical climate zone and the strongest in the tropical savanna and dry climate zones (Figure 9). This is consistent with the results in Section 3.2 (weakest correlations for EBF and strongest correlations for SAV and DBF). Overall, ALEXI had again the weakest correlations. The biases of MOD16 ET varied markedly across climate zones (Figure 9). When compared with ET_{orig}, MOD16 tended to overestimate ET in the wet tropical and mild temperate climate zones and underestimate ET in the dry climate zone. This result is consistent with the positive biases observed in Figure 5 for forest vegetation (dominating the wet tropical and mild temperate climate zones) and the negative biases for non-forest vegetation (dominating the dry climate zone). Biases in GLEAM showed the same tendency but were generally much smaller in size. For ALEXI, the bias with respect to ET_{orig} was practically zero in the wet tropical climate zone but positive in the other climate zones.

The seasonal trend analysis (Figure 10) revealed the same patterns as found earlier in Section 3.2, again reflecting the close correspondence between climate and vegetation. For the tropical wet climate zone, MOD16 C5 ET closely followed the observed seasonal changes in ET. This was also the case for GLEAM, except for the positive bias during the wet season in the southern hemisphere (traced back mainly to the Brazilian rainforests; Section 3.2). The positive bias of ALEXI ET at the end of the dry season in the southern hemisphere can also be observed again (traced back mainly to EBF in northeastern Australia; Section 3.2). In addition, ALEXI ET showed large, seemingly erratic, variations in the northern hemisphere. For the mild temperate climate zone, all products represented the observed seasonal trend in ET fairly well. For the tropical savanna climate, both MOD16 and GLEAM had a strong negative bias during the dry season, which was also observed in the plots for GRA and SAV in Figure 7. The positive bias for ALEXI during the dry period can also be observed again. For the dry climate zone, GLEAM ET closely followed the observed seasonal trend in ET. ALEXI had again a positive bias during the dry period. MOD16 had a strong negative bias during the wet season in the southern hemisphere.

<Figure 10>

3.4. Sensitivity to the choice of reference dataset

The IPE scores based on the common reference dataset (Figure S3) show similar results to those obtained using all data (Figure 6, ET_{orig}). For both the comparisons by land cover type and climate zone, GLEAM ET generally showed the best agreement with ET_{orig}. Furthermore, the differences among the two MOD16 collections and ALEXI were generally too small to identify a second best performing product. Differences in ranking results between the two approaches were observed only for MOD16 and ALEXI (Figure 6).

3.5. Product performance versus VMI

Figure 11 shows binned scatter plots between the performance metrics (R^2, RMSE, PBIAS) and the vegetation match index for the different products. In addition, the regression lines and the *p*-values indicating the statistical significance of the regression slopes are shown. The VMIs for each individual site are given in Table S2. The average VMI was 0.77 ± 0.41 for MOD16, 0.71 ± 0.40 for ALEXI, and 0.51 ± 0.23 for GLEAM. These results indicate a decreasing vegetation-match between pixel and site with increasing pixel size, although the VMIs of GLEAM and the other products cannot be directly compared as they are based on different data. For none of the products there was an improvement in performance (i.e., increasing R^2 or decreasing RMSE or PBIAS) with increasing VMI (Figure 11). Moreover, for none of the regressions the slope was statistically significant.

<Figure 11>

3.6. Latitudinal comparison of product performance

Figure 12 shows zonal averages (southern and northern low latitudes and northern middle latitudes) of the performance metrics grouped by forest and non-forest vegetation for MOD16 and GLEAM and ALEXI. The averages were calculated using evaluation results from this study and from the literature (Tables S2 and S4). For ALEXI, no data on NRMSE and PBIAS were available for the northern middle latitudes (Table 4). Figure 12 should be interpreted with caution because the number of evaluation results (N_{ER}) available varied considerably among latitudinal zones, products, and vegetation categories (Table 4). In the case of forest vegetation, correlations (all products) seem to be weaker while PBIAS and NRMSE scores (MOD16 and

GLEAM) seem to be better at low latitudes than at northern middle latitudes. Both MOD16 and GLEAM seem to overestimate ET_{orig} in all latitudinal zones. In contrast, in the case of non-forest vegetation the performance metrics show no clear variation with latitude. Moreover, both MOD16 and GLEAM seem to underestimate ET_{orig} in all latitudinal zones.

<Figure 12>

4. Discussion

4.1. The effect of the energy balance closure problem on product evaluation results

The average energy balance ratio for the 40 sites in this study (0.83) is nearly identical to that reported by Stoy et al. (2013) for 173, mainly mid-latitude, FLUXNET sites (0.84). When grouped by land cover type or climate zone, the ranges of EBR values were fairly small (0.79–0.87 or 0.77–0.85, respectively). A greater range was observed for the 173 FLUXNET sites grouped by land cover type (0.70–0.94; Table 2 in Stoy et al., 2013).

As mentioned in Section 1, the reasons for the energy balance closure problem and the extent to which it affects the ET fluxes are not yet clear. Despite being long recognized, the effect of the energy balance closure problem on the evaluation results of remote sensing-based ET products has rarely been examined (Michel et al., 2016; Melton et al., 2021). This study found that the performance and ranking of the evaluated products depended on whether the unadjusted or the energy balance closure corrected ET fluxes were used. When using ET_{orig}, GLEAM showed the best overall performance with the strongest correlations and smallest biases. However, when using ET_{ebc}, none of the products was superior to the others. Not surprisingly, the use of ET_{ebc} instead of ET_{orig} affected the product biases more than the correlations. Overall, MOD16 and GLEAM underestimated ET compared to ET_{ebc}. For ALEXI, PBIAS decreased when using ET_{ebc}

instead of ET_{orig}, but the corresponding RMSEs tended to increase rather than decrease. For most SEB approaches used in the OpenET project, cumulative totals of ET over the growing season or water year also agreed better with ET_{ebc} than with ET_{orig} (Melton et al., 2021).

Both MOD16 and GLEAM include parameters that were calibrated using field observations of ET. MOD16 was calibrated using ET obtained from eddy covariance-based estimates of water use efficiency (WUE) and MODIS-based gross primary production (GPP), with WUE being calculated as the ratio between GPP and ET fluxes not corrected for energy balance closure (Mu et al., 2011). The use of ET_{orig} could possibly explain the negative bias of MOD16 with respect to ET_{ebc} (Michel et al., 2016). However, the GPP fluxes may have been underestimated for the same reason as ET_{orig} (Foken et al., 2011). In that case, the estimated WUE would not (or only partly) be affected. In GLEAM, fixed values are used for the Priestley-Taylor coefficient for short (α = 1.26) and tall (α = 0.97) vegetation (Martens et al., 2017). These values are averages of α values published in the literature, which in turn were obtained by comparing field measurements of ET under well-watered conditions with potential ET. Some of the α values were derived with ET_{orig}, but others were obtained using ET estimates based on other methods, such as the weighing lysimeter and bowen ratio energy balance techniques (see references cited in Martens et al., 2017). Hence, also the negative bias error of GLEAM with respect to ET_{ebc} cannot be directly linked to calibration with ET_{orig}. ALEXI ET had smaller PBIAS when using ET_{ebc} than when using ET_{orig}. In contrast to the other models, ALEXI is not calibrated with field data. However, no conclusions can be drawn from this observation without a better understanding of the effect of the energy balance closure problem on ET_{orig}. In addition, the RMSEs of ALEXI tended to increase rather than decrease when using ET_{ebc}.

The literature review showed that most studies evaluated the products using ET_{orig} (Table S4). It is recommended to use both ET_{orig} and ET_{ebc} as long as the effect of the energy balance closure problem on ET_{orig} is not clear. The remainder of the discussion will focus on the results obtained with ET_{orig} to facilitate comparisons with the literature.

4.2. Relative performance of the evaluated products

Similar results were obtained when grouping the data by land cover type or climate zone, showing the close relationship between the two (see, for example, Cui et al., 2021). Therefore, the results of these two analyses will be discussed together and interchangeably. As explained in Section 1, a distinction should be made between product and model evaluation studies. The former evaluate the published ET products while the latter evaluate the performance of the underlying models using a common input dataset. Because modeled ET is sensitive to the input data, the results of the model evaluation studies do not necessarily apply to the final ET products (see Section 1 for references).

Overall, GLEAM had the best performance across different land cover types and climate zones in the low latitudes; neither MOD16 nor ALEXI could be identified as the second best performing product. These results were obtained regardless of whether the comparisons were made using all data or a common reference dataset. There are very few product evaluation studies that have compared the performance of the products assessed in this study. Khan et al. (2018, 2020) compared the performance of GLEAM and MOD16 at nine and five eddy covariance sites, respectively; in both studies, about one third of sites were located in the low latitudes and about two thirds in the middle latitudes (see also Table S4). Khan et al. (2018) used ET_{orig} to evaluate the products, while Khan et al. (2020) used ET_{ebc}. Consistent with the results of this study, Khan et al. (2018) found that GLEAM generally corresponded better with ET_{orig} than MOD16. Also when using ET_{ebc}, Khan et al. (2020) found that GLEAM outperformed MOD16 (as opposed to this study where differences were small when using ET_{ebc}). In the absence of other comparative studies, the compilation of product evaluation results for the latitudinal analysis (Figure 12, Table S4) allows for an indirect comparison of product performance. Focusing only on the results for the northern middle latitudes (to exclude the evaluation data from this study), the overall better match of GLEAM with ET_{orig} as compared to MOD16 is also evident from the studies included in this analysis (see Table S4 for references). The stronger correlations of GLEAM as compared to ALEXI are also noticeable when comparing the R^2 values from studies that evaluated these products separately.

An overall better performance of GLEAM as compared to MOD16 was also observed in the model evaluation studies by McCabe et al. (2016) and Michel et al. (2016). GLEAM also performed better than the surface energy balance approach evaluated in these studies (i.e., the SEBS model of Su, 2002). Similarly, Vinukollu et al. (2011a) obtained better performance results for a Priestley-Taylor approach developed by NASA's Jet Propulsion Laboratory (PT-JPL; Fisher et al., 2008) than for MOD16 and SEBS. However, all these studies concluded that no single model was superior in all cases. The same conclusion was reached in a recent model evaluation study for South America (Melo et al., 2021). Such a conclusion is not supported by the results of this study as GLEAM performed better than MOD16 and ALEXI in all land cover types and climate zones. It is not known whether this is a feature of the low latitudes (i.e., the other studies focused mainly on the middle latitudes), whether it is related to the differences between product and model evaluation studies discussed earlier, or whether it is a result specific to the products studied here.

All products had the weakest correlations in the wet tropical climate zone (dominated by EBF) and the strongest in the tropical savanna and dry climate zones (dominated by DBF and SAV). For the most part this reflects differences in seasonality (i.e., the greater the variation in ET, the stronger the correlations; Miralles et al., 2011; Yilmaz et al., 2014) rather than differences related to the performance of the products. ALEXI generally had the weakest correlations of all products. Comparing the R^2 values from the studies used for the latitudinal analysis shows values for ALEXI between those of MOD16 and GLEAM (northern middle latitudes, Figure 12). Although this is an indirect comparison (because it involves studies that evaluated the products separately), it suggests that the low correlations observed in this study for ALEXI are not a general feature of the product. A known challenge for thermal-based approaches is the filtering of cloud-contaminated data and the resulting gaps between successful retrievals. Failure to detect cloud-contaminated data can lead to large errors in ALEXI ET estimates (Anderson et al., 2007; Yilmaz et al., 2014). The uncertainty in gap-filled ALEXI ET can be twice as large as that in ET generated by the algorithm under clear-sky conditions (Anderson et al., 2007). These cloud-related problems could be responsible for the weak correlations of ALEXI, but that still does not explain the difference with the northern middle latitudes where the data are also affected by clouds. A possible approach to solving these problems is to use cloud-tolerant microwave-based land surface temperature in ALEXI (Holmes et al., 2018). Finally, Holmes et al. (2018) found that averaging 0.05° ALEXI ET estimates to 0.25° spatial resolution generally improved correlations with flux tower data. They attributed this to the presence of noise in the MODIS land surface temperature data that outweighed the benefits of the higher resolution compared to a 0.25° average. However, the overall effect reported in that study is too small to explain the differences in correlation with the other products observed here. More work is needed to understand the reasons for the low correlations of ALEXI observed at the low-latitude sites studied here.

Both MOD16 and GLEAM had a positive bias for forest vegetation and a negative bias for non-forest vegetation. No such land cover type-dependent biases were observed for ALEXI. The biases of MOD16 and GLEAM are also evident from the evaluation results of other studies (northern middle latitudes, Figure 12). Both Kun Zhang et al. (2019) and Brust et al. (2021) showed that the biases in MOD16 can be significantly reduced when calibrating the algorithm with more and a greater diversity of sites than used in the original calibration. Brust et al. (2021) found that the accuracy of MOD16 can also be improved by including the effect of soil moisture on ET. Although the apparent vegetation type-dependent biases were less pronounced in GLEAM, more work is needed to understand the causes. In a model evaluation study, Miralles et al. (2016) found the opposite pattern, i.e., MOD16 and GLEAM underestimated ET (determined from rainfall and streamflow data) in wet regions (dominated by forest vegetation) and overestimated ET in dry regions (dominated by non-forest vegetation types). A similar pattern was observed in the model evaluation study by Michel et al. (2016) (using ET_{orig} as reference data). In the model evaluation study for South America, GLEAM underestimated ET in both wet and dry regions (Melo et al., 2021). The biases of MOD16 were small in that study. This shows again that the results of model evaluation studies do not necessarily apply to the actual ET products.

None of the products were able to correctly represent the seasonal trend in ET in all land cover types and climate zones. Detailed analyses such as this one can help identify the causes of the biases discussed above. For example, the negative biases of MOD16 and GLEAM in GRA and SAV seemed to occur mainly during the dry season. This may indicate an overestimation of the effect of water stress on ET. In contrast, ALEXI seemed to overestimate ET during the dry season. In some cases, the differences between the product-based and the observed trends could be traced to individual sites. The overestimation of ET of Brazilian rainforest by GLEAM during the wet season was also observed by Chen et al. (2022). These authors suggested the lack of an atmospheric moisture control on transpiration as a possible cause of this overestimation. It is likely that the erratic variation observed in the ALEXI data for the tropical wet sites was caused

by the cloud-related problems discussed above. However, to correctly identify possible seasonal biases, a more comprehensive analysis that includes more sites and site years is needed.
The results showed that the differences between the MOD16 C5 and C6 products were generally smaller than the differences with the other products. Differences between C5 and C6 were to be expected because of differences in input data and spatial resolutions (Mu et al., 2013; Running et al., 2019). The differences persisted when using the common reference dataset (Figure S3). Future work can focus on quantifying the level of consistency between these two collections.

4.3. Latitudinal comparison of product performance

The literature review revealed that remote sensing-based ET products have been evaluated primarily in the northern middle latitudes. The bias is the result of geographic differences in the availability of eddy covariance data due to uneven distribution of flux towers (see, for example, Schimel et al., 2015) and regional differences in data sharing (Villareal and Vargas, 2021). With the results of this study, the availability of evaluation data for the low latitudes was significantly improved. This allowed a comparison of product performance across latitudes. The results of this analysis should be interpreted with caution, however, because the number of evaluation results available varied considerably among latitudes, products, and vegetation categories (Table 4). Smaller normalized RMSEs and smaller PBIAS values suggest better performance of MOD16 and GLEAM for low-latitude forests than for northern mid-latitude forests. The weaker correlations at low latitudes are thought to be the result of differences in seasonality rather than differences in performance (see below). The similarity between the latitudinal trends in NRMSE and PBIAS of MOD16 and GLEAM is striking considering the different approaches, forcing data, and resolutions of the underlying models. More work is needed to understand the causes of the apparent latitudinal dependence of these products. In the case of non-forest vegetation, none of the performance metrics showed a clear trend with latitude. Noteworthy is that both MOD16 and GLEAM seem to overestimate ET of forest vegetation and underestimate ET of non-forest vegetation in all latitudinal bands (see also discussion above).

A limitation of the current analysis is that regional differences were not detected because of the broad zonal bands used. For example, NRMSEs were considerably larger for seasonally dry DBF (0.36 and 0.84 for GLEAM and MOD16, respectively; data not shown) than for wet tropical EBF (0.27 and 0.46, respectively). Similarly, Vinukollu et al. (2011b) and Miralles et al. (2016) found higher relative uncertainties for the subtropics than for the tropics. In these studies, relative uncertainty was estimated from the spread between different model outputs.

The weaker correlations for low-latitude forests are most likely explained by the small seasonal variation in ET of EBF. The seasonal variation in ET of temperate forests is much greater due to stronger seasonal variations in radiation and temperature (Baldocchi and Ryu, 2011). Again, however, differences among forests in the low latitudes were large. For example, the R^2 values for DBF were about twice as high as those for EBF (Figure 5). There were no clear latitudinal differences in R^2 for non-forest vegetation. At low latitudes, non-forest vegetation occurs mainly in regions with high seasonality of rainfall (e.g., savanna regions) and thus large variations in ET. Likewise, temperate non-forest vegetation types such as grass and crops show large variations in ET due to seasonal variation in radiation and temperature (e.g., Monteith and Moss, 1977).

4.4. Product performance versus vegetation-match between pixel and site

The linear regression analyses across all 40 sites showed that there was no relationship between the product evaluation results and the vegetation-match between pixel and site. Indirect evidence for this was also provided by the finding that the product with the largest pixel size and the lowest average VMI (GLEAM) performed best overall. Similar results were obtained by Hobeichi et al. (2018) and Jiménez et al. (2018). Hobeichi et al. (2018) investigated the effect of a vegetation mismatch between pixel and site on the performance of a merged ET product. For this they divided the eddy covariance sites in two groups, those for which the IGBP land cover type was the same as that of the pixel and those for which it was not. They used MODIS land cover data at the same spatial resolution (0.5°) as the merged ET product. No clear differences in the performance of the product were observed between the two groups of sites. Jiménez et al. (2018) investigated the effect of a vegetation mismatch between pixel and site on the performance of the GLEAM, PT-JPL, and MOD16 algorithms. The models were run with a common input dataset at a spatial resolution of 0.25°. For all three models a single vegetation match index was used (called homogeneity index). This index was calculated using MODIS IGBP land cover data (MCD12Q1) and MODIS vegetation cover data (MOD44B). Also in their study, no significant relationships were found between model performance and the homogeneity index. A challenge is to correctly define the vegetation match index (Hobeichi et al., 2018). In this study, only two vegetation categories were considered (forest and non-forest vegetation), as the land cover type-dependent parameters in MOD16 and ALEXI can be broadly grouped into these two categories (and GLEAM only distinguishes between these two categories). In the other studies, a match was only obtained if the specific IGBP land cover type corresponded. This may be too stringent if the parameters are similar among certain cover types. Understanding the sensitivity of the model outputs to the land cover type-specific parametrizations can help determine which of these approaches is more adequate. Nevertheless, the results obtained so far suggest that errors other than those caused by a vegetation mismatch between pixel and site are more important (Jiménez et al., 2018).

5. Conclusions

There is a geographical bias in the availability of evaluation data for remote sensing-based ET products in favor of the northern middle latitudes. To address this bias, three products (GLEAM, MOD16, ALEXI) were evaluated at 40 eddy covariance sites in the low latitudes. From MOD16, the discontinued collection 5 (C5) and the latest collection (C6) were evaluated. Two potential problems need to be considered when using eddy covariance observations of ET as reference data. First, eddy covariance data suffer from uncertainties related to the energy balance closure problem. Second, scale differences and classification errors can lead to a mismatch in vegetation between pixel and site (which in turn can complicate the comparisons). Because of the geographical bias in evaluation studies, it is unknown whether the products perform equally well at all latitudes.

The differences between MOD16 C5 and C6 were generally smaller than the differences with the other products. More work is needed, however, to determine the degree of consistency between the two collections.

Performance and ranking of the evaluated products depended on whether or not the eddy covariance ET data were corrected for the lack of energy balance closure. When using the unadjusted fluxes (ET_{orig}), GLEAM showed the best overall performance across the studied land cover types and climate zones, with the strongest correlations and smallest biases. Neither MOD16 nor ALEXI consistently outperformed the other. When using the corrected fluxes (ET_{ebc}), there was no product that stood out in terms of both low bias and strong correlations. The uncertainty associated with the energy balance closure problem affected the product biases more than the correlations. Most product evaluation studies use ET_{orig} as reference data. Use of both ET_{orig} and ET_{ebc} is recommended until a better understanding of the effect of the energy balance closure problem on ET is obtained.

Few studies have compared the performance of the products examined here. However, a comparison of results from studies that evaluated these products separately seems to confirm that GLEAM generally outperforms the other products (when using ET_{orig} as reference data). Latitudinal dependence of product performance was examined using the results of this study and those published in the literature. The comparison suggests that MOD16 and GLEAM perform better for low-latitude forests than for northern mid-latitude forests. However, regional differences, such as between the tropics and subtropics, can be large and were not detected because of the broad zonal bands used in this analysis. In the case of non-forest vegetation, the products show no clear latitudinal differences in performance.

No relationship was found between the product evaluation results and the degree of match between the vegetation at the flux tower site and that detected in the model pixel. More work is needed to understand the effect of a vegetation mismatch between pixel and site on product performance.

CRediT authorship contribution statement

Diego Salazar-Martínez: Formal analysis, Investigation, Visualization, Writing - Original Draft. **Friso Holwerda:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Supervision, Writing - Original Draft, Writing - Review & Editing. **Thomas R.H. Holmes:** Conceptualization, Investigation, Writing - Review & Editing. **Enrico A. Yépez:** Investigation, Resources, Writing - Review & Editing. **Christopher R. Hain:** Investigation, Writing - Review & Editing. **Susana Alvarado-Barrientos:** Investigation, Writing - Review & Editing. **Rest of the authors (alphabetical):** Investigation.

Data availability

The 8-day mean ET_{orig} and ET_{ebc} from the MexFlux sites (Dataset S1), as well as the 8-day mean ALEXI ET estimates for all but four eddy covariance sites (Dataset S2; see Section 2.1.3 and Table S2 for more details regarding the missing data) can be accessed in the supporting information. The rest of the data used in this study can be accessed through the open data portals as explained in Section 2.

Declaration of Competing Interest

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

Acknowledgements

This work used eddy covariance data acquired and shared by the FLUXNET community, including these networks: AmeriFlux, AfriFlux, AsiaFlux, CarboAfrica, CarboEuropeIP, CarboItaly, CarboMont, ChinaFlux, Fluxnet-Canada, GreenGrass, ICOS, KoFlux, LBA, NECC, OzFlux-TERN, TCOS-Siberia, and USCCC. The FLUXNET eddy covariance data processing and harmonization was carried out by the ICOS Ecosystem Thematic Center, AmeriFlux Management Project and Fluxdata project of FLUXNET, with the support of CDIAC, and the OzFlux, ChinaFlux and AsiaFlux offices.

We are grateful to Dr. Takashi Hirano and Dr. Shih-Chieh Chang for their permission to use data from the Palangkaraya Drained Forest (PDF) and Chi-Lan Mountain (CLM) sites, respectively (AsiaFlux), and to Dr. Michael Liddell for making available data from the Cape Tribulation, Cow Bay, and Robson Creek sites (OzFlux).

D.S.M. was supported by a graduate scholarship from CONACYT (number 595847). E.A.Y. acknowledges support from CONACYT (grant number 221014) and PROFAPI-ITSON.

N.E.R.R. was supported by a graduate scholarship from CONACYT (number 278991). G.A.P. acknowledges support from CONACYT PN (grant number 2017-6231) and USAID-USFS (grant number 12-IJ-11242306-033). E.G.C. acknowledges support from the Inter-American Institute for Global Change Research (grant numbers CRN2, CRN3-025) and CONACYT (grant number 175725). J.D.B. and T.A. acknowledge support from SEMARNAT-CONACYT (grant numbers 108000, CB 2008- 01 102855, CB 2013 220788). J.D.B. was supported by a graduate scholarship from CONACYT (number 211819). F.H. acknowledges support from CONACYT (grant number 187646) and UNAM-PAPIIT (grant number IB100113). S.A.B. acknowledges support from CONACYT (grant number INFR-2016-01-269269). B.F.E and J.U.S. acknowledge support from LANRESC-CONACYT (grant number 271544) and UNAM-PINCC-2020. J.U.S. was supported by a graduate scholarship from CONACYT (number 415123). J.G.P. acknowledges support from PROFAPI-ITSON.

Finally, we are grateful to the two anonymous reviewers whose comments and suggestions led to significant improvements in this work.

Supplementary materials

Supporting information associated with this article can be found, in the online version, at...

References

Aguilar, A.L., Flores, H., Crespo, G., Marín, M.I., Campos, I., Calera, A., 2018. Performance assessment of MOD16 in evapotranspiration evaluation in Northwestern Mexico. Water 10, 901. https://doi.org/10.3390/w10070901.

Allen, R.G., 2008. Quality assessment of weather data and micrometeorological flux—impacts on evapotranspiration calculation. Journal of Agricultural Meteorology 64 (4), 191–204. https://doi.org/10.2480/agrmet.64.4.5.

Alvarado-Barrientos, M.S., López-Adame, H., Lazcano-Hernández, H.E., Arellano-Verdejo, J., Hernández-Arana, H.A., 2021. Ecosystem- Atmosphere Exchange of CO_2, Water, and Energy in a Basin Mangrove of the Northeastern Coast of the Yucatan Peninsula. Journal of Geophysical Research Biogeosciences 126 (2), e2020JG005811. https://doi.org/10.1029/2020JG005811.

Anderson, M.C., Norman, J.M., Diak, G.R., Kustas, W.P., Mecikalski, J.R., 1997. A two-source time-integrated model for estimating surface fluxes using thermal infrared remote sensing. Remote Sensing of Environment 60(2), 195–216. https://doi.org/10.1016/S0034-4257(96)00215-5.

Anderson, M.C., Norman, J.M., Mecikalski, J.R., Otkin, J.A., Kustas, W.P., 2007. A climatological study of evapotranspiration and moisture stress across the continental United States based on thermal remote sensing: 1. Model formulation. Journal of Geophysical Research Atmospheres 112, DD10117. https://doi.org/10.1029/2006JD007506.

Anderson, M.C., Kustas, W.P., Norman, J.M., Hain, C.R., Mecikalski, J.R., Schultz, L., González-Dugo, M.P., Cammalleri, C., d'Urso, G., Pimstein, A., Gao, F., 2011. Mapping daily evapotranspiration at field to continental scales using geostationary and polar orbiting satellite imagery, Hydrology and Earth System Sciences 15, 223–239. https://doi.org/10.5194/hess-15-223-2011.

Ardö, J., Mölder, M., El-Tahir, B.A., Elkhidir, H.A.M., 2008. Seasonal variation of carbon fluxes in a sparse savanna in semi arid Sudan. Carbon Balance and Management 3, 7. https://doi.org/10.1186/1750-0680-3-7.

Badgley, G., Fisher, J.B., Jiménez, C., Tu, K.P., Vinukollu, R., 2015. On uncertainty in global terrestrial evapotranspiration estimates from choice of input forcing datasets. Journal of Hydrometeorology 16 (4), 1449–1455. http://dx.doi.org/10.1175/JHM-D-14-0040.1.

Baldocchi, D.D., Ryu, Y., 2011. A synthesis of forest evaporation fluxes—from days to years—as measured with eddy covariance. In: Levia, D.F., Carlyle-Moses, D. (Eds.), Forest Hydrology and Biogeochemistry: Synthesis of Past Research and Future Directions. Springer Verlag, New York, pp. 101–116.

Barr, A.G., Van der Kamp, G., Black, T.A., McCaughey, J.H., Nesic, Z., 2012. Energy balance closure at the BERMS flux towers in relation to the water balance of the White Gull Creek watershed 1999–2009. Agricultural and Forest Meteorology 153, 3–13. https://doi.org/10.1016/j.agrformet.2011.05.017.

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

Beringer, J., Hutley, L.B., Tapper, N.J., Cernusak, L.A., 2007. Savanna fires and their impact on net ecosystem productivity in North Australia. Global Change Biology 13 (5), 990–1004.

https://doi.org/10.1111/j.1365-2486.2007.01334.x.

Beringer, J., Hacker, J., Hutley, L.B., Leuning, R., Arndt, S.K., Amiri, R., Bannehr, L., Cernusak, L.A., Grover, S., Hensley, C., Hocking, D., Isaac, P., Jamali, H., Kanniah, K., Livesley, S., Neininger, B., Paw U, K.T., Sea, W., Straten, D., Tapper, N., Weinmann, R., Wood, S., Zegelin, S., 2011. SPECIAL—Savanna Patterns of Energy and Carbon Integrated across the Landscape. Bulletin of the American Meteorological Society 92 (11), 1467–1485. https://doi.org/10.1175/2011BAMS2948.1.

Beringer, J., Livesley, S.J., Randle, J., Hutley, L.B., 2013. Carbon dioxide fluxes dominate the greenhouse gas exchanges of a seasonal wetland in the wet–dry tropics of northern Australia. Agricultural and Forest Meteorology 182–183, 239–247. https://doi.org/10.1016/j.agrformet.2013.06.008.

Beringer, J., Hutley, L.B., McHugh, I., Arndt, S.K., Campbell, D., Cleugh, H.A., Cleverly, J., Resco de Dios, V., Eamus, D., Evans, B., Ewenz, C., Grace, P., Griebel, A., Haverd, V., Hinko-Najera, N., Huete, A., Isaac, P., Kanniah, K., Leuning, R., Liddell, M.J., Macfarlane, C., Meyer, W., Moore, C., Pendall, E., Phillips, A., Phillips, R.L., Prober, S.M., Restrepo-Coupe, N., Rutledge, S., Schroder, I., Silberstein, R., Southall, R., Yee, M.S., van Gorsel, E., Vote, C., Walker, J., Wardlaw, T., 2016. An introduction to the Australian and New Zealand flux tower network–OzFlux. Biogeosciences 13, 5895–5916. https://doi.org/10.5194/bg-13-5895-2016.

Bonal, D., Bosc, A., Ponton, S., Goret, J.-Y., Burban, B., Gross, P., Bonnefond, J.-M., Elbers, J., Longdoz, B., Epron, D., Guehl, J.-M., Granier, A., 2008. Impact of severe dry season on net ecosystem exchange in the Neotropical rainforest of French Guiana. Global Change Biology 14, 1917–1933. https://doi.org/10.1111/j.1365-2486.2008.01610.x.

Bristow, M., Hutley, L.B., Beringer, J., Livesley, S.J., Edwards, A.C., Arndt, S.K., 2016. Quantifying the relative importance of greenhouse gas emissions from current and future savanna land use change across northern Australia. Biogeosciences 13, 6285–6303. https://doi.org/10.5194/bg-13-6285-2016.

Bruijnzeel, L.A., Mulligan, M., Scatena, F.N., 2011. Hydrometeorology of tropical montane cloud forests: emerging patterns. Hydrological Processes 25, 465–498. https://doi.org/10.1002/hyp.7974.

Brust, C., Kimball, J.S., Maneta, M.P., Jencso, K., He, M., Reichle, R.H., 2021. Using SMAP Level-4 soil moisture to constrain MOD16 evapotranspiration over the contiguous USA. Remote Sensing of Environment 255, 112277. https://doi.org/10.1016/j.rse.2020.112277.

Brutsaert, W., 1982. Evaporation Into the Atmosphere: Theory, History, and Applications. Springer, Dordrecht. https://doi.org/10.1007/978-94-017-1497-6.

Cernusak, L.A., Hutley, L.B., Beringer, J., Holtum, J.A.M., Turner, B.L., 2011. Photosynthetic physiology of eucalypts along a sub-continental rainfall gradient in northern Australia. Agricultural and Forest Meteorology 151 (11), 1462–1470. https://doi.org/10.1016/j.agrformet.2011.01.006.

Chang, J.C., Hanna, R.S., 2004. Air quality model performance evaluation. Meteorology and Atmospheric Physics 87, 167–196. https://doi.org/10.1007/s00703-003-0070-7.

Chen, J.M., Liu, J., 2020. Evolution of evapotranspiration models using thermal and shortwave remote sensing data. Remote Sensing of Environment 237, 111594. https://doi.org/10.1016/j.rse.2019.111594.

Chen, H., Zhu, G., Shang, S., Qin, W., Zhang, Y., Su, Y., Zhang, K., Zhu, Y., Xu, C., 2022. Uncertainties in partitioning evapotranspiration by two remote sensing-based models. Journal of Hydrology 604, 127223. https://doi.org/10.1016/j.jhydrol.2021.127223.

Chu, H.-S., Chang, S.-C., Klemm, O., Lai, C.-W., Lin, Y.-Z., Wu, C.-C., Lin, J.-Y., Jiang, J.-Y., Chen, J., Gottgens, J.F., Hsia, Y.-J., 2014. Does canopy wetness matter? Evapotranspiration from a subtropical montane cloud forest in Taiwan. Hydrological Processes 28, 1190–1214. https://doi.org/10.1002/hyp.9662.

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

Cleverly, J., Boulain, N., Villalobos-Vega, R., Grant, N., Faux, R., Wood, C., Cook, P.G., Yu, Q., Leigh, A., Eamus, D., 2013. Dynamics of component carbon fluxes in a semi-arid Acacia woodland, central Australia. Journal of Geophysical Research Biogeosciences 118 (3), 1168–1185. https://doi.org/10.1002/jgrg.20101.

Cleverly, J., Eamus, D., Isaac, P., 2016. FLUXNET2015 AU-TTE Ti Tree East. Australia: N. p. https://doi.org/10.18140/FLX/1440205.

Cui, D., Liang, S., Wang, D., 2021. Observed and projected changes in global climate zones based on Köppen climate classification. WIREs Clim Change 12, e701. https://doi.org/10.1002/wcc.701.

Dawson, C.W., Mount, N.J., Abrahart, R.J., Shamseldin, A.Y., 2012. Ideal point error for model assessment in data-driven river flow forecasting. Hydrology and Earth System Sciences 16, 3049–3060. https://doi.org/10.5194/hess-16-3049-2012.

Delgado-Balbuena, J., Yépez, E.A., Paz-Pellat, F., Ángeles-Pérez, G., Aguirre-Gutiérrez, C., Alvarado-Barrientos, M.S., Arredondo, T., Ayala-Niño, F., Bullock, S.H., Castellanos, A.E., Cueva, A., Figueroa-Espinoza, B., Garatuza-Payán, J., González-del Castillo, E., González-Sosa, E., Guevara-Escobar, A., Hinojo-Hinojo, C., Paw U, K.-T., Lizárraga-Celaya, C., Maya-Delgado, Y., Oechel, W., Pérez-Ruiz, E.R., Quesada-Avendaño, M., Robles-Zazueta, C.A., Rodríguez, J.C., Rojas-Robles, N.E., Tarin-Terrazas, T., Troyo-Diéguez, E., Uuh-Sonda, J., Vargas-Terminel, M.L., Vargas, R., Vega-Puga, M.G., Verduzco, V.S., Vivoni, E.R., Watts, C.J., 2018. Base de datos de flujos verticales de dioxido de carbono en ecosistemas terrestres y costeros en México. Elementos para Políticas Públicas 2 (2), 93–108.

Delgado- Balbuena, J., Arredondo, J.T., Loescher, H.W., Pineda- Martínez, L.F., Carbajal, J.N., Vargas, R., 2019. Seasonal precipitation legacy effects determine the carbon balance of a semiarid grassland. Journal of Geophysical Research Biogeosciences 124, 987–1000. https://doi.org/10.1029/2018JG004799.

Denager, T., Looms, M.C., Sonnenborg, T.O., Jensen, K.H., 2020. Comparison of evapotranspiration estimates using the water balance and the eddy covariance methods. Vadose Zone Journal 19, e20032. https://doi.org/10.1002/vzj2.20032.

Doelling, D., 2012. CERES Level 3 SYN1DEG-DAYTerraCAqua netCDF file – Edition 3A. NASA Langley Atmospheric Science Data Center DAAC.

https://doi.org/10.5067/TERRA+AQUA/CERES/SYN1DEGDAY_L3.003A.

Dolman, A.J., Miralles, D.G., De Jeu, R.A.M., 2014. Fifty years since Monteith's 1965 seminal paper: the emergence of global ecohydrology. Ecohydrology 7, 897–902. https://doi.org/10.1002/eco.1505.

Drake, B., Hinkle, R., 2003–2006. FLUXNET2015 US-KS2 Kennedy Space Center (scrub oak). Dataset. https://doi.org/10.18140/FLX/1440075.

Eamus, D., Cleverly, J., Boulain, N., Grant, N., Faux, R., Villalobos-Vega, R., 2013. Carbon and water fluxes in an arid-zone Acacia savanna woodland: An analyses of seasonal patterns and responses to rainfall events. Agricultural and Forest Meteorology 182–183, 225– 238. http://dx.doi.org/10.1016/j.agrformet.2013.04.020.

Elshorbagy, A., Corzo, G., Srinivasulu, S., Solomatine, D.P., 2010. Experimental investigation of the predictive capabilities of data driven modeling techniques in hydrology - Part 1: Concepts and methodology. Hydrology and Earth System Sciences 14, 1931–1941. https://doi.org/10.5194/hess-14-1931-2010.

Figueroa-Espinoza, B., Sánchez-Mejía, Z., Uuh-Sonda, J.M., Salles, P., Méndez-Barroso, L., Gutiérrez-Jurado, H.A., 2021. Friction Velocity estimation using a 2D Sonic Anemometer in Coastal Zones. Atmósfera. https://doi.org/10.20937/ATM.52960.

Fisher, J.B., Tu, K.P., Baldocchi, D.D., 2008. Global estimates of the land–atmosphere water flux based on monthly AVHRR and ISLSCP-II data, validated at 16 FLUXNET sites. Remote Sensing of Environment 112, 901 – 919. https://doi.org/10.1016/j.rse.2007.06.025.

Fisher, J.B., Malhi, Y., Bonal, D., Da Rocha, H.R., De Araújo, A.C., Gamo, M., Goulden, M.L., Hirano, T., Huete, A.R., Kondo, H., Kumagai, T., Loescher, H.W., Miller, S., Nobre, A.D., Nouvellon, Y., Oberbauer, S.F., Panuthai, S., Roupsard, O., Saleska, S., Tanaka, K., Tanaka, N., Tu, K.P., Von Randow, C., 2009. The land–atmosphere water flux in the tropics. Global Change Biology 15, 2694–2714. https://doi.org/10.1111/j.1365-2486.2008.01813.x.

Fisher, J.B., Melton, F., Middleton, E., Hain, C., Anderson, M., Allen, R., McCabe, M.F., Hook, S., Baldocchi, D., Townsend, P.A., Kilic, A., Tu, K., Miralles, D.D., Perret, J., Lagouarde, J.-P., Waliser, D., Purdy, A.J., French, A., Schimel, D., Famiglietti, J.S., Stephens, G., Wood, E.F., 2017. The future of evapotranspiration: Global requirements for ecosystem functioning, carbon and climate feedbacks, agricultural management, and water resources, Water Resources Research 53, 2618–2626. https://doi.org/10.1002/2016WR020175.

Foken, T., Aubinet, M., Finnigan, J.J., Leclerc, M.Y., Mauder, M., Paw U, K.T., 2011. Results of a panel discussion about the energy balance closure correction for trace gases. Bulletin of the American Meteorological Society 92 (4), ES13–ES18.

Frank, J.M., Massman, W.J., Swiatek, E., Zimmerman, H.A., Ewers, B.E., 2016. All sonic anemometers need to correct for transducer and structural shadowing in their velocity measurements. Journal of Atmospheric and Oceanic Technology 33 (1), 149–167. https://doi.org/10.1175/JTECH-D-15-0171.1.

Gao, Z., Liu, H., Katul, G.G., Foken, T., 2017. Non-closure of the surface energy balance explained by phase difference between vertical velocity and scalars of large atmospheric eddies. Environmental Research Letters 12, 034025. https://doi.org/10.1088/1748–9326/aa625b.

Gash, J.H.C., 1979. An analytical model of rainfall interception by forests. Quarterly Journal of the Royal

Meteorological Society 105, 43–55. https://doi.org/10.1002/qj.49710544304.

González del Castillo, E., Paw U, K.T., Sánchez-Azofeifa, A., 2018. Turbulence scales for eddy covariance quality control over a tropical dry forest in complex terrain. Agricultural and Forest Meteorology 249, 390–406. https://doi.org/10.1016/j.agrformet.2017.11.014.

Goulden, M. 2000–2004. FLUXNET2015 BR-Sa3 Santarem-Km83-Logged Forest, Dataset. https://doi.org/10.18140/FLX/1440033.

Hansen, M.C., Song, X., DiMiceli, C., Carroll, M., Sohlberg, R., Kim, D.-H., Townshend, J., 2018. MEaSURES Vegetation Continuous Fields ESDR Algorithm Theoretical Basis Document (ATBD) Version 2.0. University of Maryland, College Park, MD. Available at: https://lpdaac.usgs.gov/documents/144/VCF5KYR_ATBD.pdf (last access: 29 August 2021).

Hidalgo-Sánchez, M., Ángeles-Pérez, G., Yépez, E.A., Plascencia-Escalante, F.O., Delgado-Balbuena, J., González-Martínez, T.M., 2021. Evapotranspiration and energy exchange in a temperate forest in Mexico. Tecnología y Ciencias del Agua 12(2), 490–537. https://doi.org/10.24850/j-tyca-2021-02-11.

Hirano, T., Kusin, K., Limin, S., Osaki, M., 2015. Evapotranspiration of tropical peat swamp forests. Global Change Biology 21, 1914–1927. https://doi.org/10.1111/gcb.12653.

Hobeichi, S., Abramowitz, G., Evans, J., Ukkola, A., 2018. Derived Optimal Linear Combination Evapotranspiration (DOLCE): a global gridded synthesis ET estimate. Hydrology and Earth System Sciences 22, 1317–1336. https://doi.org/10.5194/hess-22-1317-2018.

Hain, C.R., Anderson, M.C., 2017. Estimating morning change in land surface temperature from MODIS day/night observations: Applications for surface energy balance modeling. Geophysical Research Letters 44, 9723–9733. https://doi.org/10.1002/2017GL074952.

Holmes, T.R.H., Hain, C.R., Crow, W.T., Anderson, M.C., Kustas, W.P., 2018. Microwave implementation of two-source energy balance approach for estimating evapotranspiration. Hydrology and Earth System Sciences 22, 1351–1369. https://doi.org/10.5194/hess-22-1351-2018.

Holwerda, F., Alvarado-Barrientos, M.S., González-Martínez, T.M., 2016. Surface energy exchange in a tropical montane cloud forest environment: flux partitioning, and seasonal and land cover-related variations. Agricultural and Forest Meteorology 228, 13–28. https://doi.org/10.1016/j.agrformet.2016.06.011.

Holwerda, F., Meesters, A.G.C.A., 2019. Soil evaporation in a shaded coffee plantation derived from eddy covariance measurements. Journal of Geophysical Research Biogeosciences 124 (6), 1472–1490. https://doi.org/10.1029/2018JG004911.

Hsieh, C.-I., Huang, C.-W., Kiely, G., 2009. Long-term estimation of soil heat flux by single layer soil temperature. International Journal of Biometeorology 53, 113–123. https://doi.org/10.1007/s00484-008-0198-8.

Hu, G., Jia, L., Menenti, M., 2015. Comparison of MOD16 and LSA-SAF MSG evapotranspiration products over Europe for 2011. Remote Sensing of Environment 156, 510–526. http://dx.doi.org/10.1016/j.rse.2014.10.017.

Hutley, L.B., Beringer, J., Isaac, P.R., Hacker, J.M., Cernusak, L.A., 2011. A sub-continental scale living laboratory: Spatial patterns of savanna vegetation over a rainfall gradient in northern Australia.

Agricultural and Forest Meteorology 151, 1417–1428. https://doi.org/10.1016/j.agrformet.2011.03.002.

Isaac, P., Cleverly, J., McHugh, I., Van Gorsel, E., Ewenz, C., Beringer, J., 2017. OzFlux data: network integration from collection to curation. Biogeosciences 14, 2903–2928. https://doi.org/10.5194/bg-14-2903-2017.

Jiménez, C., Martens, B., Miralles, D.M., Fisher, J.B., Beck, H.E., Fernández-Prieto, D., 2018. Exploring the merging of the global land evaporation WACMOS-ET products based on local tower measurements. Hydrology and Earth System Sciences 22, 4513–4533. https://doi.org/10.5194/hess-22-4513-2018.

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. https://doi.org/10.1016/j.agrformet.2018.01.022.

Khan, M.S., Baik, J., Choi, M., 2020. Inter-comparison of evapotranspiration datasets over heterogeneous landscapes across Australia. Advances in Space Research 66 (3), 533–545. https://doi.org/10.1016/j.asr.2020.04.037.

Kim, H.W., Hwang, K., Mu, Q., Lee, S.O., Choi, M., 2012. Validation of MODIS 16 global terrestrial evapotranspiration products in various climates and land cover types in Asia. KSCE Journal of Civil Engineering 16 (2), 229–238. https://doi.org/10.1007/s12205-012-0006-1.

Komatsu, H., 2005. Forest categorization according to dry-canopy evaporation rates in the growing season: comparison of the Priestley–Taylor coefficient values from various observation sites. Hydrological Processes 19, 3873– 3896. https://doi.org/10.1002/hyp.5987.

Liddell, M., 2013. Cape Tribulation OzFlux tower site OzFlux: Australian and New Zealand Flux Research and Monitoring hdl: 102.100.100/14242.

Lu, Y., Lu, S., Horton, R., Ren, T., 2014. An empirical model for estimating soil thermal conductivity from texture, water content, and bulk density. Soil Science Society of America Journal 78, 1859–1868. https://doi.org/10.2136/sssaj2014.05.0218.

Martens, B., Miralles, D.G., Lievens, H., Van der Schalie, R., De Jeu, R.A.M., Fernández-Prieto, D., Beck, H.E., Dorigo, W.A., Verhoest, N.E.C., 2017. GLEAM v3: satellite-based land evaporation and root-zone soil moisture. Geoscientific Model Development 10, 1903–1925. https://doi.org/10.5194/gmd-10-1903-2017.

Mauder, M., Genzel, S., Fu, J., Kiese, R., Soltani, M., Steinbrecher, R., Zeeman, M., Banerjee, T., De Roo, F., Kunstmann, H., 2018. Evaluation of energy balance closure adjustment methods by independent evapotranspiration estimates from lysimeters and hydrological simulations. Hydrological Processes 32, 39–50. https://doi.org/10.1002/Hyp.11397.

Mauder, M., Foken, T., Cuxart, J., 2020. Surface-energy-balance closure over land: a review. Boundary Layer Meteorology 177, 395–426. https://doi.org/10.1007/s10546-020-00529-6.

Mayocchi, C.L., Bristow, K.L., 1995. Soil surface heat flux: some general questions on measurements. Agricultural and Forest Meteorology 75, 43–50. https://doi.org/10.1016/0168-1923(94)02198-S.

McCabe, M.F., Ershadi, A., Jimenez, C., Miralles, D.G., Michel, D., Wood, E.F., 2016. The GEWEX LandFlux project: evaluation of model evaporation using tower-based and globally gridded forcing data. Geoscientific Model Development 9, 283–305. https://doi.org/10.5194/gmd-9-283-2016.

Melo, D.C.D., Anache, J.A.A., Borges, V.P., Miralles, D.G., Martens, B., Fisher, J.B., Nóbrega, R.L.B., Moreno, A., Cabral, O.M.R., Rodrigues, T.R., Bezerra, B., Silva, C.M.S., Meira Neto, A.A., Moura, M.S.B., Marques, T.V., Campos, S., Nogueira, J.S., Rosolem, R., Souza, R.M.S., Antonino, A.C.D., Holl, D., Galleguillos, M., Perez-Quezada, J.F., Verhoef, A., Kutzbach, L., Lima, J.R.S., Souza, E.S., Gassman, M.I., Perez, C.F., Tonti, N., Posse, G., Rains, D., Oliveira, P.T.S., Wendland, E., 2021. Are remote sensing evapotranspiration models reliable across South American ecoregions? Water Resources Research 57, e2020WR028752. https://doi.org/10.1029/2020WR028752.

Melton, F.S., Huntington, J., Grimm, R., Herring, J., Hall, M., Rollison, D., Erickson, T., Allen, R., Anderson, M., Fisher, J.B., Kilic, A., Senay, G.B., Volk, J., Hain, C., Johnson, L., Ruhoff, A., Blankenau, P., Bromley, M., Carrara, W., Daudert, B., Doherty, C., Dunkerly, C., Friedrichs, M., Guzman, A., Halverson, G., Hansen, J., Harding, J., Kang, Y., Ketchum, D., Minor, B., Morton, C., Ortega-Salazar, S., Ott, T., Ozdogan, M., ReVelle, P.M., Schull, M., Wang, C., Yang, Y., Anderson, R.G., 2021. OpenET: Filling a critical data gap in water management for the western United States. Journal of the American Water Resources Association 1–24. https://doi.org/10.1111/1752-1688.12956.

Merbold, L., Ardö, J., Arneth, A., Scholes, R.J., Nouvellon, Y., de Grandcourt, A., Archibald, S., Bonnefond, J.M., Boulain, N., Brueggemann, N., Bruemmer, C., Cappelaere, B., Ceschia, E., El-Khidir, H.A.M., El-Tahir, B.A., Falk, U., Lloyd, J., Kergoat, L., Le Dantec, V., Mougin, E., Muchinda, M., Mukelabai, M.M., Ramier, D., Roupsard, O., Timouk, F., Veenendaal, E.M., Kutsch, W.L., 2009. Precipitation as driver of carbon fluxes in 11 African ecosystems. Biogeosciences 6 (6), 1027–1041. https://doi.org/10.5194/bg-6-1027-2009.

Michel, D., Jiménez, C., Miralles, D.G., Jung, M., Hirschi, M., Ershadi, A., Martens, B., McCabe, M.F., Fisher, J.B., Mu, Q., Seneviratne, S.I., Wood, E.F., Fernández-Prieto, D., 2016. The WACMOS-ET project – Part 1: Tower-scale evaluation of four remote-sensing-based evapotranspiration algorithms. Hydrology and Earth System Sciences 20, 803–822. https://doi.org/10.5194/hess-20-803-2016.

Miralles, D.G., Gash, J.H., Holmes, T.R.H., De Jeu, R.A.M., Dolman, A.J., 2010. Global canopy interception from satellite observations. Journal of Geophysical Research Atmospheres 115, D16122. https://doi.org/10.1029/2009JD013530.

Miralles, D.G., Holmes, T.R.H., De Jeu, R.A.M., Gash, J.H., Meesters, A.G.C.A., Dolman, A.J., 2011. Global land-surface evaporation estimated from satellite-based observations. Hydrology and Earth System Sciences 15, 453–469. https://doi.org/10.5194/hess-15-453-2011.

Miralles, D.G., Jiménez, C., Jung, M., Michel, D., Ershadi, A., McCabe, M.F., Hirschi, M., Martens, B., Dolman, A.J., Fisher, J.B., Mu, Q., Seneviratne, S.I., Wood, E.F., Fernández-Prieto, D., 2016. The WACMOS-ET project – Part 2: Evaluation of global terrestrial evaporation data sets. Hydrology and Earth System Sciences 20, 823–842. https://doi.org/10.5194/hess-20-823-2016.

Monteith, J.L., 1965. Evaporation and environment. Symposia of the Society for Experimental Biology 19, 205–234.

Monteith, J.L., Moss, C.J., 1977. Climate and the efficiency of crop production in Britain. Philosophical Transactions of the Royal Society B 281, 277–294. https://doi.org/10.1098/rstb.1977.0140.

Mu, Q., Heinsch, F.A., Zhao, M., Running, S.W., 2007. Development of a global evapotranspiration algorithm based on MODIS and global meteorology data. Remote Sensing of Environment 111, 519–536. https://doi.org/10.1016/j.rse.2007.04.015.

Mu, Q., Zhao, M., Running, S.W., 2011. Improvements to a MODIS global terrestrial evapotranspiration algorithm. Remote Sensing of Environment 115, 1781–1800. https://doi.org/10.1016/j.rse.2011.02.019.

Mu, Q., Zhao, M., Running, S.W., 2013. MODIS Global Terrestrial Evapotranspiration (ET) Product (NASA MOD16A2/A3). Algorithm Theoretical Basis Document, Collection 5.

Myneni, R.B., Hoffman, S., Knyazikhin, Y., Privette, J.L., Glassy, J., Tian, Y., Wang, Y., Song, X., Zhang, Y., Smith, G.R., Lotsch, A., Friedl, M., Morisette, J.T., Votava, P., Nemani, R.R., Running, S.W., 2002. Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data. Remote Sensing of Environment 83 (1–2), 214–231. https://doi.org/10.1016/S0034-4257(02)00074-3.

MY-PSO, 2003–2009. FLUXNET2015 MY-PSO Pasoh Forest Reserve (PSO), Dataset. https://doi.org/10.18140/FLX/1440240.

Pastorello, G., Agarwal, D., Papale, D., Samak, T., Trotta, C., Ribeca, A., Poindexter, C., Faybishenko, B., Gunter, D., Hollowgrass, R., Canfora, E., 2014. Observational data patterns for time series data quality assessment. In: 2014 IEEE 10th International Conference on e-Science, pp. 271–278. https://doi.org/10.1109/eScience.2014.45.

Pastorello, G., et al., 2020. The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data. Scientific Data 7, 225. https://doi.org/10.1038/s41597-020-0534-3.

Pérez-Ruiz, E.R., Vivoni, E.R., Yépez, E.A., Rodríguez, J.C., Gochis, D.J., Robles-Morua, A., Delgado-Balbuena, J., Adams, D.K., 2021. Landscape controls on water-energy-carbon fluxes across different ecosystems during the North American monsoon. Journal of Geophysical Research Biogeosciences 126, e2020JG005809. https://doi.org/10.1029/2020JG005809.

Piñeiro, G., Perelman, S., Guerschman, J.P., Paruelo, J.M., 2008. How to evaluate models: Observed vs. predicted or predicted vs. observed? Ecological modelling 216, 316–322. https://doi.org/10.1016/j.ecolmodel.2008.05.006.

Posse, G., Lewczuk, N., Richter, K., Cristiano, P., 2016. Carbon and water vapor balance in a subtropical pine plantation. iForest - Biogeosciences and Forestry 9 (5), 736–742. https://doi.org/10.3832/ifor1815-009.

Priestley, C.H.B., Taylor, R.J., 1972. On the assessment of surface heat flux and evaporation using large-scale parameters. Monthly Weather Review 100, 81–92.

Ramoelo, A., Majozi, N., Mathieu, R., Jovanovic, N., Nickless, A., Dzikiti, S., 2014. Validation of global evapotranspiration product (MOD16) using flux tower data in the African Savanna, South Africa. Remote Sensing 6, 7406–7423. http://dx.doi.org/10.3390/rs6087406.

Reitz, M., Senay, G.B., Sanford, W.E., 2017. Combining remote sensing and water-balance evapotranspiration estimates for the conterminous United States. Remote Sensing 9, 1181. https://doi.org/10.3390/rs9121181.

Rojas- Robles, N.E., Garatuza- Payán, J., Álvarez- Yépiz, J.C., Sánchez- Mejía, Z.M., Vargas, R., Yépez, E.A., 2020. Environmental controls on carbon and water fluxes in an old- growth tropical dry forest. Journal of Geophysical Research Biogeosciences 125, e2020JG005666. https://doi.org/10.1029/2020JG005666.

Ruhoff, A.L., Paz, A.R., Aragao, L.E.O.C., Mu, Q., Malhi, Y., Collischonn, W., Rocha, H.R., Running, S.W., 2013. Assessment of the MODIS global evapotranspiration algorithm using eddy covariance measurements and hydrological modelling in the Rio Grande basin. Hydrological Sciences Journal 58 (8), 1658–1676. http://dx.doi.org/10.1080/02626667.2013.837578.

Running, S.W., Mu, Q., Zhao, M., Moreno, A., 2019. MODIS Global Terrestrial Evapotranspiration (ET) Product (MOD16A2/A3 and Year- end Gap- filled MOD16A2GF/A3GF) NASA Earth Observing System MODIS Land Algorithm (for Collection 6). Version 2.0, Jan 3rd, 2019. https://modis-land.gsfc.nasa.gov/pdf/MOD16UsersGuideV2.022019.pdf.

Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., Behringer, D., Hou, Y.-T., Chuang, H.-Y, Iredell, M., Ek, M., Meng, J., Yang, R., Mendez, M.P., van den Dool, H., Zhang, Q., Wang, W., Chen, M., Becker, E., 2011. NCEP Climate Forecast System Version 2 (CFSv2) 6-hourly Products. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory. https://doi.org/10.5065/D61C1TXF.

Saleska, S., 2002–2011. FLUXNET2015 BR-Sa1 Santarem-Km67-Primary Forest, Dataset. https://doi.org/10.18140/FLX/1440032.

Santanello, J.A., Jr., Friedl, M.A., 2003. Diurnal covariation in soil heat flux and net radiation. Journal of Applied Meteorology 42 (6), 851–862.

Sauer, T.J., 2002. 5.5 Heat flux density. In: Dane, J.H., Topp, G.C. (Eds.), Methods of Soil Analysis: Part 4 Physical Methods. Soil Science Society of America, Madison, WI, pp. 1233–1248. https://doi.org/10.2136/sssabookser5.4.c52.

Schimel, D., Pavlick, R., Fisher, J.B., Asner, G.P., Saatchi, S., Townsend, P., Miller, C., Frankenberg, C., Hibbard, K., Cox, P., 2015. Observing terrestrial ecosystems and the carbon cycle from space. Global Change Biology 21, 1762 –1776. https://doi.org/10.1111/gcb.12822.

Schroder, I., Kuske, T., Zegelin, S., 2014. Eddy Covariance Dataset for Arcturus (2011–2013). Geoscience Australia, Canberra. https://doi.org/102.100.100/14249.

Senay, G.B., Bohms, S., Singh, R.K., Gowda, P.H., Velpuri, N.M., Alemu, H., Verdin, J.P., 2013. Operational evapotranspiration mapping using remote sensing and weather datasets: A new parameterization for the SSEB approach. Journal of the American Water Resources Association 49 (3), 577–591. https://doi.org/10.1111/jawr.12057.

Senay, G.B., Kagone, S., Velpuri, N.M., 2020. Operational global actual evapotranspiration: development, evaluation, and dissemination. Sensors 20, 1915. https://doi.org/10.3390/s20071915.

Sheffield, J., Wood, E.F., Pan, M., Beck, H., Coccia, G., Serrat-Capdevila, A., Verbist, K., 2018. Satellite remote sensing for water resources management: Potential for supporting sustainable development in data-poor regions. Water Resources Research 54, 9724–9758. https://doi.org/10.1029/2017WR022437.

Shuttleworth, W.J., Gurney, R.J., Hsu, A.Y., Ormsby, J.P., 1989. FIFE: the variation in energy partitioning at surface flux sites. IAHS Red Book Series no. 186, 67–74.

Shuttleworth, W.J., 1993. Evaporation. In: Maidment, D.R. (Ed.), Handbook of Hydrology. McGraw-Hill, New York, pp. 4.1–4.53.

Souza, V.D.A., Roberti, D.R., Ruhoff, A.L., Zimmer, T., Adamatti, D.S., Gonçalves, L.G.G.D., Diaz, M.B., Alves, R.D.C.M., Moraes, O.L.L.D., 2019. Evaluation of MOD16 algorithm over irrigated rice paddy using flux tower measurements in Southern Brazil. Water 11, 1911. https://doi.org/10.3390/w11091911.

Stoy, P.C., Mauder, M., Foken, T., Marcolla, B., Boegh, E., Ibrom, A., Arain, M.A., Arneth, A., Aurela, M., Bernhofer, C., Cescatti, A., Dellwik, E., Duce, P., Gianelle, D., Van Gorsel, E., Kiely, G., Knohl, A., Margolis, H., McCaughey, H., Merbold, L., Montagnani, L., Papale, D., Reichstein, M., Saunders, M., Serrano-Ortiz, P., Sottocornola, M., Spano, D., Vaccari, F., Varlagin, A., 2013. A data-driven analysis of energy balance closure across FLUXNET research sites: the role of landscape scale heterogeneity. Agricultural and Forest Meteorology 171, 137–152. http://dx.doi.org/10.1016/j.agrformet.2012.11.004.

Su, Z., 2002. The Surface Energy Balance System (SEBS) for estimation of turbulent heat fluxes. Hydrology and Earth System Sciences 6, 85–100. https://doi.org/10.5194/hess-6-85-2002.

Tang, R., Shao, K., Li, Z.-L., Wu, H., Tang, B.-H., Zhou, G., Zhang, L., 2015. Multiscale validation of the 8-day MOD16 evapotranspiration product using flux data collected in China. IEEE Journal of selected topics in applied earth observations and remote sensing 8, 1478–1486. https://doi.org/10.1109/JSTARS.2015.2420105.

Twine, T.E., Kustas, W.P., Norman, J.M., Cook, D.R., Houser, P.R., Meyers, T.P., Prueger, J.H., Starks, P.J., Wesely, M.L., 2000. Correcting eddy-covariance flux underestimates over a grassland. Agricultural and Forest Meteorology 103, 279–300. https://doi.org/10.1016/S0168-1923(00)00123-4.

Uuh-Sonda, J.M., Figueroa-Espinoza, B., Gutiérrez-Jurado, H.A., Méndez-Barroso, L.A., 2022. Ecosystem productivity and evapotranspiration dynamics of a seasonally dry tropical forest of the Yucatan Peninsula. Journal of Geophysical Research Biogeosciences 127, e2019JG005629. https://doi.org/10.1029/2019JG005629.

Vargas, R., Yépez, E.A., Andrade, J.L., Ángeles, G., Arredondo, T., Castellanos, A.E., Delgado-Balbuena, J., Garatuza-Payán, J., González del Castillo, E., Oechel, W., Rodríguez, J.C., Sánchez-Azofeifa, A., Velasco, E., Vivoni, E.R., Watts, C., 2013. Progress and opportunities for monitoring greenhouse gases fluxes in Mexican ecosystems: the MexFlux network. Atmósfera 26, 325–336. https://doi.org/10.1016/S0187-6236(13)71079-8.

Velpuri, N.M., Senay, G.B., Singh, R.K., Bohms, S., Verdin, J.P., 2013. A comprehensive evaluation of two MODIS evapotranspiration products over the conterminous United States: Using point and gridded FLUXNET and water balance ET. Remote Sensing of Environment 139, 35–49. http://dx.doi.org/10.1016/j.rse.2013.07.013.

Verduzco, V.S., Garatuza-Payán, J., Yépez, E.A., Watts, C.J., Rodríguez, J.C., Robles-Morua, A., Vivoni, E.R., 2015. Variations of net ecosystem production due to seasonal precipitation differences in a tropical dry forest of northwest Mexico. Journal of Geophysical Research Biogeosciences 120, 2081–2094. https://doi.org/10.1002/2015JG003119.

Verduzco, V.S., Vivoni, E.R., Yépez, E.A., Rodríguez, J.C., Watts, C.J., Tarin, T., Garatuza-Payán, J., Robles-Morua, A., Ivanov, V.Y., 2018. Climate change impacts on net ecosystem productivity in a subtropical shrubland of northwestern México. Journal of Geophysical Research Biogeosciences 123, 688–711. https://doi.org/10.1002/2017JG004361.

Villarreal, S., Vargas, R., 2021. Representativeness of FLUXNET sites across Latin America. Journal of Geophysical Research Biogeosciences 126, e2020JG006090. https://doi.org/10.1029/2020JG006090.

Vinukollu, R.V., Wood, E.F., Ferguson, C.R., Fisher, J.B., 2011a. Global estimates of evapotranspiration for climate studies using multi-sensor remote sensing data: Evaluation of three process-based approaches. Remote Sensing of Environment 115, 801–823. https://doi.org/10.1016/j.rse.2010.11.006.

Vinukollu, R.V., Meynadier, R., Sheffield, J., Wood, E.F., 2011b. Multi-model, multi-sensor estimates of global evapotranspiration: climatology, uncertainties and trends. Hydrological Processes 25, 3993–4010. https://doi.org/10.1002/hyp.8393.

Wan, Z., Zhang, Y., Zhang, Q., Li, Z.-L., 2004. Quality assessment and validation of the MODIS global land surface temperature. International Journal of Remote Sensing 25 (1), 261–274. https://doi.org/10.1080/0143116031000116417.

Wang, Z.H., Bou-Zeid, E., 2012. A novel approach for the estimation of soil ground heat flux. Agricultural and Forest Meteorology 154−155, 214−221. https://doi.org/10.1016/j.agrformet.2011.12.001.

Wang, K., Dickinson, R.E., 2012. A review of global terrestrial evapotranspiration: Observation, modeling, climatology, and climatic variability. Reviews of Geophysics 50, RG2005. https://doi.org/10.1029/2011RG000373.

Willmott, C.J., 1982. Some comments on the evaluation of model performance. Bulletin of the American Meteorological Society 63 (11), 1309–1313. https://doi.org/10.1175/1520-0477(1982)063<1309:SCOTEO>2.0.CO;2.

Wilson, K., Goldstein, A., Falge, E., Aubinet, M., Baldocchi, D., Berbigier, P., Bernhofer, C., Ceulemans, R., Dolman, H., Field, C., Grelle, A., Ibrom, A., Law, B.E., Kowalski, A., Meyers, T., Moncrieff, J., Monson, R., Oechel, W., Tenhunen, J., Valentini, R., Verma, S., 2002. Energy balance closure at FLUXNET sites. Agricultural and Forest Meteorology 113 (1–4), 223–243.

https://doi.org/10.1016/S0168-1923(02)00109-0.

Wolf, S., Eugster, W., Potvin, C., Buchmann, N., 2011. Strong seasonal variations in net ecosystem CO_2 exchange of a tropical pasture and afforestation in Panama. Agricultural and Forest Meteorology 151, 1139–1151. https://doi.org/10.1016/j.agrformet.2011.04.002.

Yilmaz, M.T., Anderson, M.C., Zaitchik, B., Hain, C.R., Crow, W.T., Ozdogan, M., Chun, J.A., Evans, J., 2014. Comparison of prognostic and diagnostic surface flux modeling approaches over the Nile River basin. Water Resources Research 50 (1), 386-408. https://doi.org/10.1002/2013WR014194.

Yu, G.-R., Wen, X.-F., Sun, X.-M., Tanner, B.D., Lee, X., Chen, J.-Y., 2006. Overview of ChinaFLUX and evaluation of its eddy covariance measurement. Agricultural and Forest Meteorology 137, 125–137. https://doi.org/10.1016/j.agrformet.2006.02.011.

Zhang, K.[Ke], Kimball, J.S., Running, S.W., 2016. A review of remote sensing based actual evapotranspiration estimation. WIREs Water 3, 834–853. https://doi.org/10.1002/wat2.1168.

Zhang, K.[Kun], Zhu, G., Ma, J., Yang, Y., Shang, S., Gu, C., 2019. Parameter analysis and estimates for the MODIS evapotranspiration algorithm and multiscale verification. Water Resources Research 55, 2211–2231. https://doi.org/10.1029/2018WR023485.

Zhang, Y.[Yongqiang], Peña-Arancibia, J.L., McVicar, T.R., Chiew, F.H.S., Vaze, J., Liu, C., Lu, X., Zheng, H., Wang, Y., Liu, Y.Y., Miralles, D.G., Pan, M., 2016. Multi-decadal trends in global terrestrial evapotranspiration and its components. Scientific Reports 6, 19124. https://doi.org/10.1038/srep19124.

Zhang, Y.[Yongqiang], Kong, D., Gan, R., Chiew, F.H.S., McVicar, T.R., Zhang, Q., Yang, Y., 2019.

Coupled estimation of 500 m and 8-day resolution global evapotranspiration and gross primary production in 2002–2017. Remote Sensing of Environment 222, 165–182. https://doi.org/10.1016/j.rse.2018.12.031.

CRediT authorship contribution statement

Diego Salazar-Martínez: Formal analysis, Investigation, Visualization, Writing - Original Draft.
Friso Holwerda: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Supervision, Writing - Original Draft, Writing - Review & Editing.
Thomas R.H. Holmes: Conceptualization, Investigation, Writing - Review & Editing.
Enrico A. Yépez: Investigation, Resources, Writing - Review & Editing.
Christopher R. Hain: Investigation, Writing - Review & Editing.
Susana Alvarado-Barrientos: Investigation, Writing - Review & Editing.
Rest of the authors (alphabetical): Investigation.

Declaration of 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.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

March 24, 2022, Mexico City, Mexico

Dear Dr Anagnostou,

We are pleased to submit the revision of the manuscript entitled “Evaluation of remote sensing-based evapotranspiration products at low-latitude eddy covariance sites” for publication in Journal of Hydrology.

The manuscript has been submitted and reviewed before by Journal of Hydrology under the manuscript number HYDROL43166_R1. The present submission is the revised version of this manuscript. We also provide point-by-point replies to the reviewers comments.

We look very much forward to your response.

Sincerely,

Dr. Friso Holwerda (corresponding author)

Instituto de Ciencias de la Atmósfera y Cambio Climático
Universidad Nacional Autónoma de México
Circuito de la Investigación s/n
Ciudad Universitaria
Coyoacan, 04510
México D.F.
México
E-mail address: friso.holwerda@gmail.com
Phone: 0052-55-56224088

Figure 1

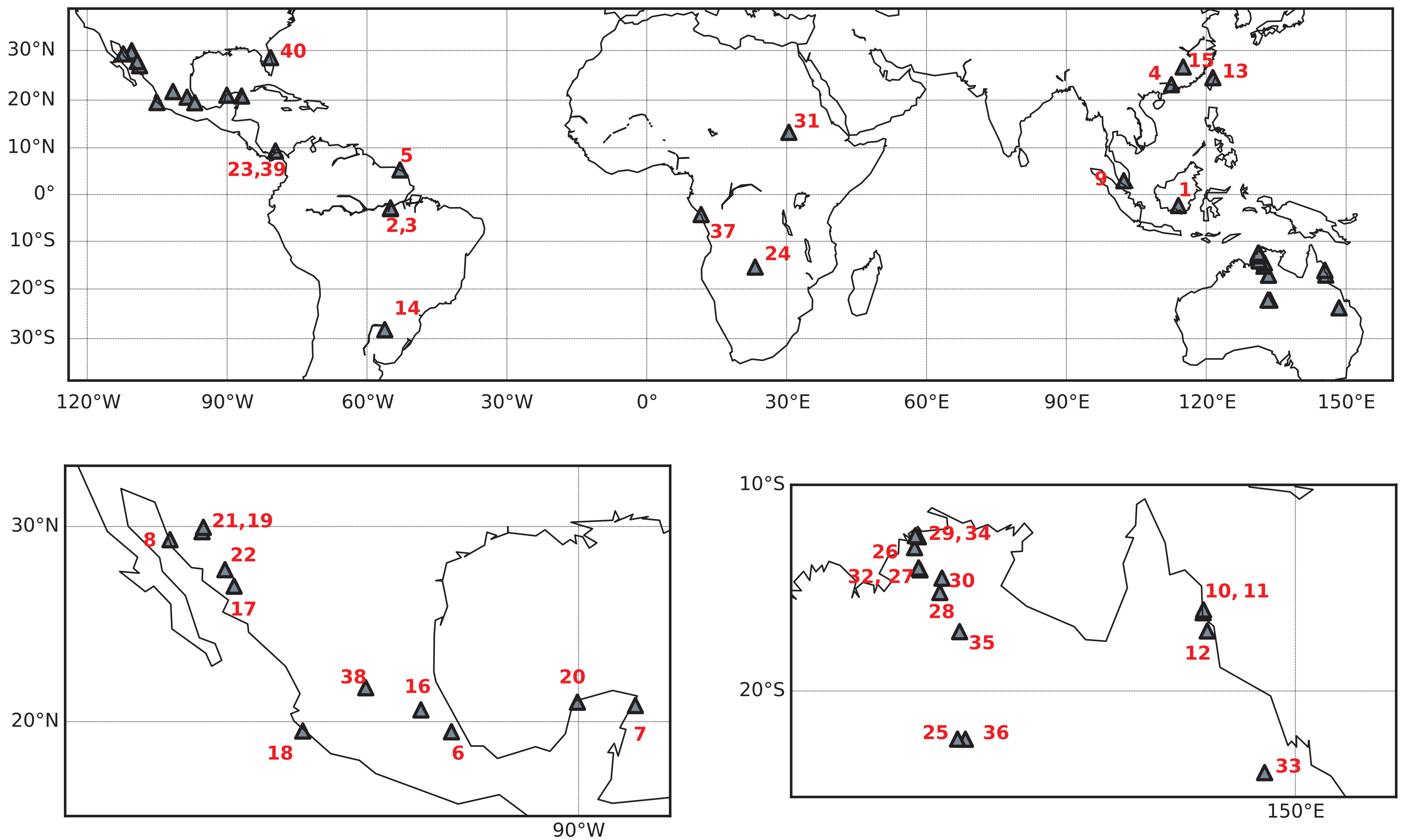
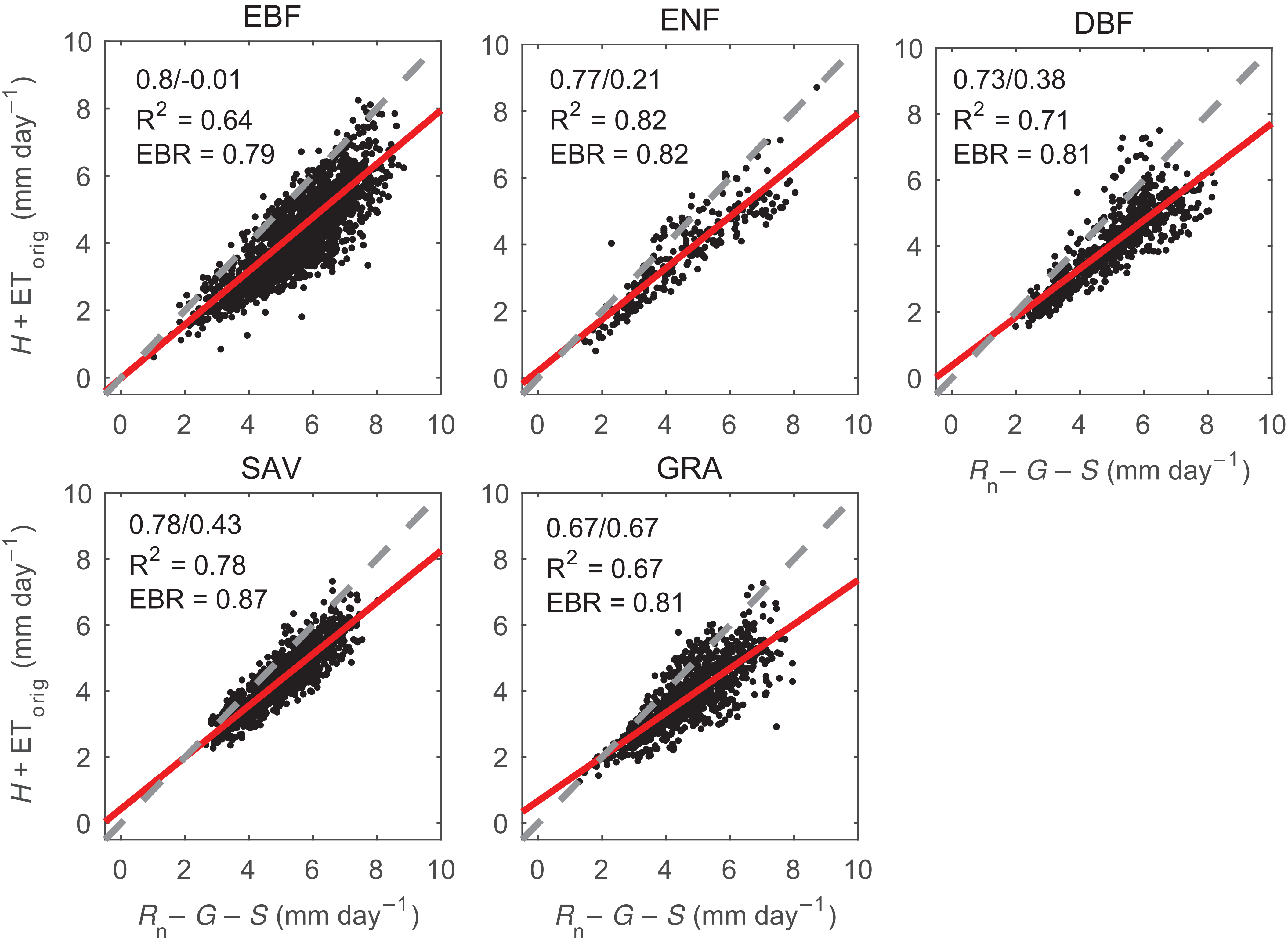
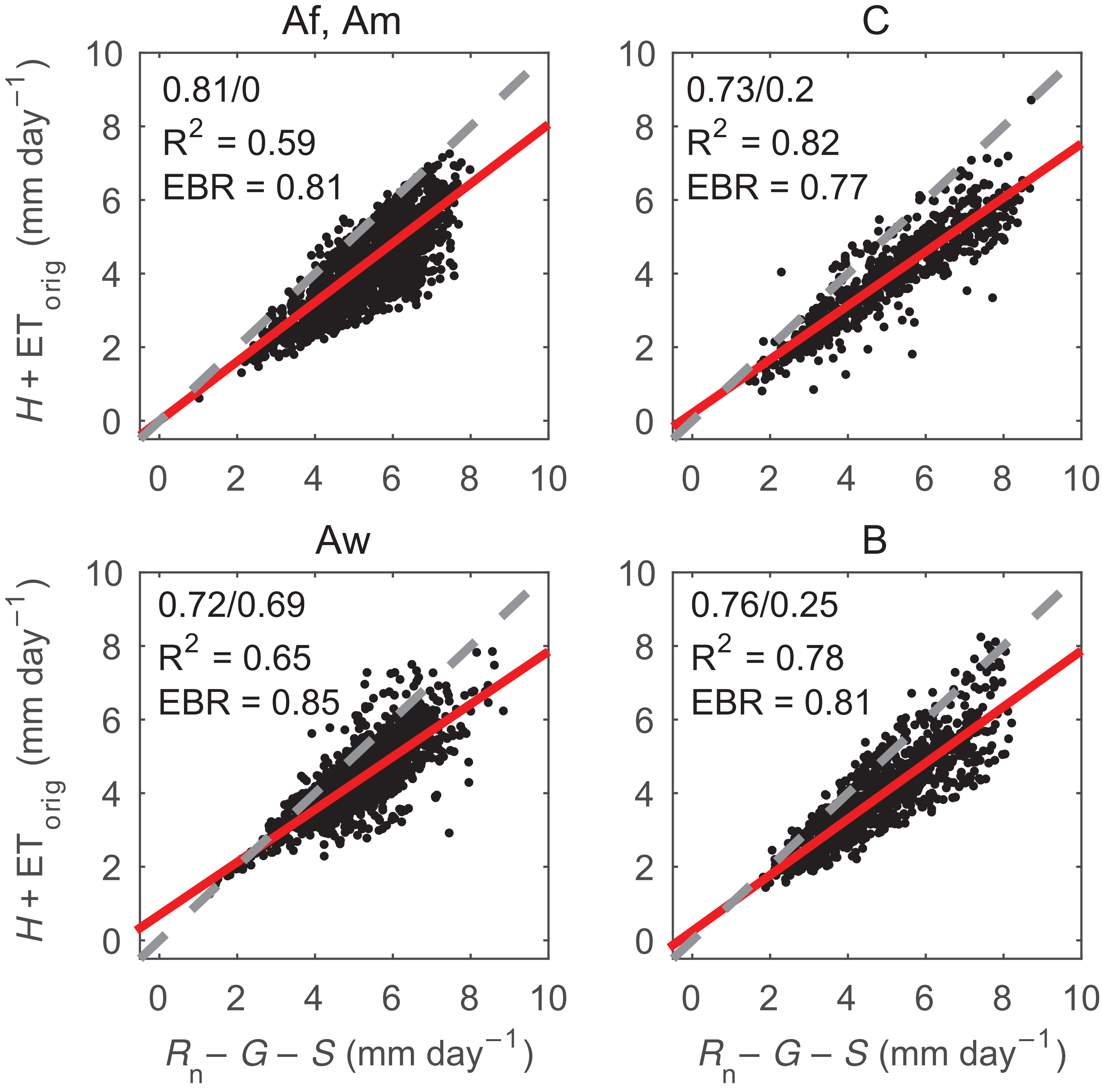

30°N
20°N
10°N
0°
10°S
20°S
30°S
120°W
90°W
60°W
30°W
0°
30°E
60°E
90°E
120°E
150°E
40
23,39
5
2,3
14
31
37
24
4
15
13
9
1
30°N
20°N
90°W
21,19
8
22
17
38
16
20
7
18
6
10°S
20°S
150°E
29, 34
26
32, 27
30
28
35
10, 11
12
25
36
33

Figure 2



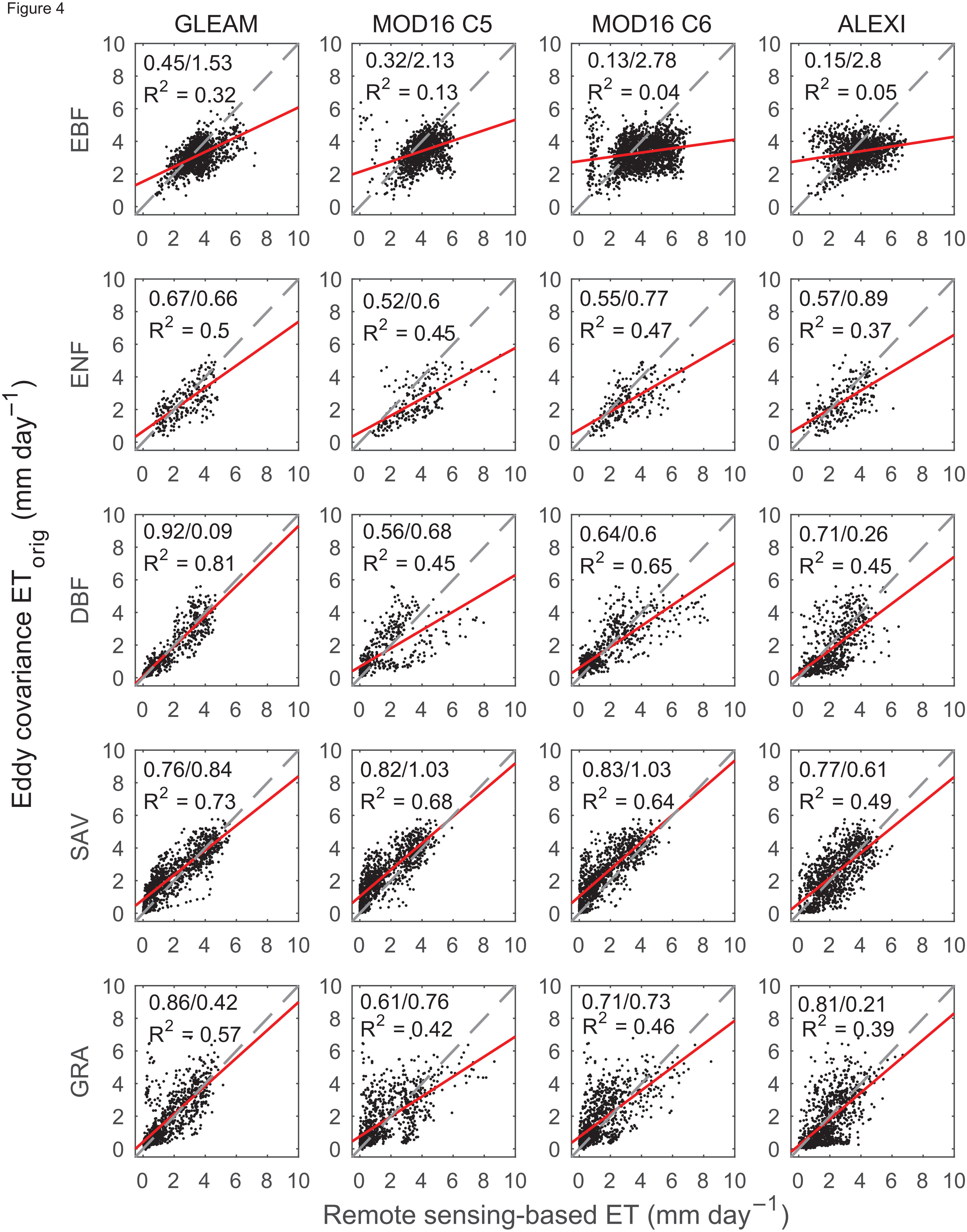
EBF
0.8/-0.01
R^2 = 0.64
EBR = 0.79
ENF
0.77/0.21
R^2 = 0.82
EBR = 0.82
DBF
0.73/0.38
R^2 = 0.71
EBR = 0.81
SAV
0.78/0.43
R^2 = 0.78
EBR = 0.87
GRA
0.67/0.67
R^2 = 0.67
EBR = 0.81
$H + ET_{orig}$ (mm day^{-1})
$R_n - G - S$ (mm day^{-1})

Figure 3



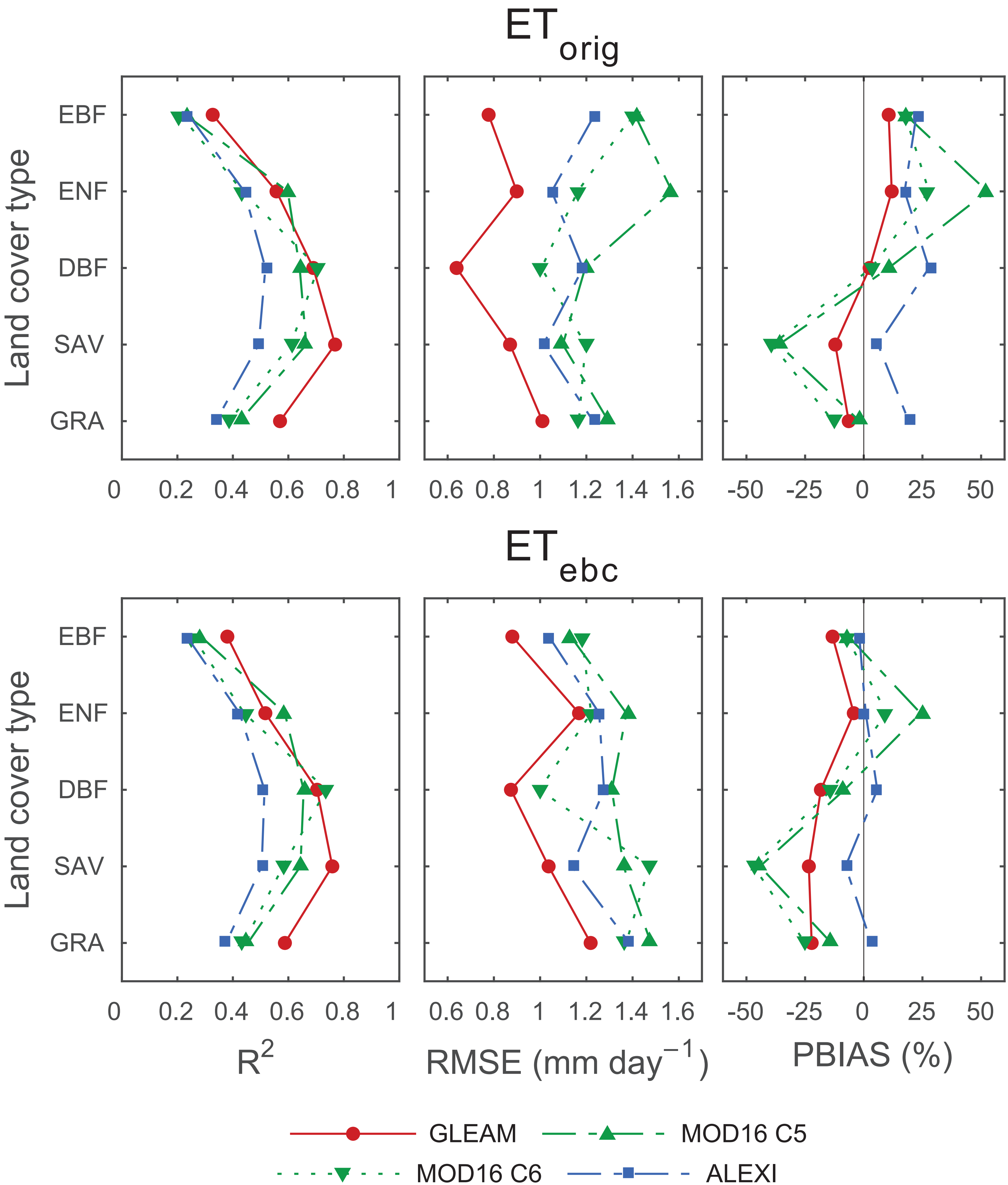
Af, Am
0.81/0
R^2 = 0.59
EBR = 0.81
C
0.73/0.2
R^2 = 0.82
EBR = 0.77
Aw
0.72/0.69
R^2 = 0.65
EBR = 0.85
B
0.76/0.25
R^2 = 0.78
EBR = 0.81
$H + ET_{orig}$ (mm day^{-1})
$R_n - G - S$ (mm day^{-1})
0
2
4
6
8
10

Figure 4



GLEAM
MOD16 C5
MOD16 C6
ALEXI
EBF
ENF
DBF
SAV
GRA
0.45/1.53
R^2 = 0.32
0.32/2.13
R^2 = 0.13
0.13/2.78
R^2 = 0.04
0.15/2.8
R^2 = 0.05
0.67/0.66
R^2 = 0.5
0.52/0.6
R^2 = 0.45
0.55/0.77
R^2 = 0.47
0.57/0.89
R^2 = 0.37
0.92/0.09
R^2 = 0.81
0.56/0.68
R^2 = 0.45
0.64/0.6
R^2 = 0.65
0.71/0.26
R^2 = 0.45
0.76/0.84
R^2 = 0.73
0.82/1.03
R^2 = 0.68
0.83/1.03
R^2 = 0.64
0.77/0.61
R^2 = 0.49
0.86/0.42
R^2 = 0.57
0.61/0.76
R^2 = 0.42
0.71/0.73
R^2 = 0.46
0.81/0.21
R^2 = 0.39
Eddy covariance ET_{orig} (mm day^{-1})
Remote sensing-based ET (mm day^{-1})

Figure 5



ET_{orig}
ET_{ebc}
Land cover type
EBF
ENF
DBF
SAV
GRA
0
0.2
0.4
0.6
0.8
1
0.6
0.8
1
1.2
1.4
1.6
-50
-25
0
25
50
R^2
RMSE (mm day^{-1})
PBIAS (%)
GLEAM
MOD16 C5
MOD16 C6
ALEXI

Figure 6

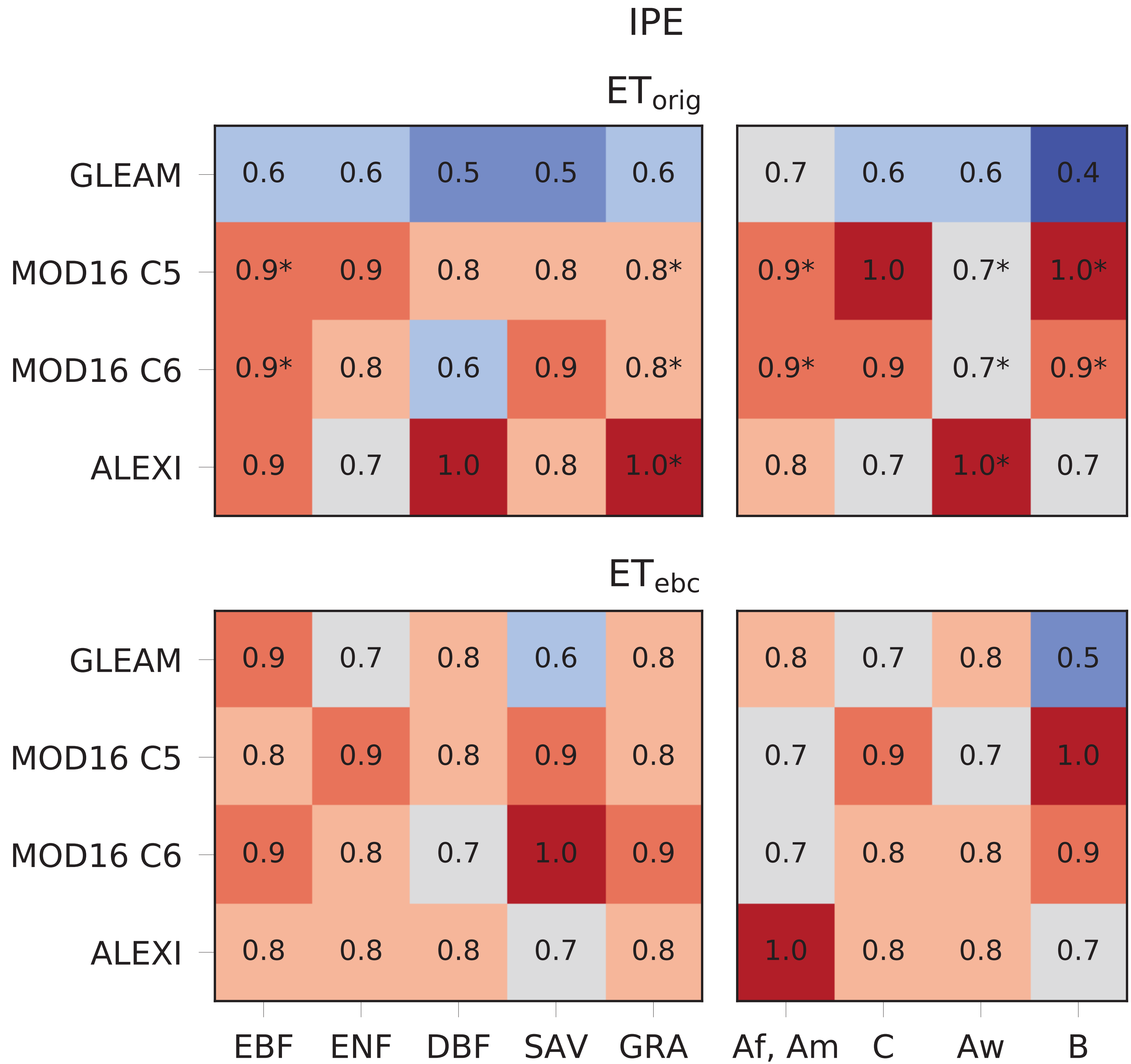
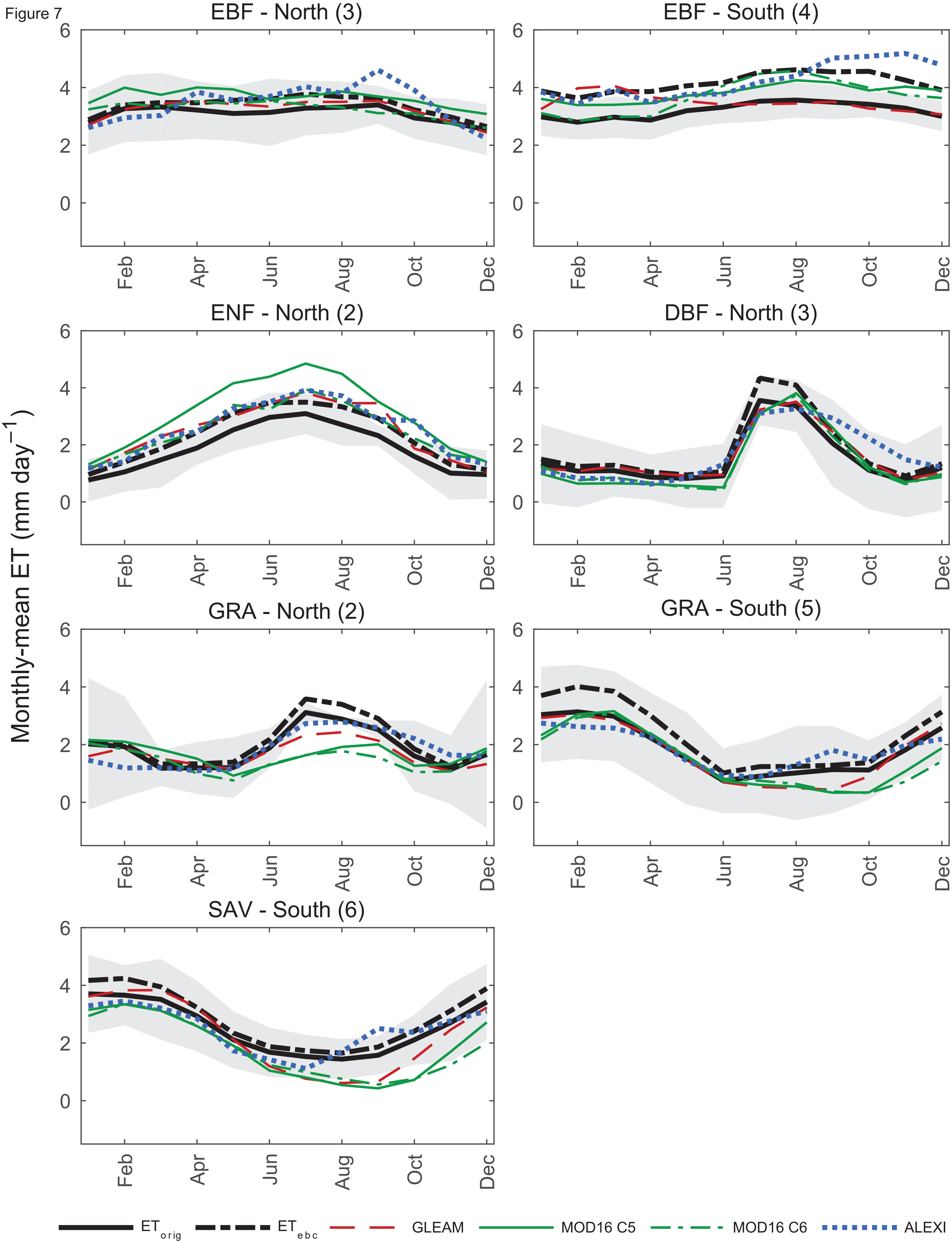
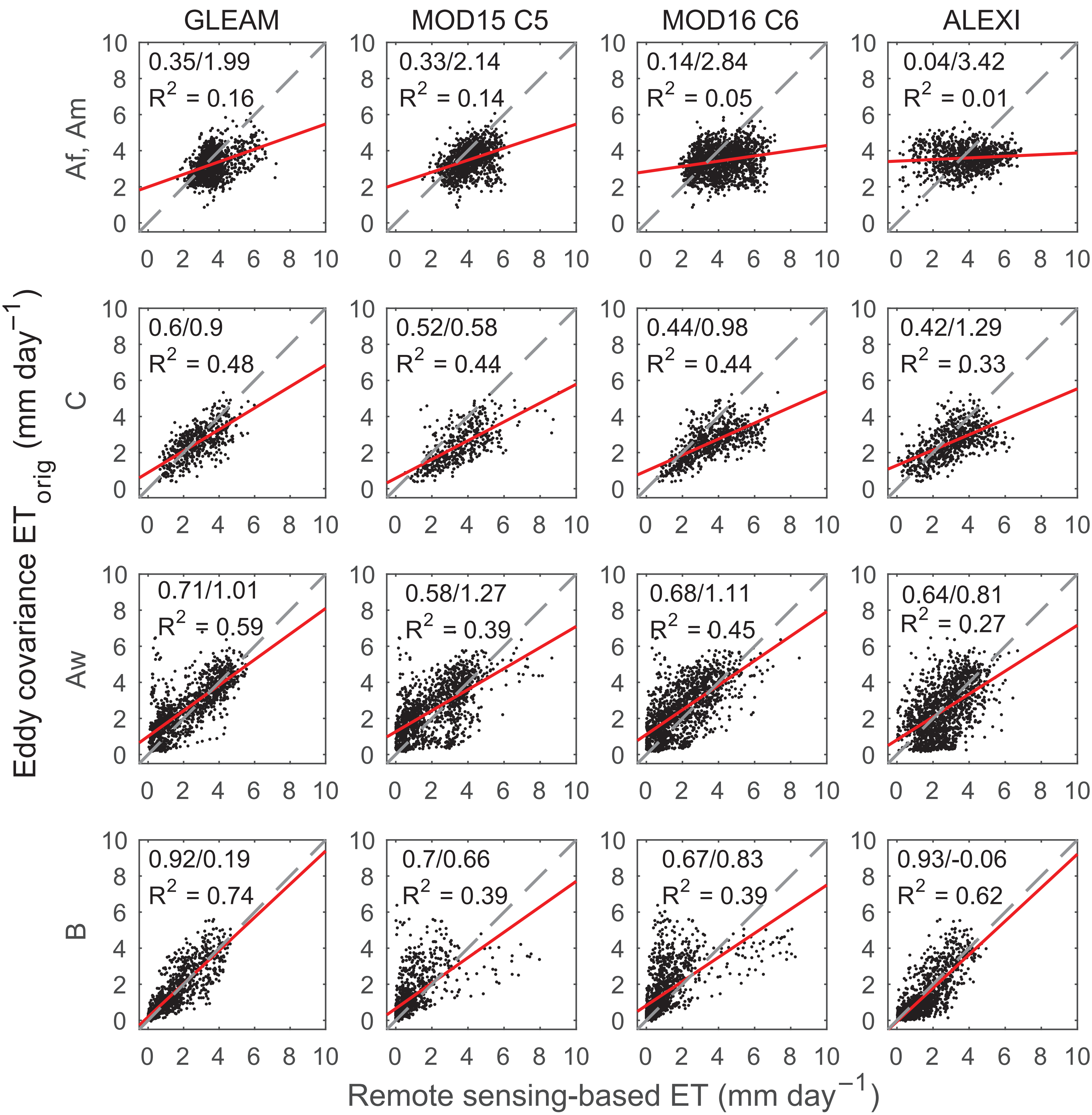

IPE
ET orig
GLEAM 0.6 0.6 0.5 0.5 0.6 0.7 0.6 0.6 0.4
MOD16 C5 0.9* 0.9 0.8 0.8 0.8* 0.9* 1.0 0.7* 1.0*
MOD16 C6 0.9* 0.8 0.6 0.9 0.8* 0.9* 0.9 0.7* 0.9*
ALEXI 0.9 0.7 1.0 0.8 1.0* 0.8 0.7 1.0* 0.7
ET ebc
GLEAM 0.9 0.7 0.8 0.6 0.8 0.8 0.7 0.8 0.5
MOD16 C5 0.8 0.9 0.8 0.9 0.8 0.7 0.9 0.7 1.0
MOD16 C6 0.9 0.8 0.7 1.0 0.9 0.7 0.8 0.8 0.9
ALEXI 0.8 0.8 0.8 0.7 0.8 1.0 0.8 0.8 0.7
EBF ENF DBF SAV GRA Af, Am C Aw B

Figure 7



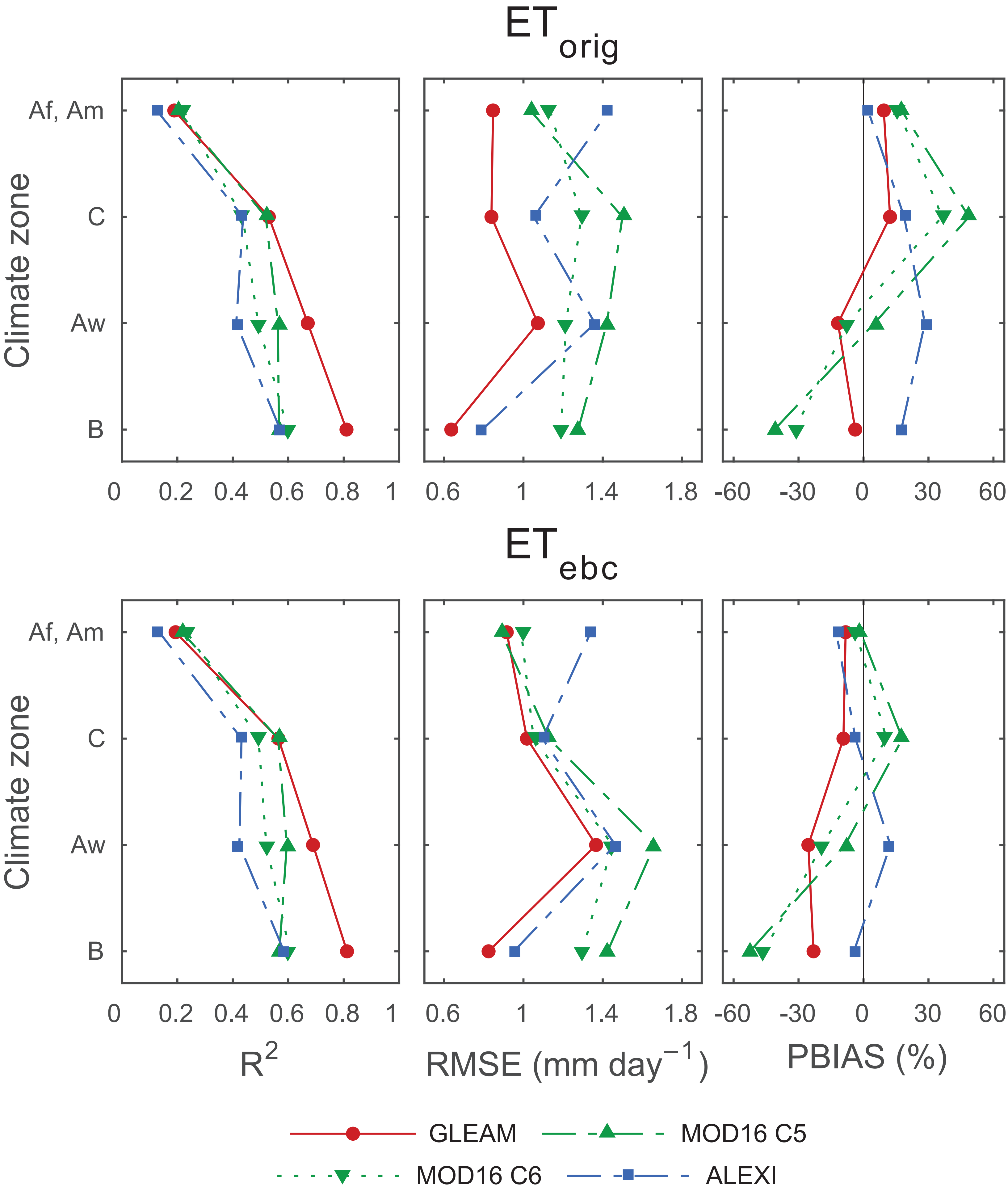
EBF - North (3)
EBF - South (4)
ENF - North (2)
DBF - North (3)
GRA - North (2)
GRA - South (5)
SAV - South (6)
Monthly-mean ET (mm day^{-1})
6
4
2
0
Feb
Apr
Jun
Aug
Oct
Dec
ET_{orig}
ET_{ebc}
GLEAM
MOD16 C5
MOD16 C6
ALEXI

Figure 8



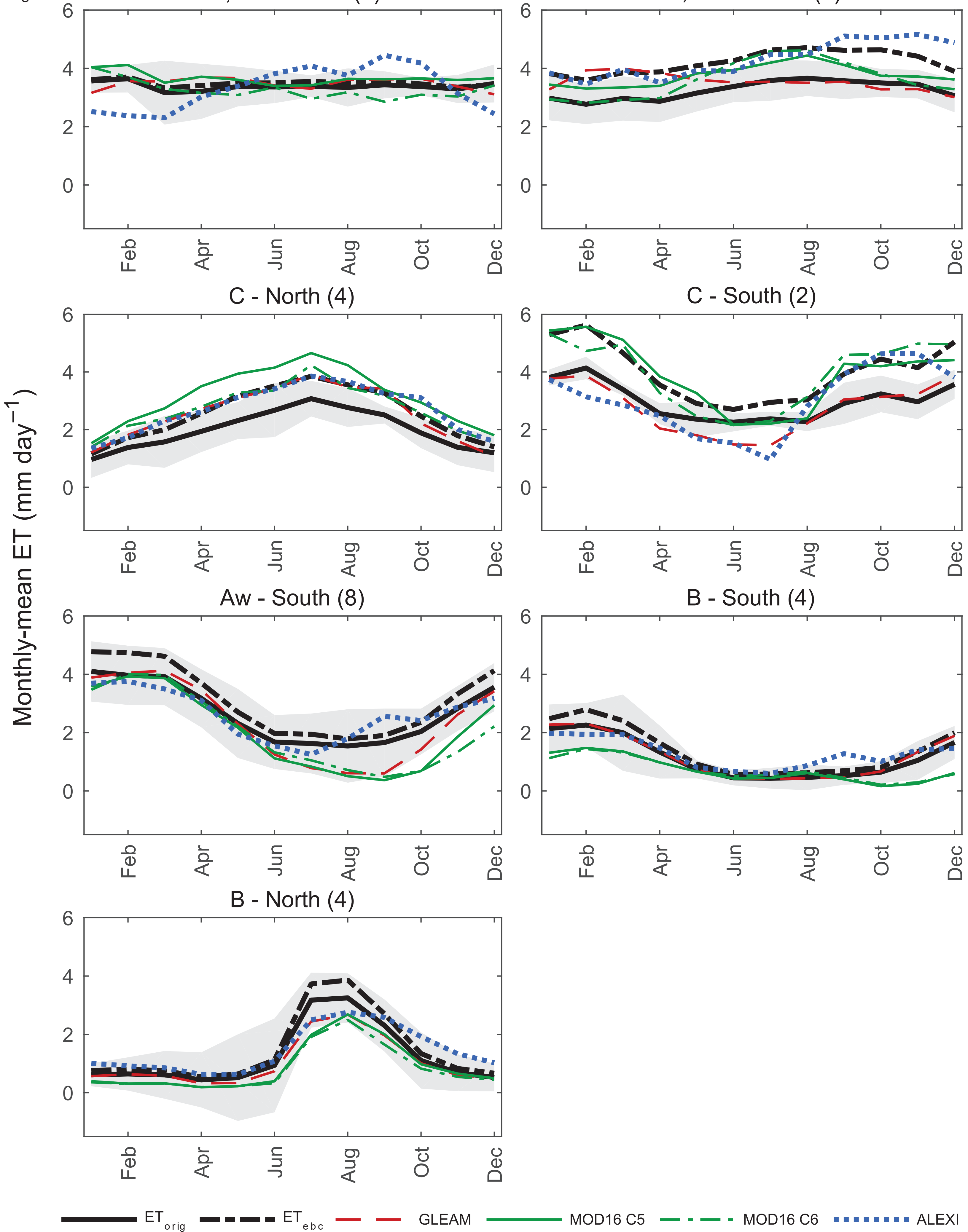
GLEAM
MOD15 C5
MOD16 C6
ALEXI
Af, Am
C
Aw
B
0.35/1.99
R2 = 0.16
0.33/2.14
R2 = 0.14
0.14/2.84
R2 = 0.05
0.04/3.42
R2 = 0.01
0.6/0.9
R2 = 0.48
0.52/0.58
R2 = 0.44
0.44/0.98
R2 = 0.44
0.42/1.29
R2 = 0.33
0.71/1.01
R2 = 0.59
0.58/1.27
R2 = 0.39
0.68/1.11
R2 = 0.45
0.64/0.81
R2 = 0.27
0.92/0.19
R2 = 0.74
0.7/0.66
R2 = 0.39
0.67/0.83
R2 = 0.39
0.93/-0.06
R2 = 0.62
Eddy covariance ET orig (mm day−1)
Remote sensing-based ET (mm day−1)

Figure 9



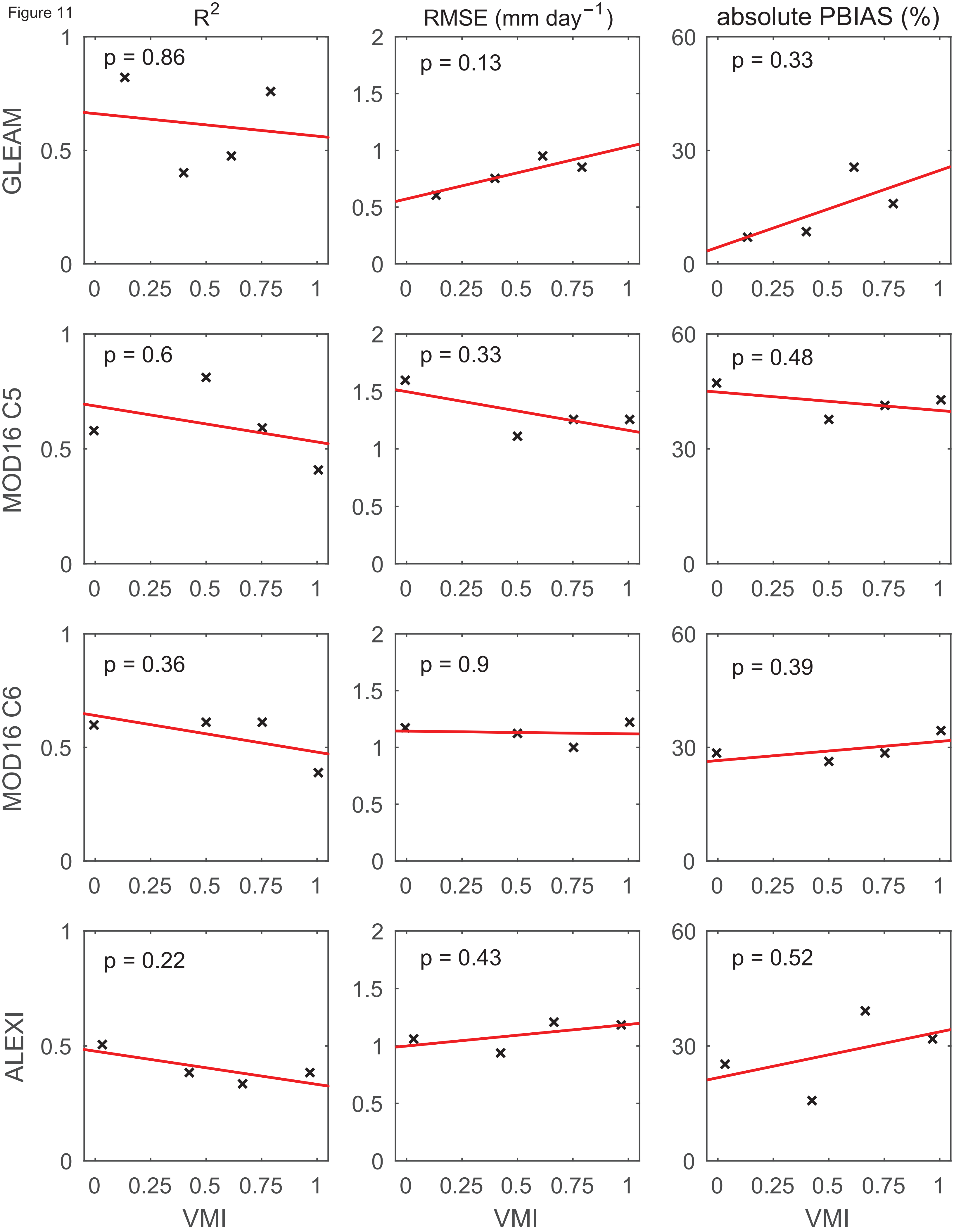
ET orig
ET ebc
Climate zone
Af, Am
C
Aw
B
0
0.2
0.4
0.6
0.8
1
0.6
1
1.4
1.8
-60
-30
0
30
60
R2
RMSE (mm day−1)
PBIAS (%)
GLEAM
MOD16 C5
MOD16 C6
ALEXI

Figure 10



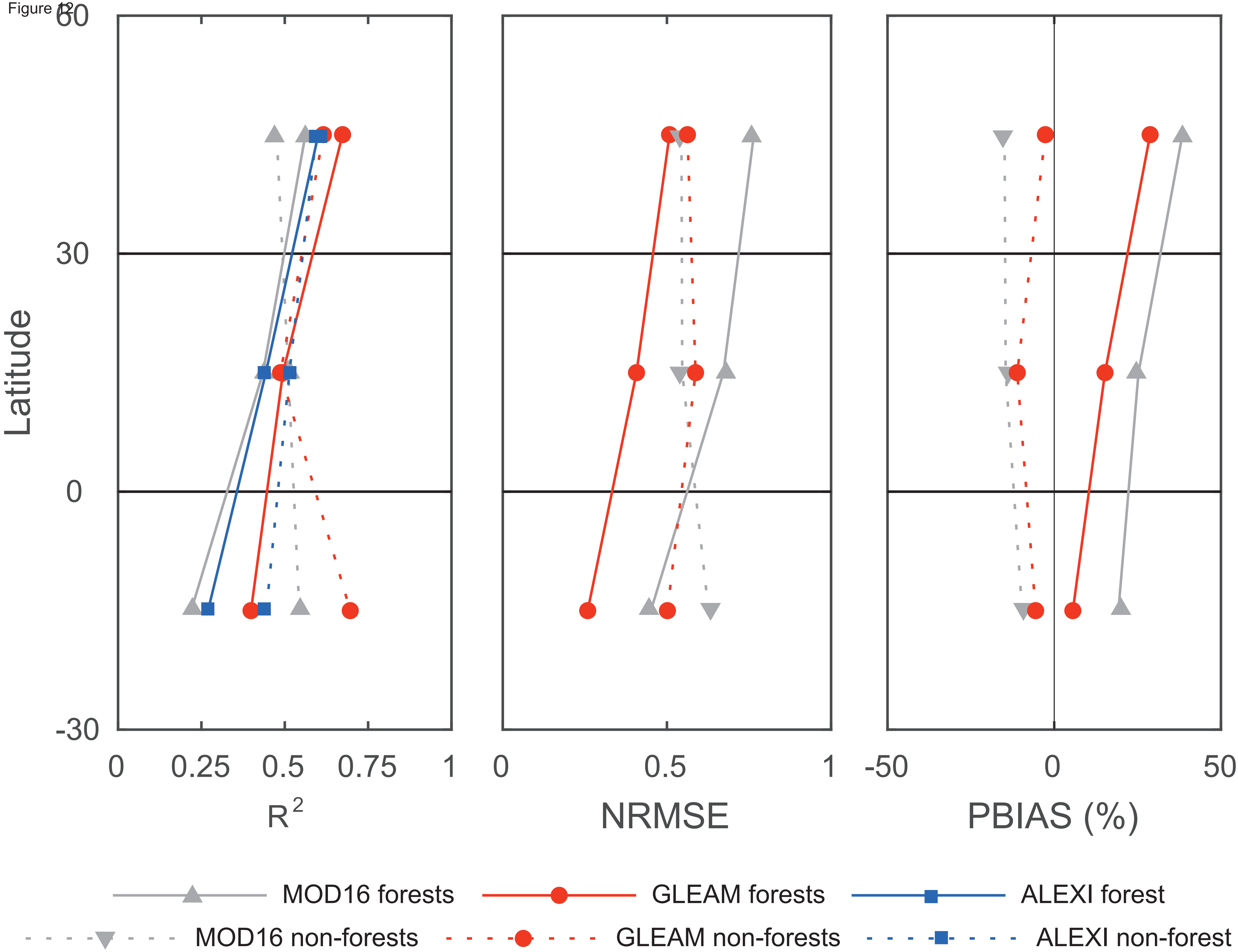
Af, Am - North (3)
Af, Am - South (3)
C - North (4)
C - South (2)
Aw - South (8)
B - South (4)
B - North (4)
Monthly-mean ET (mm day⁻¹)
6
4
2
0
Feb
Apr
Jun
Aug
Oct
Dec
ET orig
ET ebc
GLEAM
MOD16 C5
MOD16 C6
ALEXI

Figure 11



R2
RMSE (mm day−1)
absolute PBIAS (%)
GLEAM
MOD16 C5
MOD16 C6
ALEXI
p = 0.86
p = 0.13
p = 0.33
p = 0.6
p = 0.33
p = 0.48
p = 0.36
p = 0.9
p = 0.39
p = 0.22
p = 0.43
p = 0.52
VMI
VMI
VMI

Figure 12



Latitude
60
30
0
-30
0
0.25
0.5
0.75
1
R^2
0
0.5
1
NRMSE
-50
0
50
PBIAS (%)
MOD16 forests
GLEAM forests
ALEXI forest
MOD16 non-forests
GLEAM non-forests
ALEXI non-forest

Fig. 1. Map showing the geographic location of the 40 eddy covariance sites used in the evaluation, zoomed in for Mexico and northern Australia. The numbers identify the sites in Table 2.

Fig. 2. Scatter plots of daytime sums of sensible heat flux (*H*) and evapotranspiration (ET_{orig}) versus available energy (R_n−G−S; all terms in units of millimetres) for different land cover types for the eddy covariance sites used in the evaluation. Shown are the regression slope (value before the slash), the intercept (value after the slash), the coefficient of determination (R^2), the energy balance ratio (EBR), the linear regression line (solid red line), and the 1:1 line (dashed line).

Fig. 3. Scatter plots of daytime sums of sensible heat flux (*H*) and evapotranspiration (ET_{orig}) versus available energy (R_n−G−S; all terms in units of millimetres) for different climate zones for the eddy covariance sites used in the evaluation. Shown are the regression slope (value before the slash), the intercept (value after the slash), the coefficient of determination (R^2), the energy balance ratio (EBR), the linear regression line (solid red line), and the 1:1 line (dashed line).

Fig. 4. Unadjusted eddy covariance ET observations (ET_{orig}) versus remote sensing-based ET for each land cover type for each of the evaluated products. Shown are the regression slope (value before the slash), the intercept (value after the slash), the coefficient of determination (R^2), the linear regression line (solid red line), and the 1:1 line (dashed line).

Fig. 5. Mean performance statistics (R^2, RMSE, PBIAS) by land cover type for each of the evaluated products for the unadjusted eddy covariance ET observations (ET_{orig}) and those corrected for the lack of energy balance closure (ET_{ebc}).

Fig. 6. Heat maps of the Ideal Point Error (IPE) for each of the evaluated products for each of the comparisons by land cover type and climate zone for the unadjusted eddy covariance ET observations (ET_{orig}) and those corrected for the lack of energy balance closure (ET_{ebc}). The IPE values are shown on the plot. The lower the IPE, the better the relative performance of the product. Blue/red colors indicate best/worst IPE scores. The asterisks in the heatmap for ET_{orig} indicate where the ranking of a product differed from that based on the IPE scores for the common reference dataset (Figure S3; Section 2.2.4).

Fig. 7. Average monthly ET for the four ET products together with the average monthly unadjusted ET observations (ET_{orig}) and those corrected for the lack of energy balance closure (ET_{ebc}) for different land cover types in the northern and southern hemispheres. Curves were calculated using the common reference dataset. Only land cover-hemisphere combinations for which data from at least two sites were available are shown. The number of sites in each land cover-hemisphere combination is given between parentheses. The error band represents the standard deviation of the mean monthly ET_{orig} at the different sites.

Fig. 8. Unadjusted eddy covariance ET observations (ET_{orig}) versus remote sensing-based ET for each climate zone for each of the evaluated products. Shown are the regression slope (value before the slash), the intercept (value after the slash), the coefficient of determination (R^2), the linear regression line (solid red line), and the 1:1 line (dashed line).

Fig. 9. Mean performance statistics (R^2, RMSE, PBIAS) by climate zone for each of the evaluated products for the unadjusted eddy covariance ET observations (ET_{orig}) and those corrected for the lack of energy balance closure (ET_{ebc}).

Fig. 10. Average monthly ET for the four ET products together with the average monthly unadjusted ET observations (ET_{orig}) and those corrected for the lack of energy balance closure (ET_{ebc}) for different climate zones in the northern and southern hemispheres. Curves were calculated using the common reference dataset. Only climate zone-hemisphere combinations for which data from at least two sites were available are shown. The number of sites in each climate zone-hemisphere combination is given between parentheses. The error band represents the standard deviation of the mean monthly ET_{orig} at the different sites.

Fig. 11. Binned scatter plots between the performance metrics (R^2, RMSE, PBIAS) and the vegetation match index for each of the evaluated products. Shown are the regression lines and the *p*-values indicating the statistical significance of the regression slopes.

Fig. 12. Zonal averages (southern and northern low latitudes and northern middle latitudes) of the performance metrics grouped by forest and non-forest vegetation for MOD16 and GLEAM (R^2, NRMSE, PBIAS) and ALEXI (only R^2). Averages were calculated using evaluation results from this study and from the literature. See Section 2.4 for further details.

1 Table 1. General characteristics of the remote sensing-based ET products evaluated in this study.

	GLEAM v3.3a	MOD16 C5	MOD16 C6	ALEXI
Spatial resolution	0.25°	1 km	500 m	0.05°
Temporal resolution	daily	8-day	8-day	daily
Temporal coverage	1980–2018	2000–2014	2000–present	2002–2019
Principle	Priestley-Taylor	Penman-Monteith	Penman-Monteith	Two-source energy balance
Public access	yes	yes	yes	no
References	Miralles et al. (2011); Martens et al. (2017)	Mu et al. (2011, 2013)	Mu et al. (2011); Running et al. (2019)	Anderson et al. (1997, 2007, 2011); Hain and Anderson (2007)

Table 2. Eddy covariance sites used in the evaluation of the remote sensing-based ET products. Shown for each site are: number to locate site on map in Figure 1; site ID or site name used by the flux network; latitude and longitude (decimal degrees); period with data availability; flux network; country; IGBP land cover type; Köppen-Geiger climate class; evaporative fraction (EF); energy balance ratio (EBR); site elevation (m); and reference(s) to article(s) with additional information or to dataset.

	Site ID or site name	Lat.	Lon.	Data period	Network	Country	Land cover	Climate class	EF	EBR	Elev. (m)	Reference
1	PDF	−2.35	114.03	2002–2005	AsiaFlux	Indonesia	EBF	Af	0.77	0.82	30	Hirano et al. (2015)
2	BR-Sa1	−2.86	−54.96	2002–2011	FLUXNET	Brazil	EBF	Am	0.76	0.71	88	Saleska (2002–2011)
3	BR-Sa3	−3.02	−54.97	2000–2004	FLUXNET	Brazil	EBF	Am	0.77	0.82	100	Goulden (2000–2004)
4	CN-Din	23.17	112.54	2003–2005	FLUXNET	China	EBF	Cwa	0.60	0.68	300	Yu et al. (2006)
5	GF-Guy	5.28	−52.92	2004–2014	FLUXNET	French Guiana	EBF	Am	0.75	0.85	48	Bonal et al. (2008)
6	La Orduña	19.47	−96.93	2014–2018	MexFlux	Mexico	EBF	Cfa	0.56	0.82	1210	Holwerda et al. (2016); Holwerda and Meesters (2019)
7	Puerto Morelos	20.85	−86.90	2017–2018	MexFlux	Mexico	EBF	Aw	0.51	0.75	10	Alvarado-Barrientos et al. (2021)
8	El Sargento	29.34	−112.28	2014–2016	MexFlux	Mexico	EBF	BWh	0.59	0.99	0	Delgado-Balbuena et al. (2018)
9	MY-PSO	2.97	102.31	2003–2009	FLUXNET	Malaysia	EBF	Af	0.64	0.98	112	MY-PSO (2003–2009)
10	Cape Tribulation	−16.10	145.45	2012–2018	OzFlux	Australia	EBF	Am	0.72	0.77	40	Liddell (2013)
11	Cow Bay	−16.24	145.43	2011–2019	OzFlux	Australia	EBF	Am	0.72	0.63	86	Liddell (2013)
12	Robson Creek	−17.12	145.63	2014–2019	OzFlux	Australia	EBF	Cwa	0.61	0.71	710	Liddell (2013)
13	CLM	24.59	121.42	2007–2009	AsiaFlux	Taiwan	ENF	Cfb	0.50	0.86	1638	Chu et al. (2014)
14	AR-Vir	−28.24	−56.19	2009–2012	FLUXNET	Argentina	ENF	Cfa	0.79	0.72	127	Posse et al. (2016)
15	CN-Qia	26.74	115.06	2003–2004	FLUXNET	China	ENF	Cfa	0.70	0.76	100	Yu et al. (2006)
16	Atopixco	20.61	−98.59	2017–2018	MexFlux	Mexico	ENF	Cwb	0.43	0.99	2064	Hidalgo-Sánchez et al. (2021)
17	Álamos	27.00	−108.79	2015–2017	MexFlux	Mexico	DBF	BSh	0.57	0.71	368	Rojas-Robles et al. (2020)
18	Chamela	19.51	−105.04	2007–2010	MexFlux	Mexico	DBF	Aw	0.30	0.82	73	González del Castillo et al. (2018)
19	Sierra Los Locos	29.96	−110.46	2010–2014	MexFlux	Mexico	DBF	BSh	0.42	0.73	1314	Pérez-Ruiz et al. (2021)

20	El Palmar	21.02	−90.06	2016–2018	MexFlux	Mexico	DBF	Aw	0.40	1.03	8	Figueroa-Espinoza et al. (2021); Uuh-Sonda et al. (2022)
21	Rayón	29.74	−110.53	2008–2015	MexFlux	Mexico	DBF	BSh	0.23	0.81	632	Verduzco et al. (2018); Pérez-Ruiz et al. (2021)
22	Tesopaco	27.83	−109.28	2004–2008	MexFlux	Mexico	DBF	BSh	0.31	0.76	426	Verduzco et al. (2015)
23	PA-SPn	9.32	−79.63	2007–2009	FLUXNET	Panama	DBF	Am	0.66	0.91	78	Wolf et al. (2011)
24	ZM-Mon	−15.44	23.25	2007–2009	FLUXNET	Zambia	DBF	Aw	0.45	0.77	1053	Merbold et al. (2009)
25	AU-ASM	−22.28	133.25	2010–2014	FLUXNET	Australia	SAV	BWh	0.16	0.83	600	Cleverly et al. (2013)
26	AU-Ade	−13.08	131.12	2007–2009	FLUXNET	Australia	SAV	Aw	0.58	0.96	100	Beringer et al. (2011)
27	AU-DaS	−14.16	131.39	2008–2014	FLUXNET	Australia	SAV	Aw	0.51	0.90	110	Hutley et al. (2011)
28	AU-Dry	−15.26	132.37	2009–2014	FLUXNET	Australia	SAV	Aw	0.46	0.80	175	Cernusak et al. (2011)
29	AU-How	−12.49	131.15	2003–2014	FLUXNET	Australia	WSA	Aw	0.63	0.89	64	Beringer et al. (2007)
30	AU-RDF	−14.56	132.48	2011–2013	FLUXNET	Australia	SAV	Aw	0.40	0.87	171	Bristow et al. (2016)
31	SD-Dem	13.28	30.48	2007–2009	FLUXNET	Sudan	SAV	BWh	0.40	0.83	500	Ardö et al. (2008)
32	AU-DaP	−14.06	131.32	2007–2013	FLUXNET	Australia	GRA	Aw	0.45	0.74	67	Hutley et al. (2011)
33	AU-Emr	−23.86	148.47	2011–2013	FLUXNET	Australia	GRA	BSh	0.32	0.67	170	Schroder et al. (2014)
34	AU-Fog	−12.55	131.31	2006–2008	FLUXNET	Australia	GRA	Aw	0.72	0.80	4	Beringer et al. (2013)
35	AU-Stp	−17.15	133.35	2008–2014	FLUXNET	Australia	GRA	BSh	0.33	0.82	225	Beringer et al. (2011)
36	AU-TTE	−22.29	133.64	2012–2014	FLUXNET	Australia	GRA	BWh	0.12	1.03	553	Cleverly et al. (2016)
37	CG-Tch	−4.29	11.66	2006–2009	FLUXNET	Congo	GRA	Aw	0.48	0.98	82	Merbold et al. (2009)
38	Ojuelos	21.78	−101.61	2011–2016	MexFlux	Mexico	GRA	BSk	0.41	0.83	2228	Delgado-Balbuena et al. (2019)
39	PA-SPs	9.31	−79.63	2007–2009	FLUXNET	Panama	GRA	Am	0.67	0.96	68	Wolf et al. (2011)
40	US-KS2	28.61	−80.67	2003–2006	FLUXNET	United States	CSH	Cfa	0.59	0.81	3	Drake and Hinkle (2003–2006)

Table 3. Number of eddy covariance sites and site years (between parentheses) available in the complete dataset for each product by land cover type and climate zone. Note that the number of site years corresponds to the length of the flux tower records. Actual data availability was lower due to, for example, missing or erroneous data.

		GLEAM	MOD16 C5	MOD16 C6	ALEXI
Land cover type	EBF	9 (49)	11 (47)	12 (67)	8 (47)
	ENF	4 (7)	3 (6)	4 (7)	4 (7)
	DBF	6 (23)	5 (18)	8 (27)	8 (27)
	SAV	7 (33)	7 (33)	7 (33)	7 (33)
	GRA	7 (28)	8 (29)	8 (31)	8 (31)
Climate zone	Af, Am	8 (42)	9 (46)	9 (55)	7 (39)
	Aw	8 (37)	10 (43)	12 (46)	11 (45)
	C	7 (19)	7 (15)	8 (24)	8 (24)
	B	10 (42)	9 (33)	11 (45)	10 (42)

Table 4. Number of evaluation results (N_{ER}) from this study and from the literature, broken down into results for R^2, NRMSE and PBIAS, and grouped by latitudinal zone, ET product and vegetation category (see Section 2.4 for further explanation). NA is not Not Available.

		MOD16		GLEAM		ALEXI	
		Forest	Non-forest	Forest	Non-forest	Forest	Non-forest
30° N–60° N	R^2	19	30	2	9	27	26
	NRMSE	11	23	2	9	NA	NA
	PBIAS	11	23	2	9	NA	NA
0°–30° N	R^2	15	9	16	4	14	5
	NRMSE	13	9	16	4	14	4
	PBIAS	13	9	16	4	14	4
0°–30° S	R^2	9	16	8	11	6	19
	NRMSE	8	16	8	11	6	12
	PBIAS	8	16	8	11	6	12