Hydrol. Earth Syst. Sci., 20, 803–822, 2016
www.hydrol-earth-syst-sci.net/20/803/2016/
doi:10.5194/hess-20-803-2016


Hydrology and Earth System Sciences
Open Access
EGU

# The WACMOS-ET project – Part 1: Tower-scale evaluation of four remote-sensing-based evapotranspiration algorithms

**D. Michel[1], C. Jiménez[2,3], D. G. Miralles[4,5], M. Jung[6], M. Hirschi[1], A. Ershadi[7], B. Martens[5], M. F. McCabe[7], J. B. Fisher[8], Q. Mu[9], S. I. Seneviratne[1], E. F. Wood[10], and D. Fernández-Prieto[11]**

[1]Institute for Atmospheric and Climate Science, ETH Zürich, Zürich, Switzerland
[2]Estellus, Paris, France
[3]LERMA, Paris Observatory, Paris, France
[4]Department of Earth Sciences, VU University Amsterdam, Amsterdam, the Netherlands
[5]Laboratory of Hydrology and Water Management, Ghent University, Ghent, Belgium
[6]Max Planck Institute for Biogeochemistry, Jena, Germany
[7]Division of Biological and Environmental Sciences and Engineering, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia
[8]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, USA
[9]Department of Ecosystem and Conservation Sciences, University of Montana, Missoula, Montana, USA
[10]Department of Civil and Environmental Engineering, Princeton University, Princeton, New Jersey, USA
[11]ESRIN, European Space Agency, Frascati, Italy

*Correspondence to:* D. Michel (dominik.michel@env.ethz.ch)

Received: 2 October 2015 – Published in Hydrol. Earth Syst. Sci. Discuss.: 20 October 2015
Accepted: 2 February 2016 – Published: 23 February 2016

**Abstract.** The WAter Cycle Multi-mission Observation Strategy – EvapoTranspiration (WACMOS-ET) project has compiled a forcing data set covering the period 2005–2007 that aims to maximize the exploitation of European Earth Observations data sets for evapotranspiration (ET) estimation. The data set was used to run four established ET algorithms: the Priestley–Taylor Jet Propulsion Laboratory model (PT-JPL), the Penman–Monteith algorithm from the MODerate resolution Imaging Spectroradiometer (MODIS) evaporation product (PM-MOD), the Surface Energy Balance System (SEBS) and the Global Land Evaporation Amsterdam Model (GLEAM). In addition, in situ meteorological data from 24 FLUXNET towers were used to force the models, with results from both forcing sets compared to tower-based flux observations. Model performance was assessed on several timescales using both sub-daily and daily forcings. The PT-JPL model and GLEAM provide the best performance for both satellite- and tower-based forcing as well as for the considered temporal resolutions. Simulations using the PM-MOD were mostly underestimated, while the SEBS performance was characterized by a systematic overestimation. In general, all four algorithms produce the best results in wet and moderately wet climate regimes. In dry regimes, the correlation and the absolute agreement with the reference tower ET observations were consistently lower. While ET derived with in situ forcing data agrees best with the tower measurements ($R^2 = 0.67$), the agreement of the satellite-based ET estimates is only marginally lower ($R^2 = 0.58$). Results also show similar model performance at daily and sub-daily (3-hourly) resolutions. Overall, our validation experiments against in situ measurements indicate that there is no single best-performing algorithm across all biome and forcing types. An extension of the evaluation to a larger selection of 85 towers (model inputs resampled to a common grid to facilitate global estimates) confirmed the original findings.

## 1 Introduction

Research on climate variability and the development of predictive capabilities relies largely on globally available reference data time series of the various components of the energy

and water cycles. Turbulent fluxes of sensible and latent heat determine the development of the planetary boundary layer and thus govern the interactions between the Earth surface and the atmosphere. Evapotranspiration (ET) represents the time-integrated flux of latent heat and is an essential component of the energy and water cycle, playing a key role in meteorology and climate as well as agriculture (see, e.g., Ershadi et al., 2014).

Historically, there has been a lack of reliable estimates of turbulent fluxes, since the partitioning of the available energy at the Earth's surface into these fluxes is complex and characterized by large spatial and temporal variability. Also, these components of the energy balance cannot be monitored directly on a global scale by remote sensing techniques. Thus, efforts to produce satellite-based estimates typically involve combining multi-sensor data sets with predictive formulations of varying complexity, ranging from relatively simple empirical formulations to more complex modeling approaches (see, e.g., Courault et al. (2005), Kalma et al. (2008) and Wang and Dickinson (2012) for comprehensive reviews). In recent years, such efforts have generated global ET products (Mu et al., 2007; Fisher et al., 2008; Jung et al., 2010; Zhang et al., 2010; Vinukollu et al., 2011; Miralles et al., 2011a) that have typically been evaluated by comparing them individually to in situ data or by inter-comparing them with other existing global heat flux estimates. For example, within the Evaluation and inter-comparison of existing land evapotranspiration products (Landflux-Eval initiative) initiative of the Global Energy and Water cycle Exchanges (GEWEX) Data and Assessments Panels (GDAP), Jiménez et al. (2011) investigated 3 years (1993–1995) of global sensible and latent fluxes from a selection of 12 products, including satellite-based estimates, atmospheric reanalyses and offline land surface model simulations, while Mueller et al. (2011) examined a total of 30 observation-based ET estimates from similar sources over the longer period of 1989–1995, while also providing a comparison with global climate model simulations contributing to the Intergovernmental Panel on Climate Change (IPCC) fourth assessment report. More recently, Mueller et al. (2013) extended the previous LandFlux-EVAL evaluations and presented two monthly global ET synthesis products, merged from individual data sets spanning the periods 1989–1995 and 1989–2005. Based on the Mueller et al. (2013) methodology, Mao et al. (2015) synthesized a global ET time series for the period 1982–2010 based on a set of diagnostic ET products (including data sets produced with PT-JPL (Priestley–Taylor Jet Propulsion Laboratory model) and GLEAM (Global Land Evaporation Amsterdam Model)) to investigate the role of anthropogenic and climatic controls on ET trends. For the period 1982–2013 Zhang et al. (2015) produced a satellite-based global ET data set using remote sensing data and daily surface meteorology records for the investigation of multidecadal changes in ET, which was validated against precipitation and discharge records.

The GEWEX-LandFlux initiative is currently working towards producing an observation-based data set of heat fluxes that can be used together with related GDAP products to enable a joint analysis of the water and energy cycles (Jimenez et al., 2012). To contribute towards that goal, the European Space Agency (ESA) has conducted the Water Cycle Multi-mission Observation Strategy EvapoTranspiration project (WACMOS-ET), aimed at the identification of appropriate algorithms to develop regional and global ET products. WACMOS-ET efforts, which aimed at maximizing the use of European Earth Observation assets, have also included the compilation of a multi-sensor data set to run the ET methodologies for a 3-year period (2005-2007).

In WACMOS-ET, the methodologies by Su (2002) (Surface Energy Balance System, hereafter referred to as SEBS), Mu et al. (2011) (Penman–Monteith algorithm from the MODerate resolution Imaging Spectroradiometer (MODIS) evaporation product, PM-MOD), Fisher et al. (2008) (Priestley–Taylor Jet Propulsion Laboratory model, PT-JPL) and Miralles et al. (2011a, b) (Global Land Evaporation Amsterdam Model, GLEAM) were selected to produce ET estimates on different temporal and spatial scales. The same algorithms have also been examined at a selection of different tower sites in a recent paper by McCabe et al. (2015) in preparation for the GEWEX-LandFlux product. In McCabe et al. (2015) the algorithms are run at 3-hourly time steps with both point-scale inputs (from tower meteorological observations) and gridded inputs (from the GEWEX-LandFlux global forcing data set) over a longer time period. Here, the ET algorithms are run with the WACMOS-ET forcings (see Sect. 2.2) and the analyses of model performance are extended to evaluate different timescales (3-hourly and daily) and time integrations (nighttime, daytime and full day). In a companion paper, Miralles et al. (2016) present the second part of the WACMOS-ET study, in which PT-JPL, GLEAM and PM-MOD are evaluated on the global scale.

The paper is structured as follows. First, the WACMOS-ET input data set is described in detail, together with the tower flux data used for driving and evaluating the ET models. This is followed by an evaluation of ET model performance on the tower scale using the tower eddy-covariance (EC) fluxes as the reference data set. The model evaluation is first performed over a selection of 24 stations covering nine biomes in three continents (Europe, North America and Australia), in which models are run based on in situ and remote sensing forcing. Then the validation is extended to embrace a larger sample of 85 towers, in which models are driven only by satellite data resampled to a common grid. Finally, the main findings of our model evaluation on the pixel scale are summarized.

## 2 Methods and data

Here the ET methodologies comprising the WACMOS-ET effort together with the input data sets that have been compiled to run the models and evaluate the ET estimates are presented. A summary of the data sets and the model-specific forcing requirements is provided in Table 1.

### 2.1 ET models

The four algorithms selected here estimate the fraction of the available energy at the surface used by the soil and canopy evaporation processes. Therefore, the available energy (i.e., the difference between the surface net radiation and the ground heat flux) is a key input for all algorithms. However, this evaporative fraction is parameterized differently by each model. SEBS is an energy balance model (Su, 2002) based on a detailed parameterization of the sensible heat flux at the surface, where ET is estimated as the residual of the surface energy balance once the sensible heat flux is calculated. Therefore, key inputs are the surface temperature and the temperature gradient between the surface and the air, and a key component of the model is the aerodynamic resistance to sensible heat transfer. PM-MOD (Mu et al., 2011) derives ET directly based on the Penman–Monteith equation (Monteith, 1965), which relates the latent heat flux to the vapor pressure deficit between the surface and the overlying air and requires a resistance parameter to characterize the canopy transpiration. PT-JPL (Fisher et al., 2008) and GLEAM (Miralles et al., 2011a) are based on first determining the potential evaporation by applying the Priestley and Taylor equation (Priestley and Taylor, 1972), followed by reducing the potential evaporation to actual evaporation with a number of evaporative stress factors. The derivation of these stress factors is different between both models. PT-JPL requires the vapor pressure deficit, relative humidity and visible-infrared-related vegetation indexes, while GLEAM combines microwave data of vegetation optical depth and soil moisture. A more detailed description of the input requirements for each model can be found in Table 1, while a more comprehensive description of each individual model is given in the following sections.

#### 2.1.1 SEBS

SEBS (Su, 2002) is a one-source energy balance algorithm that is arguably one of the most widely used energy balance approaches to derive turbulent fluxes. The SEBS model calculates the sensible heat flux ($H$) based on the Monin and Obukhov theory (Monin and Obukhov, 1954):

$$\theta_0 - \theta_a = \frac{H}{k u_* \rho c_p}\left[\ln\left(\frac{z-d_0}{z_{0h}}\right) - \Gamma_h\left(\frac{z-d_0}{L}\right) + \Gamma_h\left(\frac{z_{0h}}{L}\right)\right], \quad (1)$$

$$u = \frac{u_*}{k}\left[\ln\left(\frac{z-d_0}{z_{0m}}\right) - \Gamma_m\left(\frac{z-d_0}{L}\right) + \Gamma_m\left(\frac{z_{0h}}{L}\right)\right], \quad (2)$$

$$L = -\frac{\rho c_p u_*^3 \theta_v}{k g H}, \quad (3)$$

where $u$ is the wind speed, $u_*$ is the friction velocity, $k$ is the von Kármán constant, $z$ is the height above the surface, $d_0$ is the zero plane displacement height, $z_{0m}$ and $z_{0h}$ are the roughness heights for momentum and heat transfer, and $\Gamma_m$ and $\Gamma_h$ are the stability correction functions for momentum and sensible heat transfer, respectively. $L$ refers to the Obukhov length, $\rho$ is the air density, $\theta_0$ is potential land surface temperature, $\theta_a$ is the potential air temperature at height $z$, $g$ is the gravity acceleration and $\theta_v$ is the potential virtual air temperature at level $z$. When the suitable data are available, the only unknowns are $H$, $u_*$ and $L$. This allows the calculation of $H$ and the further estimation of ET based on closing the energy balance at the surface, i.e., ET is estimated as the difference between net radiation ($R_n$) and the sum of the calculated $H$ and ground flux ($G$).

Additionally, in order to constrain $H$ estimates, two limiting cases are considered that set upper and lower bounds for the evaporative fraction. Under very dry conditions, ET becomes 0 and $H$ is at its maximum, set by $R_n - G$. Under wet conditions, ET occurs at potential rates and therefore $H$ is at its minimum. In this wet case, $H$ is calculated via reverse application of the Penman–Monteith equation (see Sect. 2.1.2) assuming that the surface resistance is zero.

SEBS has been extensively validated against tower measurements and has proved to estimate realistic evaporation rates on a variety of scales, ranging from local to regional (Jia et al., 2003; Su et al., 2005; McCabe and Wood, 2006). As an example, Chen et al. (2015) recently reported an average correlation of $\sim 0.8$ and root mean square difference (RMSD) of $0.7\,\mathrm{mm\,day^{-1}}$ against eddy-covariance measurements in a validation of monthly SEBS ET aggregates over China.

As a one-source energy model, SEBS does not separate the contributing components of ET (i.e., transpiration, interception loss, bare-soil evaporation), unlike the other models studied in WACMOS-ET, which provide independent estimates of these vapor sources. Although not examined here, we note that two-source energy balance models can also treat the soil and vegetation components separately (e.g., Kustas and Norman, 2000; Anderson et al., 2007) but have had limited application on the global scale.

#### 2.1.2 PM-MOD

PM-MOD (Mu et al., 2011) is based on the Penman–Monteith equation. It estimates ET as the sum of interception loss ($I$), transpiration ($ET_t$) and evaporation from the soil ($ET_s$). The interception loss is modeled as

**Table 1.** Table summarizing the model inputs. Listed are the main inputs, product selected, original temporal and spatial resolutions, and the satellite sensors used to derive the product.

| Variable | Models | Product | Resolution | Sensors |
|---|---|---|---|---|
| Surface radiation | All | SRB | 3-hourly, 100 km | Several VIS-IR sensors |
| Surface temperature | SEBS | IPMA | | |
| polar | | | twice daily, 1 km | AATSR |
| geostationarity | | | hourly, 5 km | MSG-2, MT-SAT, GOES-12 |
| Surface meteorology | | ERA-Interim | 3-hourly, 75 km | Assimilation of satellite and other meteo observations |
| temperature | All | | | |
| humidity | SEBS, PM-MOD, PT-JPL | | | |
| wind | SEBS | | | |
| $f$APAR, LAI | SEBS, PM-MOD, PT-JPL | From ESA GlobAlbedo | 8 days, 1 km | VEGETATION, MERIS, MODIS |
| Soil moisture | GLEAM | ESA-CCI | daily, 25 km | SSM/I, TMI, AMSR-E, ASCAT |
| Precipitation | GLEAM | CMORPH | 30 min, 15 km | AMSU-B, AMSR-E, TMI |
| Snow water | GLEAM | ESA | daily, 1 km | AMSR-E |

$f$APAR: fraction of absorbed photosynthetically active radiation; VIS-IR:visible and infrared; IPMA: Portuguese Sea and Atmosphere Institute; AATSR: Advanced Along-Track Scanning Radiometer; MSG-2: Meteosat Second Generation 2; MT-SAT: Multi-functional Transport Satellites; GOES-12: Geostationary Operational Environmental Satellite 12; VEGETATION: SPOT (Satellite Pour l'Observation de la Terre) vegetation; MERIS: Medium Resolution Imaging Spectrometer; SSM/I: Special Sensor Microwave Imager; TMI: TRMM (Tropical Rainfall Measuring Mission) Microwave Imager; AMSR-E: Advanced Microwave Scanning Radiometer – Earth Observing System; ASCAT: Advanced Scatterometer; AMSU-B: Advanced Microwave Sounding Unit-B.

$$I = f_{\mathrm{wet}} f_{\mathrm{c}} \frac{\Delta\,(R_{\mathrm{n}} - G) + \rho c_p \mathrm{VPD}/r_{\mathrm{a}}^{\mathrm{wc}}}{\lambda\left(\Delta + \gamma \frac{r_{\mathrm{s}}^{\mathrm{wc}}}{r_{\mathrm{a}}^{\mathrm{wc}}}\right)}, \tag{4}$$

where $\Delta$ is the slope of the curve relating saturated water vapor pressure to temperature, VPD is the vapor pressure deficit, $\gamma$ is the psychrometric constant, $f_{\mathrm{c}}$ is the canopy fraction, $f_{\mathrm{wet}}$ is the wet cover fraction based on the derivation by Fisher et al. (2008) (see Eq. (9) in Sect. 2.1.3), and $r_{\mathrm{a}}^{\mathrm{wc}}$ and $r_{\mathrm{s}}^{\mathrm{wc}}$ are aerodynamic and surface resistances against evaporation of intercepted water (calculated as a function of air temperature and leaf area index, LAI).

Canopy transpiration is estimated as

$$\mathrm{ET_t} = (1 - f_{\mathrm{wet}})\, f_{\mathrm{c}} \frac{\Delta\,(R_{\mathrm{n}} - G) + \rho c_p \mathrm{VPD}/r_{\mathrm{a}}^{\mathrm{t}}}{\lambda\left(\Delta + \gamma \frac{r_{\mathrm{s}}^{\mathrm{t}}}{r_{\mathrm{a}}^{\mathrm{t}}}\right)}, \tag{5}$$

where $r_{\mathrm{a}}^{\mathrm{t}}$ and $r_{\mathrm{s}}^{\mathrm{t}}$ are the aerodynamic and surface resistances against transpiration. $r_{\mathrm{a}}^{\mathrm{t}}$ is determined in a similar way to $r_{\mathrm{a}}^{\mathrm{wc}}$, and $r_{\mathrm{s}}^{\mathrm{t}}$ is a function of stomatal conductance, biome-constant values of cuticular conductance and canopy boundary-layer conductance. The values of stomatal conductance are a function of air temperature, VPD and LAI.

Evaporation from the soil surface is the sum of evaporation from wet soil and evaporation from saturated soil, which are both calculated separately based on Eq. (7) with specific values of aerodynamic and surface resistances for bare soils and a soil moisture constraint ($f_{\mathrm{sm}}$) depending on relative humidity (taken from Fisher et al. (2008), see Sect. 2.1.3).

The Mu et al. (2011) daily ET estimates have been previously validated against EC measurements from 46 FLUXNET towers in North America, reporting for the daily estimates an average RMSD of $\sim 0.9\,\mathrm{mm\,day^{-1}}$ and a $\sim 0.6$ average correlation coefficient.

### 2.1.3 PT-JPL

PT-JPL (Fisher et al., 2008) is based upon the Priestley and Taylor equation. As in PM-MOD, ET is estimated as the sum of interception loss $I$, transpiration $\mathrm{ET_t}$ and evaporation from the soil $\mathrm{ET_s}$. The driving equations in the model are

$$\mathrm{ET_t} = (1 - f_{\mathrm{wet}})\, f_{\mathrm{g}} f_T f_{\mathrm{M}} \alpha \frac{\Delta}{\Delta + \gamma} R_{\mathrm{n}}^{\mathrm{c}}, \tag{6}$$

$$\mathrm{ET_s} = f_{\mathrm{wet}} + f_{\mathrm{sm}}\,(1 - f_{\mathrm{wet}})\, \alpha \frac{\Delta}{\Delta + \gamma} \left(R_{\mathrm{n}}^{\mathrm{s}} - G\right), \tag{7}$$

$$I = f_{\mathrm{wet}} \alpha \frac{\Delta}{\Delta + \gamma} R_{\mathrm{n}}^{\mathrm{c}}, \tag{8}$$

where $\alpha$ is known as the Priestley and Taylor (PT) coefficient and is considered here as a constant value (1.26) (Priestley and Taylor, 1972) that aims to summarize the atmospheric term in the Penman–Monteith equation (Eq. 5), $\lambda$ is the latent heat of vaporization and $f_{\mathrm{wet}}$, $f_{\mathrm{g}}$, $f_{\mathrm{M}}$, $f_{\mathrm{sm}}$ and $f_T$ are ecophysiological constraint functions with values between 0 and 1 referred to as $f$ functions. The $f$ functions are given by

$$f_{\mathrm{wet}} = \mathrm{RH}^4, \quad (9)$$

$$f_{\mathrm{g}} = f\mathrm{APAR}/f\mathrm{IPAR}, \quad (10)$$

$$f_{\mathrm{M}} = f\mathrm{APAR}/f\mathrm{APAR}_{\mathrm{max}}, \quad (11)$$

$$f_{\mathrm{sm}} = \mathrm{RH}^{\mathrm{VPD}}, \quad (12)$$

$$f_T = e^{-\left(\frac{T_{\mathrm{s}}-T_{\mathrm{opt}}}{T_{\mathrm{opt}}}\right)^2}, \quad (13)$$

where $f_{\mathrm{wet}}$ is the relative surface wetness, $f_{\mathrm{g}}$ is green canopy fraction, $f\mathrm{APAR}$ ($f\mathrm{IPAR}$) is the fraction of absorbed (intercepted) photosynthetically active radiation, $f_{\mathrm{M}}$ is a plant moisture constraint, $f\mathrm{APAR}_{\mathrm{max}}$ is the maximum of $f\mathrm{APAR}$, $f_{\mathrm{sm}}$ is a soil moisture constraint, $f_T$ is a plant temperature constraint and $T_{\mathrm{opt}}$ is the optimum plant growth temperature, estimated as the air temperature at the time of peak canopy activity when the highest $f\mathrm{APAR}$ and minimum VPD occur. Note that as the input data set does not include $f\mathrm{IPAR}$; $f\mathrm{IPAR}$ is derived from the rescaled project LAI by inverting the model original relationships between LAI and $f\mathrm{IPAR}$.

Using this methodology, monthly estimates of ET were tested against EC measurements from 16 FLUXNET towers worldwide (Baldocchi et al., 2001) with a reported average RMSD of $\sim 0.4\,\mathrm{mm\,day^{-1}}$ and a $\sim 0.9$ average correlation coefficient (Fisher et al., 2008). Note that unlike the above statistics reported for SEBS and PM-MOD, these numbers come from the model run with the tower meteorology, instead of global forcings.

#### 2.1.4 GLEAM

GLEAM (Miralles et al., 2011a, b) calculates ET via the PT equation, a soil moisture stress computation and a Gash analytical model of rainfall interception loss (Gash, 1979). In the absence of snow, evaporation from land is calculated as

$$\mathrm{ET} = \mathrm{ET_{tc}} + \mathrm{ET_{sc}} + \mathrm{ET_s} + \beta I, \quad (14)$$

where $\mathrm{ET_{tc}}$ is transpiration from tall canopy, $\mathrm{ET_{sc}}$ is transpiration from short vegetation, $\mathrm{ET_s}$ is soil evaporation and $I$ is tall canopy interception loss. $\beta$ is a constant used to account for the times at which vegetation is wet; thus, transpiration occurs at lower rates ($\beta = 0.93$) (Gash and Stewart, 1977).

The first three terms in Eq. (14) are derived using the Priestley and Taylor equation, so ET becomes

$$\mathrm{ET} = \frac{\Delta\left[f_{\mathrm{tc}} S_{\mathrm{tc}} \alpha_{\mathrm{tc}} \left(R_{\mathrm{n}}^{\mathrm{tc}} - G_{\mathrm{tc}}\right) + f_{\mathrm{sc}} S_{\mathrm{sc}} \alpha_{\mathrm{sc}} \left(R_{\mathrm{n}}^{\mathrm{sc}} - G_{\mathrm{sc}}\right) + f_{\mathrm{s}} S_{\mathrm{s}} \alpha_{\mathrm{s}} \left(R_{\mathrm{n}}^{\mathrm{s}} - G_{\mathrm{s}}\right)\right]}{\lambda(\Delta+\gamma)} + \beta I., \quad (15)$$

where the subscripts "tc", "sc" and "s" correspond to tall vegetation, short vegetation and bare soil (respectively) and the fraction of each of these three cover types per pixel is represented by $f$. Different cover types have different values of $\alpha$ and parameterizations of $G$; additionally, $R_{\mathrm{n}}$ is distributed within the cover fractions using average albedo ratios from the literature. $S$ represents the evaporative stress due to soil moisture deficit and vegetation phenology. Soil moisture deficit is estimated using a multilayer running water balance to describe the infiltration of observed precipitation through the vertical soil profile. Microwave soil moisture observations are assimilated into the soil profile (Martens et al., 2016). In vegetated land covers, phenology effects on ET are based on microwave observations of vegetation optical depth, used as a proxy of vegetation water content.

$I$ is independently derived using a Gash analytical model (Gash, 1979), in which a running water balance for canopies and trunks is driven by observations of precipitation. The derivation of the parameters and the global implementation and validation of this $I$ model are described in Miralles et al. (2010). For regions covered by ice and snow, sublimation is calculated based on a PT equation parameterized for ice and supercooled waters (Murphy and Koop, 2005).

The ET estimates from GLEAM have been validated against eddy covariance towers worldwide; reported average correlations are 0.83 and 0.90 for daily and monthly estimates, respectively, with an average RMSD of $\sim 0.3\,\mathrm{mm\,day^{-1}}$, based on a sample of 43 towers (Miralles et al., 2011a), and correlations of 0.71–0.75 and 0.81–0.86 for daily and monthly estimates, respectively, based on a sample of 163 towers and different satellite products as forcing (Miralles et al., 2016).

### 2.2 Model inputs

#### 2.2.1 Surface meteorology

The European Centre for Medium-range Weather Forecasts (ECMWF) Re-Analysis-Interim (ERA-Interim) (Dee et al., 2011) was selected to provide the near-surface meteorology every 3 h at a spatial resolution of $\sim 75$ km. The use of reanalysis data is necessary as satellite observations are generally unable to retrieve the surface variables needed, such as temperature, humidity and wind speed, with sufficient accuracy or at a suitable sub-daily temporal resolution. Although products of near-surface air temperature and humidity derived from satellite sounders exist (Ferguson and Wood, 2010), atmospheric reanalyses have the advantage of providing regular sub-daily estimates for all weather conditions. ERA-Interim is also used in the derivation of the land surface temperature product (see Sect. 2.2.2), to ensure inter-product consistency between air and surface temperatures. In terms of accuracy, ERA-Interim data have been evaluated through comparison with other reanalyses and weather stations over specific areas, showing a good general performance (e.g., Mooney et al., 2011; Szczypta et al., 2011).

#### 2.2.2 Land surface temperature

Land surface temperature (LST) estimates have been internally generated by the project from level 1 radiances from the Advanced Along-Track Scanning Radiometer (AATSR)

onboard ESA's Environmental Satellite (EnviSat) polar-orbiting satellite, from Multi-functional Transport Satellites (MT-SAT) 2 (over Australia), Meteosat Second Generation (MSG) 2 and Geostationary Operational Environmental Satellite (GOES) 12. The data sets are provided over a sinusoidal grid with 1 km resolution for AATSR at the two satellite overpasses per day ($\sim$ 10:00 LT descending node) and 5 km for the remaining sensors (1-hourly estimates for MSG and MT-SAT, 3-hourly for GOES). Ancillary atmospheric information for the inversion of the L1 radiances comes from ERA-Interim. Estimates of surface emissivity are taken from the Global Infrared Land Surface Emissivity UW-Madison Baseline Fit Emissivity Database developed by Seemann et al. (2008).

#### 2.2.3 Surface radiation

The National Aeronautics and Space Administration (NASA)–GEWEX Surface Radiation Budget (SRB) satellite product version 3.1 (Stackhouse et al., 2004) is used to provide the surface net radiation input to the ET models. The SRB product is used by a large number of global ET algorithms to characterize the radiation at the surface, given its relatively long data record and sub-daily temporal resolution. SRB data sets include global 3-hourly averages of surface and top-of-atmosphere longwave and shortwave radiative parameters on a $\sim$ 100 km grid.

#### 2.2.4 LAI and $f$APAR

To characterize the vegetation state using visible and near-infrared wavebands, estimates of LAI and $f$APAR have been derived by applying the Joint Research Centre (JRC) two-stream inversion package (hereafter TIP) (Pinty et al., 2007, 2011a, b) to the ESA GlobAlbedo bihemispherical reflectances. Here, LAI is defined as the one-sided leaf area per unit ground area and $f$APAR as the fraction of absorbed photosynthetically active radiation in the 400–700 nm region.

The application of the TIP LAI and $f$APAR with our ET models required some LAI and $f$APAR calibration. The TIP LAI is a one-dimensional (1-D) equivalent LAI for solving the radiative transfer in a three-dimensional (3-D) medium, and it is consistent with the fluxes from which it is derived. It is not consistent with LAI derived using a 3-D radiative scheme that allows some form of horizontal canopy clumping (e.g., the MODIS MOD15A2 LAI product). In practical terms, this means that if an ET model was constructed to use a MODIS-like LAI and $f$APAR, a straight use of the project LAI and $f$APAR will result in the ET model producing lower values than expected for those biomes where horizontal clumping is significant (e.g., for forests). While the ET dynamics may not be highly affected (there is a high degree of correlation between different LAI and $f$APAR estimates), the absolute values would be. As the SEBS, PM-MOD and PT-JPL models have typically been used with the MODIS vegetation product, a rescaling of our TIP-derived LAI and $f$APAR products against the MODIS product has been undertaken. For running the models on the tower scale, a local rescaling is conducted by a linear regression between the MOD15A2 and the TIP values co-registered at each tower. For global model simulations, individual rescaling per biome or climate classification is conducted. For PT-JPL, given the model internal relationships between these variables and the vegetation indexes used as model inputs (see Table 1 in Fisher et al., 2008), it can be discussed whether the original TIP LAI and $f$APAR or the rescaled LAI and $f$APAR are the most appropriate to be used as model inputs. For simplicity we will apply also the rescaled LAI and $f$APAR, but this choice will be further evaluated in future applications of the model with the TIP LAI and $f$APAR inputs.

Figure 1 illustrates an example of both products at two towers. The station Quebec – Eastern Boreal, Mature Black Spruce (CA-Qfo) is located in an evergreen needleleaf forest and shows that the MOD15A2 and WACMOS-ET LAI and $f$APAR absolute values differ considerably. This is expected, as allowing some form of horizontal clumping (MODIS 3-D radiative transfer scheme) or not (TIP 1-D) can result in large differences in the estimated LAI and $f$APAR in forests. It can be seen that the local calibration of the MODIS-like product retains the dynamics of the WACMOS-ET product, while adding absolute values close to the MODIS product. The station Brookings (US-Bkg) is situated in a cropland area, where the effects of clumping are much less severe, and the different LAI and $f$APAR values are much closer.

#### 2.2.5 Vegetation height

Vegetation height on the global scale is required by SEBS. For shrubland and forest biomes the product developed by Simard et al. (2011) was used as static canopy height cover. For grassland and cropland biomes, where the temporal dynamics of canopy height can be more considerable, we approximated canopy height with the method by Chen et al. (2015), with the minimum and maximum canopy height obtained from the static vegetation table of the North American Land Data Assimilation System (NLDAS).

#### 2.2.6 Soil moisture and vegetation optical depth

A soil moisture product combining observations from active and passive microwave sensors has been developed as part of the ESA Climate Change Initiative (CCI) and is adopted here to provide the surface soil moisture data that are assimilated into GLEAM. Details on the product algorithm and evaluation can be found in Liu et al. (2011b). The data are provided on a regular grid with a resolution of $0.25^\circ \times 0.25^\circ$. The product performs well in moderately vegetated regions but shows higher uncertainties in densely vegetated regions (as vegetation attenuates the microwave signal from the ground)

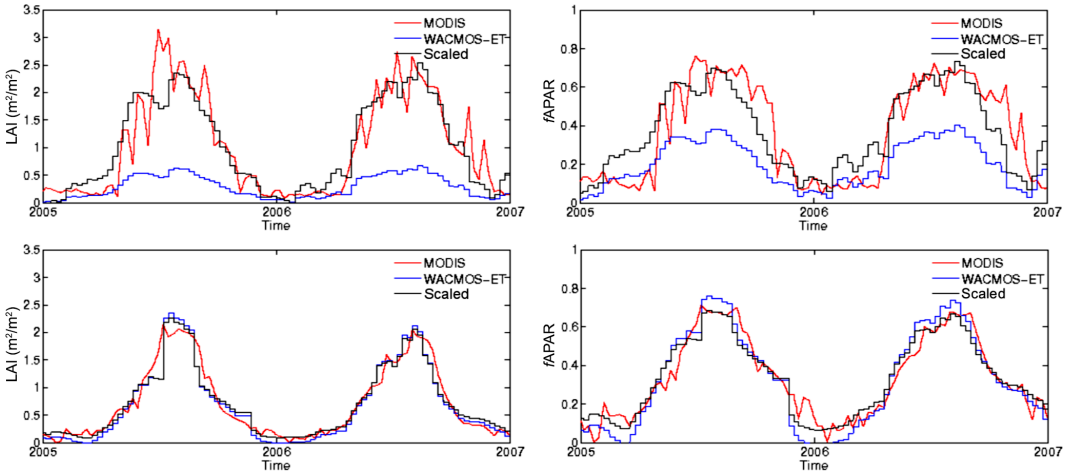


**Figure 1.** Time series for 2005–2006 of MODIS MOD15A2 LAI and $f$APAR, WACMOS-ET LAI and $f$APAR, and the MODIS-like LAI and $f$APAR (referred to as "scaled" in the figures) at the tower stations CA-Qfo (top panels) and US-Bkg (bottom panels).

and mountainous areas (due to the high surface roughness) (Liu et al., 2011b).

The retrieval of soil moisture from passive sensors discussed in Sect. 2.2.6 can be accompanied by an estimation of the vegetation optical depth (VOD). VOD can be used to account for the development of vegetation over the year as it is a good proxy of vegetation water content (Liu et al., 2015). Although most ET models traditionally use parameters derived from visible and near-infrared wavelengths, microwave VOD is used by GLEAM. Here the long-term record by Liu et al. (2011a) based on the application of the Land Parameter Retrieval Model by Owe et al. (2001) is used by GLEAM.

#### 2.2.7 Precipitation and snow

Observations of precipitation and snow water equivalent are also required by GLEAM only. Precipitation is used both to estimate the effects of soil water limitations on ET and to calculate interception loss. To run the model on the tower scale we use the Climate Prediction Center (CPC) Morphing Technique (CMORPH) (Joyce et al., 2004). CMORPH transports the features of precipitation estimates derived from low orbiter satellite microwave observations using information from geostationary satellite infrared (IR) data. Precipitation estimates are available every 30 min on a grid with a spacing of 8 km at the Equator, although the resolution of the individual satellite-derived estimates is coarser at $\sim 12\,\text{km} \times 15\,\text{km}$. The spatial coverage ranges from 60° N–60° S. To run the model globally, we use the National Centers for Environmental Prediction (NCEP) Climate Forecast System Reanalysis (CFSR) land precipitation estimates (Coccia and Wood, 2015). These precipitation estimates come from the hourly CFSR output (Saha et al., 2010) but are corrected using the observation-based data sets of the CPC (Xie and Arkin, 1997) and the Global Precipitation Climatology Project (GPCP) (Adler et al., 2003). Finally, snow water equivalent estimates come from ESA GlobSnow. Since GlobSnow covers the Northern Hemisphere only, data from the National Snow and Ice Data Center (NSIDC) are used in snow-covered regions of the Southern Hemisphere (Kelly et al., 2003). The product combines satellite passive microwave measurements with ground-based weather station data in a data assimilation scheme (Luojus and Pulliainen, 2010). The products exist at a daily resolution and a spatial resolution of 25 km.

### 2.3 Tower data

#### 2.3.1 Tower selection

Model simulations are evaluated by comparison with the turbulent latent fluxes measured by the eddy-covariance technique at a selection of tower sites from FLUXNET (Baldocchi et al., 2001). A first sample of towers was compiled by selecting those stations from the FLUXNET La Thuile synthesis data set which contain latent flux measurements in the 2005–2007 period, as well as the meteorological and radiation inputs required to run the ET models at the towers' locations. The 24 selected stations are described in Table 2 and their geographical location is displayed in Fig. 2. While some meteorological variables such as near-surface air temperature or humidity are measured at nearly all towers, other inputs such as the surface net radiation or the ground heat flux are measured at only a few towers. Some stations that were very close to the shore or in places with regular flooding were discarded. The final selection of 24 towers represents a significant number of biomes and a reasonable sample of dry and wet climate regimes.

In a later step, by removing the constraint of requiring local measurements of all the model inputs, the first selection of 24 towers is extended to a total of 85 stations. This second selection is used to evaluate model performance when

**Table 2.** Stations selected to run the models with tower inputs. From left to right: station name; longitude; latitude; Köppen–Geiger Climate Classification (KGCC); International Geosphere–Biosphere International Programme (IGBP) land cover; total number of days with data, no precipitation number of days with data; evaporative fraction (EF) for the DJF, MAM, JJA, SON 3-month periods.

| Name | Long | Lat | KGCC | IGBP | Days | EF |
|---|---|---|---|---|---|---|
| AU-How | 131.15° E | 12.49° S | Aw | SV | 114, 100 | 0.7, 0.7, 0.5, 0.3 |
| CA-Ojp | 104.69° W | 53.92° N | Dfc | ENF | 126, 101 | 0.2, 0.1, 0.3, 0.5 |
| CA-Qfo | 74.34° W | 49.69° N | Dfc | ENF | 253, 166 | 0.1, 0.1, 0.4, 0.4 |
| DE-Geb | 10.91° E | 51.1° N | Cfb | CRO | 188, 113 | 0.0, 0.4, 0.5, 0.7 |
| DE-Har | 7.60° E | 47.93° N | Cfb | MF | 105, 88 | 1.0, 0.5, 0.5, 0.7 |
| DE-Kli | 13.52° E | 50.89° N | Cfb | CRO | 275, 98 | 0.0, 0.5, 0.5, 0.0 |
| DE-Meh | 10.66° E | 51.28° N | Cfb | CRO | 444, 269 | 0.0, 0.3, 0.5, 0.5 |
| DE-Wet | 11.46° E | 50.45° N | Cfb | ENF | 384, 182 | 0.1, 0.3, 0.4, 0.6 |
| IT-MBo | 11.08° E | 46.03° N | Dfb | MF | 149, 126 | 0.0, 0.6, 0.7, 0.9 |
| IT-Noe | 8.15° E | 40.6° N | Csa | WSA | 182, 182 | 0.7, 0.3, 0.2, 0.3 |
| NL-Ca1 | 4.93° E | 51.97° N | Cfb | NVM | 38, 22 | 1.0, 0.6, 0.6, 0.9 |
| PT-Mi2 | 8.02° W | 38.48° N | Csa | SV | 275, 221 | 0.5, 0.4, 0.3, 0.4 |
| RU-Fyo | 32.92° E | 56.46° N | Dfb | MF | 374, 216 | 0.0, 0.4, 0.5, 0.4 |
| US-ARM | 97.49° W | 36.61° N | Cfa | CRO | 159, 131 | 0.4, 0.5, 0.3, 0.3 |
| US-Aud | 110.51° W | 31.59° N | BSk | OSH | 219, 219 | 0.5, 0.2, 0.3, 0.5 |
| US-Bkg | 96.84° W | 44.35° N | Dfa | CRO | 174, 172 | 0.6, 0.7, 0.9, 1.0 |
| US-Bo2 | 88.29° W | 40.01° N | Dfa | CRO | 192, 192 | 0.3, 0.3, 0.6, 0.3 |
| US-FPe | 105.1° W | 48.31° N | BSk | GRA | 184, 184 | 1.0, 0.4, 0.6, 0.4 |
| US-Goo | 89.87° W | 34.25° N | Cfa | NVM | 183, 179 | 0.7, 0.6, 0.5, 0.6 |
| US-MOz | 92.2° W | 38.74° N | Cfa | DBF | 252, 252 | 0.3, 0.4, 0.5, 0.5 |
| US-SRM | 110.87° W | 31.82° N | BSk | OSH | 139, 137 | 0.2, 0.1, 0.3, 0.4 |
| US-WCr | 90.08° W | 45.81° N | Dfb | DBF | 338, 239 | 0.1, 0.2, 0.6, 0.4 |
| US-Wkg | 109.94° W | 31.74° N | BSk | GRA | 137, 137 | 0.2, 0.1, 0.2, 0.3 |
| US-Wrc | 121.95° W | 45.82° N | Csb | ENF | 146, 107 | 0.4, 0.3, 0.3, 0.5 |

KGCC abbreviations: Aw – tropical wet and dry; BSk – semiarid midlatitudes; Cfa – humid subtropical; Cfb – marine, mild winter; Csa – interior Mediterranean; Csb – coastal Mediterranean; Dfa – humid continental hot summer, wet all year; Dfb: humid continental mild summer, wet all year; Dfc – subarctic with cool summer, wet all year. IGBP abbreviations: SV – savannas; ENF – evergreen needleleaf forests; CRO – croplands; MF – mixed forests; WSA – woody savannas; OSH – open shrublands; GRA – grasslands; DBF – deciduous broadleaf forests; NVM – natural vegetation mosaic.

the models are run with the satellite data used for the global runs.

#### 2.3.2 In situ surface energy balance

While, in principle, the surface energy balance should close at the tower, this is rarely the case: a lack of closure in the surface energy balance of about 10–30 % is commonly found when comparing the EC measurements against the energy balance residual (ER) term, i.e., the difference between net radiation and the sum of the sensible, latent and ground fluxes (e.g., Foken et al., 2006). Consequently, throughout the paper the model evaluation is discussed by comparing it with both the EC measurements and the in situ ER estimates.

#### 2.3.3 In situ LST

To run SEBS, the broadband longwave radiometer measurements need to be converted into LST estimates. This is done by inverting the equation relating the upwelling spectral radiance measured by the radiometer and the LST. Broadband emissivity is required, and it is estimated from the MODIS-based Global Infrared Land Surface Emissivity Database (Seemann et al., 2008) operated by the Cooperative Institute for Meteorological Satellite Studies (CIMSS). The estimates are calculated by following the approach suggested by Wang et al. (2005) using a linear combination of narrowband emissivities at 8.5, 11 and 12 µm.

#### 2.3.4 In situ vegetation height

SEBS also requires vegetation height to derive the surface roughness values. In most cases a mean annual value can be obtained from the tower metadata, and this value is adopted here as vegetation height at the tower. However, a clear limitation in this assumption is that it does not include dynamic changes in vegetation height over time. As discussed in Sect. 2.2, the importance of neglecting the temporal variability in height is biome-dependent; for instance, in forests the mean vegetation height is typically more constant than in, e.g., croplands, where the changes derived from agricultural practices can be large.

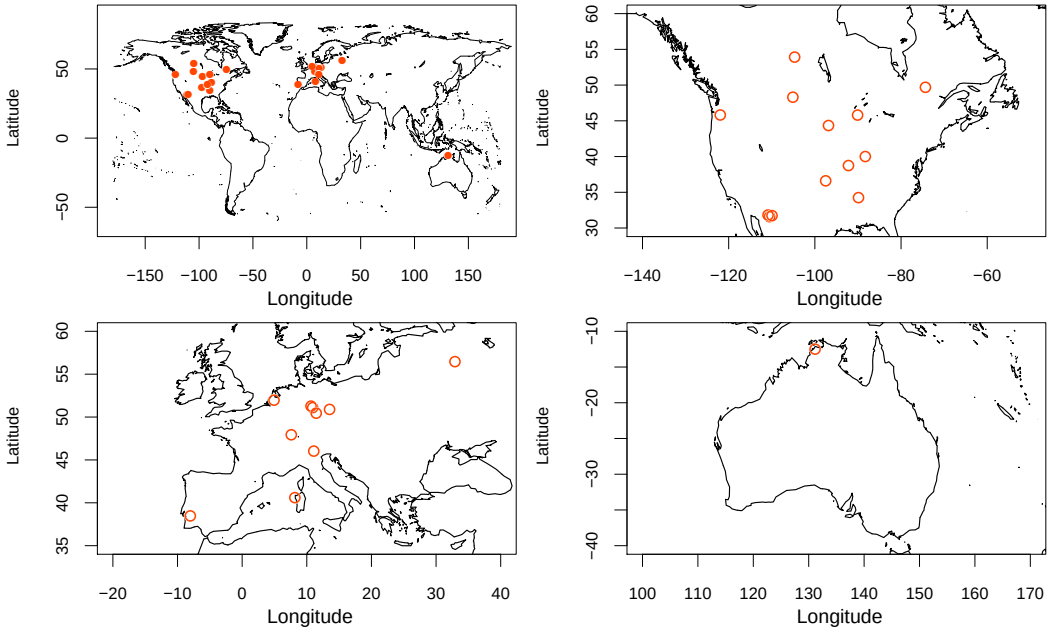


**Figure 2.** Location of the 24 FLUXNET stations used for the main analysis of the study. They are located on three different continents, encompassing nine different biomes, i.e., vegetation mosaic, croplands, mixed forests, deciduous forest, savanna, evergreen needleleaf forests, grasslands, woody savanna and shrublands.

## 2.4 ET experiments

### 2.4.1 Evaluation times

The model performance is investigated on sub-daily and daily timescales. The tower data are available at 0.5 h intervals and have been time-integrated to 3 hours in order to run the ET models at that sub-daily resolution. The satellite data have been time-matched to the 3-hourly or daily resolutions from their native resolution in different ways (see below), depending on the type of data and original resolution. The 3-hourly inputs were then aggregated to daily values in order to run the models with tower-based daily inputs. The tower data record is not always time-continuous, as in some instances there are gaps in the record. This is not a problem for the PM-MOD, PT-JPL and SEBS models because the ET estimates depends only on the instantaneous atmospheric or surface state. When inputs to the models and/or ET for the evaluation are missing, those three models are not run. Conversely, GLEAM requires continuous data records to update the soil moisture state variable. To facilitate running GLEAM with tower inputs, the tower measurements are gap-filled with the corresponding pixel data (see Sect. 2.2). ET estimates from these periods are removed after the runs, so, as before, only the time steps where tower forcing data are available are used for model evaluation.

The models are validated against the tower ET only under dry (non-raining) conditions, as EC gas analyzers are not reliable during rain events due to disturbance of the infrared signal by droplets on the sensor (Burba et al., 2010; Hirschi et al., 2016). Therefore, any days with precipitation as indicated by the tower or satellite precipitation are removed from the validation, and the interception component from PM-MOD, PT-JPL and GLEAM is not considered in the validations.

### 2.4.2 Nighttime ET

Only PM-MOD and GLEAM specifically deal with nighttime evaporation. Nevertheless, nighttime values are required from all models to integrate the 3-hourly ET estimates to daily values. For SEBS and PT-JPL negative nighttime estimates are set to 0 to allow the daily integration for those models. To separate day and night, daylight times are identified by calculating the solar zenith angle. Time intervals, where the cosine of the zenith angle is larger than 0.2, are kept as day values. This day and night separation may be less accurate than using a solar downward radiation threshold, but it allows a day–night flag for those stations without solar radiation measurements. The impact of setting ET from SEBS and PT-JPL to 0, as these models cannot specifically simulate nighttime conditions, is addressed in Sects. 3.1 and 3.2, where sub-daily periods, including daytime, are investigated.

### 2.4.3 ET production

The following ET estimates are generated to evaluate model performance.

– Tower-based ET: ET generated by the four models using the 3-hourly or daily in situ data (surface radiation, LST, air temperature, air humidity, and wind speed and precipitation), the scaled WACMOS-ET LAI and $f$APAR and gridded soil moisture and VOD data. Note that the 3-hourly tower-based ET estimates are also time-integrated to daily values, so daily ET estimates exist both from the runs with daily inputs and from the integration of the 3-hourly ET estimates.

– Original-resolution satellite-based ET: ET generated by the four models using the 3-hourly or daily satellite data (SRB surface radiation, ERA-Interim air temperature, humidity, wind speed, CMORPH precipitation, ESA-CCI soil moisture, scaled WACMOS-ET LAI and $f$APAR) at their original resolutions. In situ LST is still used here in order to have the same number of SEBS estimates as in the tower-based ET (cloudiness, satellite overpass time and revisiting times would have notably reduced the number of SEBS estimates if the satellite LST had been used). As for the tower-based ET, 3-hourly tower-based ET estimates are also time-integrated to daily values.

– Common-grid satellite-based ET: ET generated by the four models using the 3-hourly satellite data resampled to a common grid. In contrast to the previous runs, the satellite data are not applied at their original resolutions but after resampling them to the sinusoidal grid at $\sim 25$ km, adopted to produce the global model runs. Note that the CMORPH precipitation is replaced by the CFSR-Land product in order to have global coverage and that the LST is based on the AATSR observations.

## 3 Results and discussion

Here we look at the model performance against the in situ measurements, when the models are run with tower-based and satellite inputs. This section is divided into the three subsections, each of them dealing with one of the three experiments introduced in the previous Sect. 2.4.3. First, the 3-hourly and daily runs based on in situ forcing at 24 FLUXNET stations (see Table 2) are investigated. In the second part we look at the model performance at the same stations using 3-hourly and daily resolution satellite forcing. Finally, the ET estimates from the run using 3-hourly common-grid satellite forcing are compared to the in situ measurements at 85 FLUXNET stations.

### 3.1 Three-hourly and daily tower-based ET

The agreement of modeled evaporative fraction (EF) – defined here as $\lambda E/R_n$, using modeled $\lambda E$ and the net radiation from the respective forcing – with the measured EF (i.e., based on tower measurements of $\lambda E$ and $R_n$) gives an indication of the algorithm skill to model evaporative stress. Figure 3 (top panel) illustrates the agreement of modeled evaporative fraction with in situ measurements (derived using both EC and ER measurements of evapotranspiration; see Sect. 2.3.2), when models are run with tower inputs. GLEAM generally ranges between the EC and ER measurements, even at dry stations in open shrubland (OSH), woody savannas (WSA) and grassland (GRA) biomes, e.g., Sardinia/Arca di Noe (IT-Noe), Audubon Research Ranch (US-Aud), Santa Rita Mesquite (US-SRM) and Walnut Gulch Kendall Grasslands (US-Wkg). Only in the evergreen needleleaf forests (ENF) does GLEAM exceed the range of in situ measurements. PT-JPL mostly agrees with the reference as well, although it presents positive biases at some dry sites, like Wind River Crane (US-Wrc) and IT-Noe. PM-MOD underestimates EF for most stations (but it is very close to the EC measurements at six stations), while SEBS is characterized by an overall overestimation (for six stations SEBS EF is within the tower EC–ER range). In terms of the model performance per biome type, it can be stated that models generally perform the best in croplands (CRO) and deciduous broadleaf forest (DBF); at least this is the case for PT-JPL, PM-MOD and GLEAM. SEBS seems to perform better in grassland and savanna biomes (SAV). It is, however, difficult to derive robust conclusions on the model performance as a function of biome due to the low number of stations per biome type.

As the surface meteorology plays an important role in the ET production, we also compare the point-scale model performance with the gridded ERA-Interim ET data set (ERA) in Fig. 3 (top panel). ERA-Interim estimates are mostly within the range of EF measurements. The good agreement between ERA EF and the in situ measurements indicates that the ERA-Interim meteorology reliably captures the station conditions. It can also be stated that the point-scale tower-forced EF derived with PT-JPL and GLEAM match the ERA-Interim product based on a $\sim 75$ km resolution.

A statistical assessment of the model performance is given in Fig. 4, which shows the correlation ($R^2$), the RMSD and the average of the bias normalized by the reference (MBD) between modeled ET and tower measurement of ET (i.e., using the EC approach). In the left column of Fig. 4 the station averages of the statistical inferences are shown according to measured EF, i.e., from wet to dry. In general, the correlation to in situ data is high in wet and in moderately wet biomes for most sites and for all models. This is also true for SEBS, despite its substantial overestimation of EF (see Fig. 3). However, there seems to be a distinct decrease in $R^2$ from wet to dry biomes for all models; this decrease in performance is lower for GLEAM and higher for PM-MOD, which presents correlations ($R^2 < 0.4$) at dry sites. PT-JPL stands out amongst the ensembles with the highest correlation at most sites and especially in dry conditions. In comparison to the mostly underestimated evaporative fraction derived with PM-MOD (see Fig. 3), the RMSD of PM-MOD ET corre-

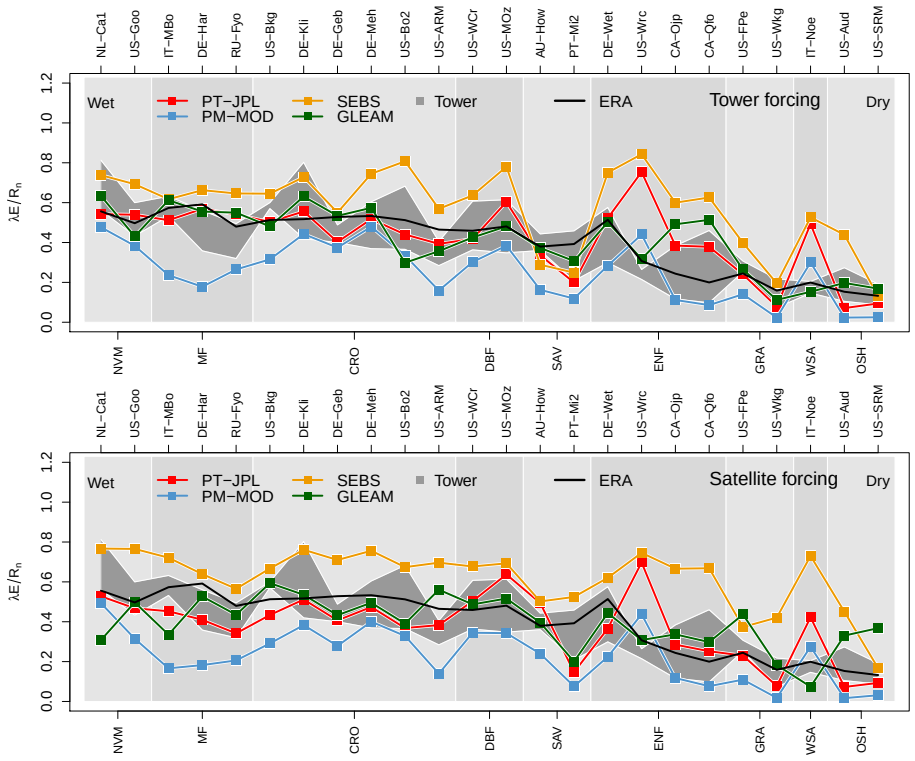


**Figure 3.** Station means of 3-hourly EC-observed and tower-forced (top panel) and satellite-forced (bottom panel) evaporative fraction against tower reference, as function of biomes, sorted from wet to dry (based on the biome average). The grey area denotes the range of evaporative fraction between EC and ER measurements. The black line denotes EF derived from ERA-Interim ET (ERA) and $R_n$.

sponds to PT-JPL and GLEAM and even produces the lowest maximum value (0.13 mm h$^{-1}$), followed by GLEAM (0.17 mm h$^{-1}$). Note that the large positive MBD values of PT-JPL and SEBS (> 200 %) may partly result from forcing ET to 0 during nighttime (see Sect. 2.4.2), when tower ET is negative, and thus leading to large relative errors, even for small negative reference ET values.

In order to evaluate the impact of using EC measurements as reference (in contrast to the ER method in Fig. 4), Table 3 shows the overall average 3-hourly model performance (i.e., the average of all station statistics) using both EC and ER data as reference. Overall, the average statistics of PT-JPL and GLEAM appear more favorable than those of SEBS and PM-Mu, although the RMSD and MBD of PM-MOD and the $R^2$ of SEBS are in general comparable to those of GLEAM and PT-JPL. This is again to a large extent affected by the overall overestimation and underestimation by SEBS and PM-MOD, respectively. The RMSD of SEBS is significantly smaller when using the ER method as reference (0.10 mm h$^{-1}$) as opposed to using EC (0.13 mm h$^{-1}$); on the other hand, the RMSD of PM-MOD is larger compared to ER (0.12 mm h$^{-1}$) than compared to EC (from 0.06 mm h$^{-1}$). Note that the transpiration resistances in PM-MOD are calibrated based on a biome-dependent annual ET derived from EC observations, which may explain the smaller RMSD and MBD when using EC as a reference. Finally, the RMSD station averages are similar to both in situ references for by PT-JPL (0.08, 0.09 mm h$^{-1}$) and GLEAM (0.08, 0.08 mm h$^{-1}$).

Here the skill of models at representing ET at specific times of the day is examined. Note that small nighttime ET values from models and measurements may produce small absolute errors and thus can improve the overall full-day model performance in comparison to daytime periods, even if the relative bias is large.

The Taylor diagrams in Fig. 5 show that correlations to in situ observations (using the EC method) considering the entire daily cycle (left panel) are very similar compared to those considering daytime values only (left panel, top row) or nighttime values only (right panel, top row). The overall $R^2$ with tower forcing including all models is 0.67 for full-day as well as daytime evaluation and 0.68 for nighttime evaluation; this indicates that the results are independent of the timescale. Note that nighttime is identified as cases when the cosine of the zenith angle is $< 0.2$.

We can see (right panel, top row) that forcing negative nighttime ET values of PT-JPL and SEBS to 0 (in contrast to specific negative ET produced by PM-MOD and GLEAM; see Sect. 2.4.2) does not have a substantial impact on the overall agreement with tower measurements. However, it should be noted that the uncertainty of nighttime EC mea-

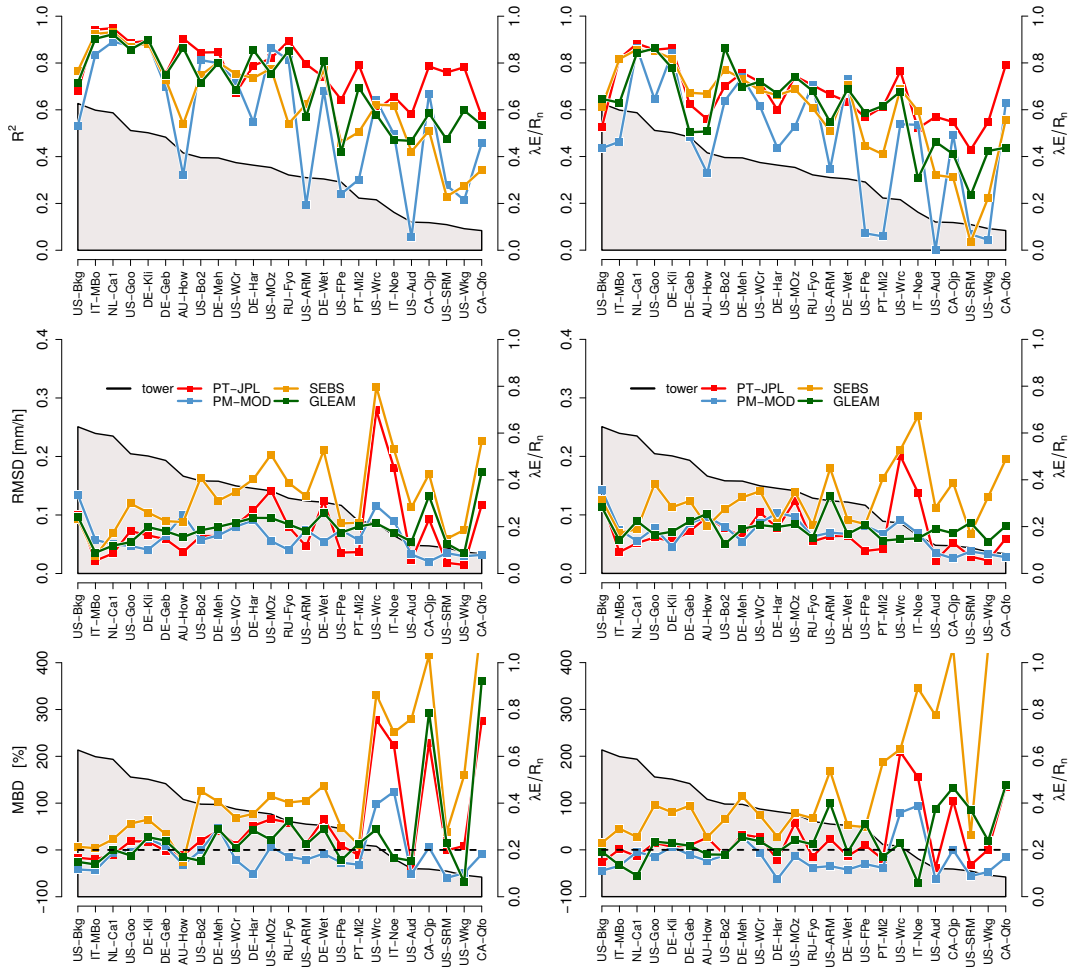


**Figure 4.** Station mean statistics of 3-hourly model data against EC reference. Left-column panels: tower-forced ET; right-column panels: satellite-forced ET; top-row panels: $R^2$ correlation coefficient (left $y$ axis); middle-row panels: mean bias deviation (MBD, left $y$ axis); bottom-row panels: root mean square difference (RMSD, left $y$ axis). For all plots the evaporative fraction is given by the grey area (right $y$ axis).

surements is high because of low turbulence. Hence, large nighttime errors can be present not only in the ET simulations but also in the EC data.

Sub-daily resolution is desirable in evaporation modeling, as it allows investigation of the underlying land–atmospheric interactions during the daily cycle of the planetary boundary layer. Given the short timescale of these interactions, one may expect that models that are able to reproduce short-term variability in ET would also be able to provide more reliable aggregates on daily timescales. Therefore, we investigate whether the model performance would benefit from solving evaporation at a 3-hourly resolution and aggregating it to daily values, as opposed to generating the estimates with daily input directly. Figure 5 (bottom row) clearly shows that not much more skill is gained by producing daily ET based on 3-hourly input (i.e., resolved diurnal cycles in the meteorological inputs) as opposed to forcing the models with the original daily input; results are almost identical when using aggregated 3-hourly output (left panel, bottom row) or using daily forcing (right panel, bottom row). In fact, for GLEAM the correlation to the EC reference is slightly higher when daily input is used, even if the standard deviation agrees marginally less well with the reference.

Figure 6 shows the statistics of the models' evaluation after forcing them with daily inputs. As expected, the general correlations become lower when daily (as opposed to 3-hourly) estimates are validated, since the daily cycle no longer plays a role in the enhancement of correlations – this was already highlighted by Table 3. Comparison of Figs. 4 and 6 shows that the decline in average $R^2$ from wet to dry stations is less evident at a daily resolution. This may be due to the smaller sample size when daily values are analyzed. PM-MOD and SEBS in particular correlate poorly at dry stations (also at other stations, such as the moderately wet AU-How). PT-JPL and GLEAM perform worse (compared with the 3-hourly resolution) at dry stations when they are run at

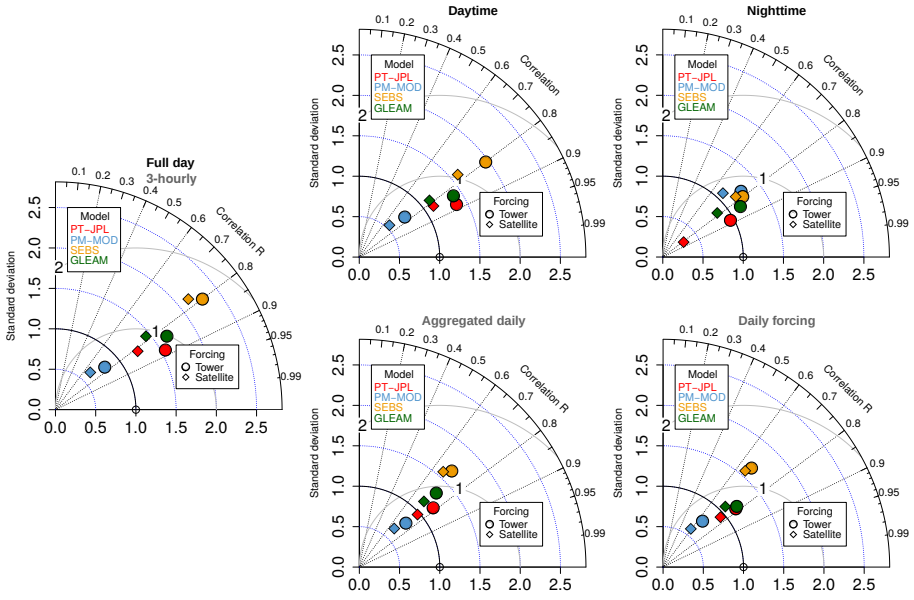


**Figure 5.** Taylor diagrams of 3-hourly model performance against EC reference in sub-daily periods (top-row panels) and as function of temporal resolution (bottom-row panels). The left panel shows the average model statistics for full-day (compare to top row) and 3-hourly output data (compare to bottom row). Daytime is defined as cases when the cosine of the sun elevation azimuth is $> 0.2$; nighttime is defined as cases when the cosine of the sun elevation azimuth is $< 0.2$. Shown are the normalized standard deviation, the normalized RMSD and the correlation coefficient ($R$).

**Table 3.** Summary of 24 stations average statistics for 3-hourly and daily tower forcing. EC denotes the model agreement with the evapotranspiration reference from eddy-covariance measurements, and ER is the model agreement with the evapotranspiration reference based on the in situ energy residual. RMSD is given in millimeters per hour for both 3-hourly data (3 h) and daily data (d).

| | | $R^2$ | | RMSD | | MBD | (%) |
|---|---|---|---|---|---|---|---|
| | | EC | ER | EC | ER | EC | ER |
| PT-JPL | 3 h | 0.77 | 0.78 | 0.08 | 0.09 | 53.1 | 37.4 |
| | d | 0.61 | 0.60 | 0.04 | 0.05 | 47.8 | 21.8 |
| PM-MOD | 3 h | 0.58 | 0.55 | 0.06 | 0.12 | −6.7 | −18.2 |
| | d | 0.43 | 0.41 | 0.04 | 0.06 | 3.8 | −11.3 |
| SEBS | 3 h | 0.64 | 0.78 | 0.13 | 0.10 | 125.9 | 78.4 |
| | d | 0.45 | 0.50 | 0.08 | 0.07 | 113.5 | 48.5 |
| GLEAM | 3 h | 0.70 | 0.80 | 0.08 | 0.08 | 31.9 | −8.7 |
| | d | 0.60 | 0.66 | 0.03 | 0.04 | 15.6 | −14.1 |

daily resolutions. In terms of the RMSD and MBD, the results are quite similar to the 3-hourly findings, but in most cases worst performance at the daily resolution is found at dry stations. An exception is GLEAM, which shows smaller RMSD at the dry stations when using daily rather than 3-hourly resolution.

The change in overall MBD (against the EC reference) from using 3-hourly tower input to using daily tower input is from 53.1 to 47.8 % for PT-JPL, from −6.7 to 3.8 % for PM-MOD, from 125.9 to 113.5 % for SEBS, and from 31.9 to 15.6 % for GLEAM. While the pattern of EF (Fig. 3) and MBD (Fig. 4) indicates a substantial underestimation of 3-hourly ET by PM-MOD, this underestimation is attenuated when daily input is used (−18.2 to −11.3 % compared to the ER reference). Note that even if we employ the term daily input, the PM-MOD model estimates day and night ET separately by using integrated day and night inputs (as opposed to PT-JPL, SEBS and GLEAM, which use daily integrated inputs) and then combines them to provide a daily value. This is how the PM-MOD model was originally used and how it is implemented in this study for daily estimation. The better agreement on a daily scale thus may reflect a more appropriate use of the inputs.

The similarity of the results for different temporal resolutions underlines the robustness of the modeling processes. PT-JPL and GLEAM agree best with the in situ measurements, while SEBS yields a good correlation in comparison to the other models yet produces the largest absolute errors due to its large overestimation. PM-MOD produces the lowest correlation but agrees rather well in terms of absolute deviations.

Table 3 summarizes the main statistics of the model evaluation for the 3-hourly and daily tower inputs.

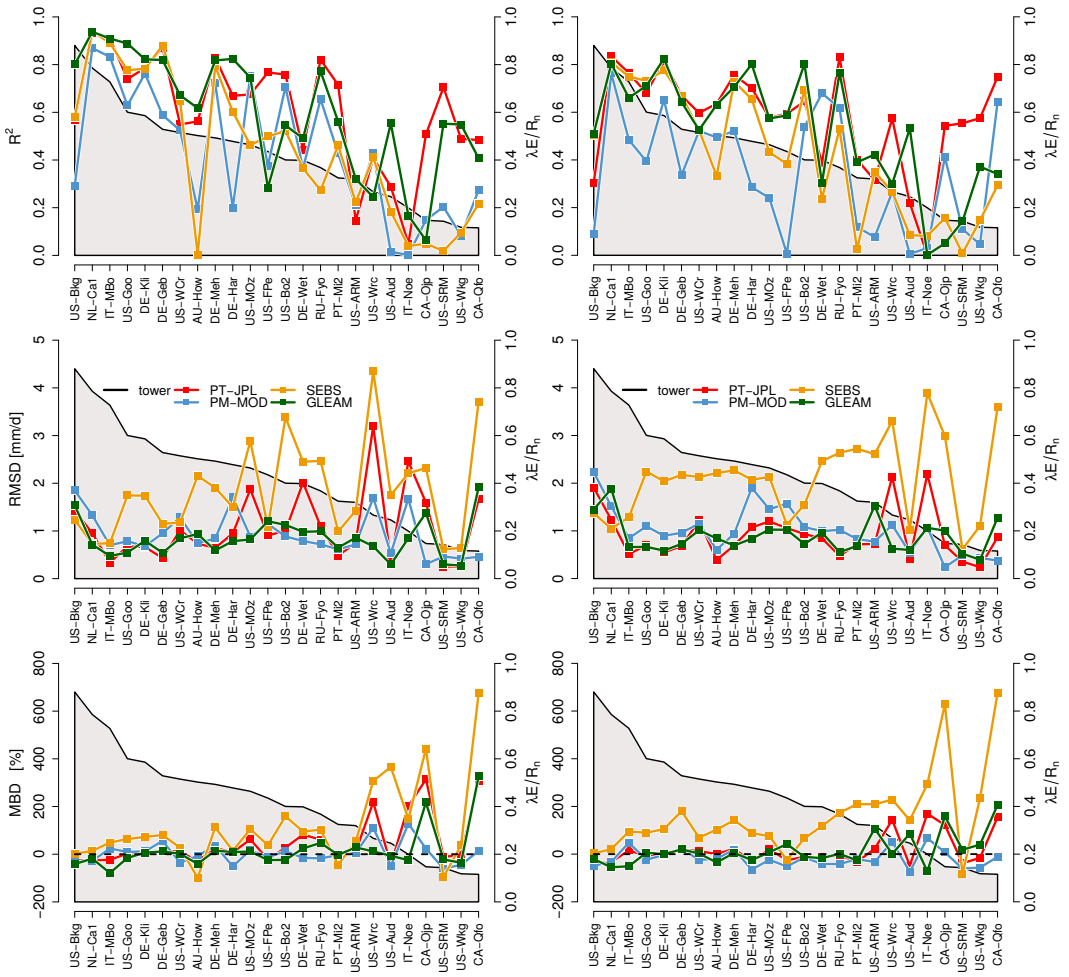


**Figure 6.** Station mean statistics of daily data from daily input against EC reference. Left-column panels: tower-forced ET; right-column panels: satellite-forced ET; top-row panels: $R^2$ correlation coefficient (left $y$ axis); middle-row panels: mean bias deviation (MBD, left $y$ axis); bottom-row panels: root mean square difference (RMSD, left $y$ axis). For all plots the evaporative fraction is given by the grey area (right $y$ axis).

### 3.2 Three-hourly and daily original-resolution satellite ET

In this section we discuss the model performance using 3-hourly and daily satellite forcing with original resolution at the selected 24 FLUXNET stations. The findings are compared to the results of the tower forcing in the previous section in order to allocate model uncertainty to either the algorithms used or the common forcing.

The evaluation of 3-hourly modeled EF using satellite forcing (Fig. 3, bottom panel) shows a very similar picture of agreement with the reference compared to the results of the tower forcing. Note that the satellite EF shown here slightly differs from tower-forced EF, as the data availability of the input time series may be different at some stations. The ET overestimation by SEBS seems to be slightly emphasized when using satellite input in comparison to the tower forcing. Note that the LST used in SEBS is still obtained from the tower measurements, as discussed in Sect. 2.4.1. EF derived with PT-JPL and GLEAM still agrees well with the reference, yet GLEAM overestimates EF in dry biomes when using satellite forcing but is more accurate at needleleaf forest sites. The good model performance of PT-JPL and GLEAM, independent of forcing type, indicates a robust performance of the models on the one hand and a reliable satellite forcing – in the sense of their meteorology comparing well with the in situ tower data – on the other hand.

In Fig. 3 (bottom panel) we also compare the model performance with the gridded ERA-Interim ET data set. Note that while the tower forcing runs (top panel) are independent of ERA-Interim, the satellite runs use ERA-Interim estimates as inputs for the surface meteorology. As shown in Sect. 3.1, the ERA-Interim EF product agrees with the in situ measurements. The correlation of the models to ERA-Interim is not substantially improved with satellite input in comparison to

the tower forcing, although the models now use the ERA-Interim meteorology as input.

The station averages of the statistical indices $R^2$, RMSD and MBD of the models forced with satellite observations (Fig. 4, bottom panel) against the in situ measurements underline the previously reported high similarity of modeled ET based on tower and satellite forcing. Only the RMSD of SEBS is slightly attenuated with remotely sensed forcing. However, the algorithm is still characterized by substantial overestimation.

In the following we compare the model performance with daily satellite forcing to the model performance with daily tower forcing. In accordance with the evaluation of 3-hourly data (see Fig. 4), Fig. 6 indicates that the daily satellite-based ET products also correspond to the tower-based modeled ET. We want to highlight, however, that in contrast to the 3-hourly runs, the RMSD of SEBS substantially increases when satellite input is used. This suggests that the SEBS physical modeling captures the ET processes more accurately with the high temporal resolution inputs (3-hourly vs. daily).

Table 4 provides a summary of the main statistics of the model evaluation for the 3-hourly and daily satellite inputs.

**Table 4.** Summary of 24 station average statistics for 3-hourly and daily satellite forcing. ERA denotes the agreement of ERA-Interim evapotranspiration with the in situ reference evapotranspiration. For other abbreviations, see Table 3. RMSD is given in millimeters per hour for both 3-hourly data (3 h) daily data (d).

| | | $R^2$ | | RMSD | | MBD | (%) |
|---|---|---|---|---|---|---|---|
| | | EC | ER | EC | ER | EC | ER |
| PT-JPL | 3 h | 0.67 | 0.68 | 0.07 | 0.11 | 25.8 | 14.6 |
| | d | 0.57 | 0.49 | 0.04 | 0.05 | 16.5 | −1.7 |
| PM-MOD | 3 h | 0.47 | 0.47 | 0.07 | 0.14 | −16.1 | −27.2 |
| | d | 0.35 | 0.29 | 0.04 | 0.06 | −17.9 | −34.2 |
| SEBS | 3 h | 0.59 | 0.71 | 0.13 | 0.11 | 145.1 | 148.6 |
| | d | 0.42 | 0.41 | 0.09 | 0.08 | 160.2 | 123.9 |
| GLEAM | 3 h | 0.61 | 0.72 | 0.08 | 0.10 | 22.7 | −4.6 |
| | d | 0.52 | 0.52 | 0.04 | 0.05 | 16.2 | −10.6 |
| ERA | 3 h | 0.51 | 0.45 | 0.10 | 0.14 | 111.3 | 87.0 |
| | d | 0.62 | 0.50 | 0.07 | 0.07 | 114.7 | 74.6 |

### 3.3 Three-hourly common-grid satellite ET

Here the ET algorithms are tested against 85 FLUXNET stations using the gridded sinusoidal ($\sim 25$ km) satellite input (as opposed to using their original input resolutions) in order to evaluate the common-gridded global ET estimates on the tower scale. Only the evaluation over the towers is discussed here, with the evaluation on the global scale discussed in the companion paper of Miralles et al. (2016). Note that the spatial mismatch between the tower fetch and the $\sim 750\,\mathrm{km}^2$ of the gridded cells is very large, and the agreement between the tower fluxes and the modeled ET certainly depends on the tower conditions being representative of the corresponding gridded pixel. This was also the case for some of the original-resolution satellite inputs used over the 24 stations, such as the SRB radiation or the ERA-Interim meteorology. The results of the satellite runs using common-grid forcing are compared to the results using the tower and satellite inputs on the tower scale presented in Sects. 3.1 and 3.2.

The top panel of Fig. 7 shows the mean 3-hourly EF over 70 stations for PM-MOD, PT-JPL and GLEAM. For 15 of the 85 stations the surface radiation or the ground flux was not available; hence, the ER reference could not be calculated. As the gridded inputs use satellite LST from AATSR, SEBS ET is only estimated at the midmorning AATSR overpass. The bottom panel of Fig. 7 shows the annual midmorning evaporative fraction, this time including SEBS. Due to the 3-day revisiting time of AATSR and the lack of measurements in cloudy conditions, the number of available SEBS ET estimates reduces drastically, compared with the previous simulations using tower LST. The bottom panel of Fig. 7 shows station averages from all models only when SEBS ET is available. Thus, it is based on fewer data and with the number of stations reduced to 67.

The 3-hourly model performances from PM-MOD, PT-JPL and GLEAM correspond closely to the performance in the analysis using the 24 towers and the original-resolution satellite inputs. The EF station averages produced by PT-JPL and GLEAM are very close at all locations and respond well to the hydrological and energetic conditions expected in the respective biome. The overall agreement with the range between EC and ER in situ measurements is comparable to what has previously been found in the smaller sample of stations (see Fig. 3). PM-MOD keeps underestimating ET, except for the cropland biome, where the majority of station averages matches well with the reference. Concerning the midmorning evaporative fractions, the PM-MOD, PT-JPL and GLEAM patterns are all very similar to the case with the full diurnal cycle. SEBS again tends to overestimate over a large number of stations, compared with the in situ measurements. Overall, it can be stated that the model accuracy and inter-model agreement obtained with in situ and satellite forcing on the tower scale could be reproduced with the common-grid satellite forcing.

Figure 8 summarizes the results above by displaying standard deviation, correlation and RMSD of the modeled ET shown in Fig. 7 against the EC reference. The Taylor plots highlight the fact that the variability of PT-JPL, PM-MOD and GLEAM is not substantially influenced by the low sample size for cases when SEBS ET is available. Again, the similarity between Fig. 5 (left panel) for satellite forcing on the tower scale and Fig. 8 for gridded input data confirms the robustness of the analyses independent of tower and time sampling.

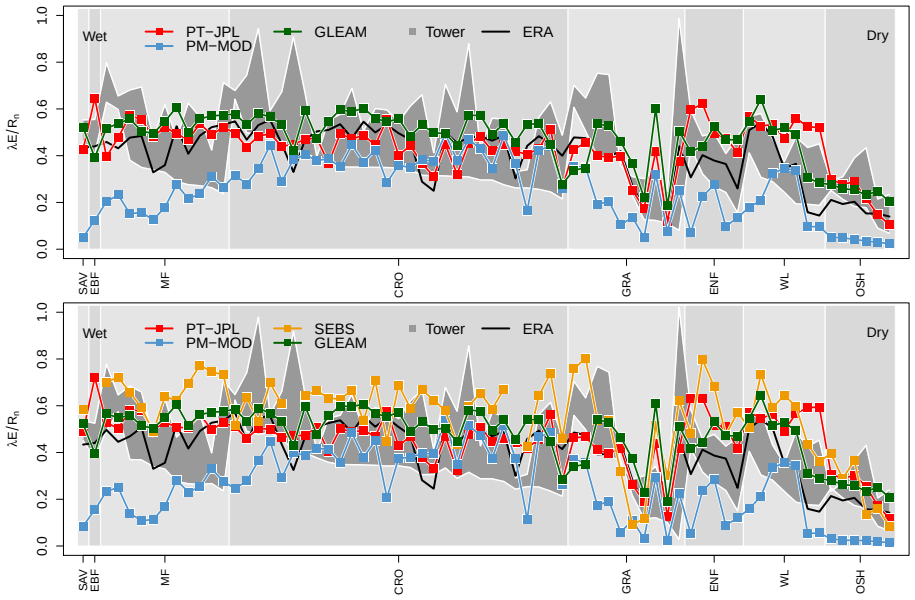


**Figure 7.** Top panel: station means of 3-hourly sinusoidal gridded satellite-forced evaporative fraction for full days (70 stations) against tower reference, as function of biomes, sorted from wet to dry (based on the biome average). Bottom panel: same as top panel but for midmorning only (from 09:00 to 13:00 LT; 67 stations). The grey area denotes the range of evaporative fraction between EC and ER measurements. The black line denotes EF derived from ERA-Interim ET and $R_n$.

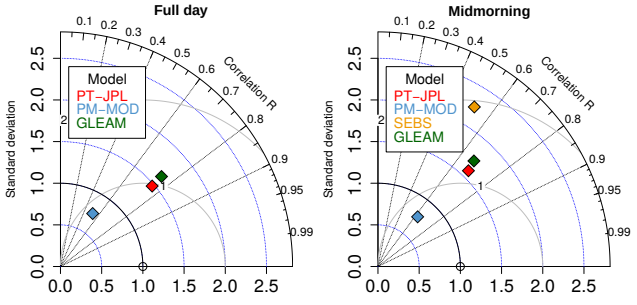


**Figure 8.** Taylor diagram plots of sinusoidal gridded model data against tower EC reference. Left panel: full-day 3-hourly data compared to 85 stations. Right panel: midmorning data (from 09:00 and 13:00 LT) compared to 82 stations. Shown are the normalized standard deviation, the normalized RMSD and the correlation coefficient ($R$).

## 4 Conclusions

In this first part of the WACMOS-ET study, the skill of the PT-JPL, PM-MOD, SEBS and GLEAM ET algorithms has been tested on the tower scale against in situ measurements at 24 FLUXNET sites. The algorithms are forced using in situ meteorological data from these towers, covering the period 2005–2007 on three continents and across nine different biomes, while ensuring spatial consistency between input and reference data. Additionally, the models are run for the same period with reanalysis and satellite forcing of varying spatial resolutions, including ERA-Interim (surface meteorology), SRB (radiation), AATSR (LST), GlobAlbedo (LAI and $f$APAR), CMORPH (precipitation) and WACMOS-CCI (soil moisture). The models were run with 3-hourly and daily input to assess the robustness of their performance for sub-daily and daily resolution.

Our analyses have shown that the four models' performance is robust in terms of changes in forcing types and temporal resolutions (i.e., the changes do not alter the model behavior at the selected stations significantly). Against the in situ 3-hourly energy residual estimates at the tower, the tower-based model simulations are ranked (according to station averages) as follows: GLEAM (0.80, 0.08), PT-JPL (0.78, 0.09), SEBS (0.78, 0.10) and PM-MOD (0.55, 0.12). The first value in the brackets denotes $R^2$ and the second value denotes RMSD in millimeters per hour. Compared to the eddy-covariance measurements, however, the station averages of RMSD do not reflect the same outcome. Due to more substantial overestimation at two stations each, the RMSD of PT-JPL (0.77, 0.08) and GLEAM (0.70, 0.08) are larger than that of PM-MOD (0.58, 0.06). However, correlations are consistently higher for GLEAM and PT-JPL. Thus, over our selection of towers and the reference period (2005–2007), we judge GLEAM and PT-JPL as the algorithms more closely matching the in situ observations. At some stations, PM-MOD and SEBS also agree well with the observations, but in general the PM-MOD and SEBS performance is characterized by under- and overestimation, respectively.

For the satellite forcing, the RMSD between the models and the reference yields very similar numbers as for tower

forcing. Correlations are closer but in most situations slightly smaller for the satellite forcings. This can be the result of discrepancies between the spatial resolution of satellite observations and tower measurements, although different inputs errors (in situ vs. satellite) may also play a role. This performance closeness between in situ and satellite-derived values can be an indication of the spatial representativeness of the tower measurements (i.e., reasonable spatial homogeneity around the tower) and the consistency of the input data set across forcing types. This is underlined by a comparison to the ∼ 75 km resolved reanalysis ET product of ERA-Interim, which agrees well with the modeled ET across the different biomes.

Regarding the analysis over the 85 stations, a similar overall picture is obtained using the ∼ 25 km common-grid ET prepared for the global runs. The evaluations of McCabe et al. (2015) of a different selection of towers (45 stations), over a more extended period (1997–2007) and with different satellite forcings (LandFlux forcings) also results in an overall similar analysis, confirming the robustness of the model performance evaluations.

Using daily input data reduces the RMSD of the models with the tower measurements but results in slightly worse correlations. This is due to the lower variability of daily values in contrast to 3-hourly data (variability accentuated by the diurnal cycle). However, the consistency of the model agreements with the reference with regard to 3-hourly and daily ET estimates highlights the robustness of the integration method applied to the models. This is also underlined by the good agreements of modeled daily ET from aggregated 3-hourly output data with modeled daily ET from daily input.

While GLEAM and PM-MOD can produce negative ET, PT-JPL and SEBS cannot operate under these conditions (mostly at nighttime when the flux of available energy reverses sign) and their negative values are forced to 0. This does not have a large impact on their full-day performance, since these values occur at night, when tower ET is negative and with generally low values. Only for the relative bias is the effect significant, since the two models consequently overestimate ET in these cases.

In terms of high and low temporal input resolution, it was found that using 3-hourly input data does not significantly increase the accuracy of the models for producing daily ET. Hence, it is sufficient to use daily input to achieve a similar result if the intended application of the ET product does not demand a reproduction of the diurnal cycle.

The ET models generally perform best in wet biomes and tend to overestimate values in dry biomes, where ET is constrained by water availability. Focusing on water stress in the model development within the community would thus provide the opportunity to obtain more robust simulations of surface fluxes for global-scale employment.

The conducted analyses based on in situ ET are useful to evaluate model performance, but there are some clear limitations. Our requirements for tower selection resulted in a somewhat limited number of stations, so it would be desirable to extend the evaluations to larger regions in order to better cover different climate and biome conditions. Therefore, in the companion paper of Miralles et al. (2016) our analyses are extended by looking at the global spatiotemporal variability of the modeled ET, the closure of regional water budgets, and the discrete estimation of land evaporation components or sources (i.e., transpiration, interception loss and direct soil evaporation).

*Acknowledgements.* This study was funded by the European Space Agency (ESA) and conducted as part of the project WACMOS-ET (Contract no. 4000106711/12/I-NB). D. G. Miralles acknowledges the financial support from the Netherlands Organization for Scientific Research through grant 863.14.004 and the Belgian Science Policy Office (BELSPO) in the framework of the STEREO III programme, project SAT-EX (SR/00/306). M. F. McCabe and A. Ershadi acknowledge the support of the King Abdullah University of Science and Technology. The SEBS team is acknowledged for facilitating discussions concerning the implementation of their model. This work used eddy-covariance data acquired by the FLUXNET community and in particular by the following networks: AmeriFlux (US Department of Energy, Biological and Environmental Research, Terrestrial Carbon Program, DE-FG02-04ER63917 and DE-FG02-04ER63911), AfriFlux, AsiaFlux, CarboAfrica, CarboEuropeIP, CarboItaly, CarboMont, ChinaFlux, Fluxnet-Canada (supported by CFCAS, NSERC, BIOCAP, Environment Canada and NRCan), GreenGrass, KoFlux, LBA, NECC, OzFlux, TCOS-Siberia and USCCC. Data and logistical support for the station US-Wrc were provided by the US Forest Service Pacific Northwest Research Station. All WACMOS-ET forcing data and ET estimates are publicly available and can be requested through the project website (http://wacmoset.estellus.eu).

Edited by: P. Gentine

## References

Adler, R., Huffman, G., Chang, A., Ferraro, R., Xie, P.-P., Janowiak, J., Rudolf, B., Schneider, U., Curtis, S., Bolvin, D., Gruber, A., Susskind, J., Arkin, P., and Nelkin, E.: The version-2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1970–Present), J. Hydrol., 4, 1147–1167, 2003.

Anderson, M., Norman, J., Mecikalski, J., Otkin, J., and Kustas, W.: A climatological study of evapotranspiration and moisture stress across the continental United States based on thermal remote sensing: 1. Model formulation, J. Geophys. Res., 112, D10117, doi:10.1029/2006JD007506, 2007.

Baldocchi, D., Falge, E., Gu, L., Olson, R., Hollinger, D., Running, S., Anthoni, P., Bernhofer, C., Davis, K., Evans, R., Fuentes, J., Goldstein, A., Katul, G., Law, B., Lee, X., Malhi, Y., Meyers, T., Munger, W., Oechel, W., Paw, K., Pilegaard, K., Schmid, H., Valentini, R., Verma, S., Vesala, T., Wilson, K., and Wofsyn, S.: FLUXNET: a new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water va-

por, and energy flux densities, B. Am. Meteorol. Soc., 82, 2415–2434, 2001.

Burba, G., McDermitt, D., Anderson, D., and Furtaw, M. D.: Novel design of an enclosed $CO_2/H_2O$ gas analyser for eddy covariance flux measurements, Tellus, 62, 743–748, 2010.

Chen, X., Su, Z., Ma, Y., Liu, S., Yu, Q., and Xu, Z.: Development of a 10-year (2001–2010) 0.1° data set of land-surface energy balance for mainland China, Atmos. Chem. Phys., 14, 13097–13117, doi:10.5194/acp-14-13097-2014, 2014.

Coccia, G. and Wood, E. F.: CFSR-Land: a new high temporal resolution global land data assimilation product, J. Geophys. Res., in preparation, 2015.

Courault, D., Seguin, B., and Olioso, A.: Review on estimation of evapotranspiration from remote sensing data: from empirical to numerical modeling approaches, Irrig. Drain. Syst., 19, 223–249, 2005.

Dee, D., Uppala, M., S., Simmons, J., A., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, A., M., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, M., A. C., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, J., A., Haimberger, L., Healy, B., S., Hersbach, H., Hólm, V., E., Isaksen, L., Kallberg, P., Khaler, M., Matricardi, M., McNally, P., A., Monge-Sanz, M., B., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thapaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Q. J. Roy. Meteorol. Soc., 137, 553–597, doi:10.1002/qj.828, 2011.

Ershadi, A., McCabe, M. F., Evans, J. P., Chaney, N. W., and Wood, E. F.: Multi-site evaluation of terrestrial evaporation models using FLUXNET data, Agr. Forest Meteorol., 187, 46–61, 2014.

Ferguson, C. R. and Wood, E. F.: An evaluation of satellite remote-sensing data products for land surface hydrology: Atmospheric Infrared Sounder (AIRS), J. Hydrometeorol., 11, 1234–1262, 2010.

Fisher, J., Tu, K., and Baldocchi, D.: Global estimates of the land-atmosphere water flux based on monthly AVHRR and ISLSCP-II data, validated at 16 FLUXNET sites, Remote Sens. Environ., 112, 901–919, 2008.

Foken, T., Wimmer, F., Mauder, M., Thomas, C., and Liebethal, C.: Some aspects of the energy balance closure problem, Atmos. Chem. Phys., 6, 4395–4402, doi:10.5194/acp-6-4395-2006, 2006.

Gash, J.: An analytical model of rainfall interception by forests, Q. J. Roy. Meteorol. Soc., 105, 43–45, 1979.

Gash, J. and Stewart, J.: The evaporation from Thetford Forest during 1975, J. Hydrol., 35, 385–396, 1977.

Hirschi, M., Michel, D., Lehner, I., and Seneviratne, S. I.: A site-level comparison of lysimeter and eddy-covariance flux measurements of evapotranspiration, Hydrol. Earth Syst. Sci., in preparation, 2016.

Jia, L., Su, Z., van den Hurk, B., Menenti, M., Moene, A., De Bruin, H. A. R., and Yrisarry, J. J. B.: Estimation of sensible heat flux using the Surface Energy Balance System (SEBS) and ATSR measurements, Phys. Chem. Earth, 28, 75–88, 2003.

Jiménez, C., Prigent, C., Mueller, B., Seneviratne, S. I., McCabe, M. F., Wood, E. F., Rossow, W. B., Balsamo, G., Betts, A. K., Dirmeyer, P. A., Fisher, J. B., Jung, M., Kanamitsu, M., Reichle, R. H., Reichstein, M., Rodell, M., Sheffield, J., Tu, K., and Wang, K.: Global intercomparison of 12 land surface heat flux estimates, J. Geophys. Res., 116, 6809–6866, D02102, doi:10.1029/2010JD014545, 2011.

Jimenez, C., McCabe, M., Seneviratne, S. I., Wood, E., and Rossow, W.: 4th LandFlux Workshop, GEWEX News, 22, 17–18, 2012.

Joyce, R. J., Janowiak, J. E., Arkin, P. A., and Xie, P.: CMORPH: a method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution, J. Hydrometeorol., 5, 487–503, 2004.

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

Kalma, J., McVicar, T., and McCabe, M.: Estimating land surface evaporation: a review of methods using remotely sensed surface temperature data, Surv. Geophys., 29, 421–469, doi:10.1007/s10712-008-9037-z, 2008.

Kelly, R., Chang, A., Tsang, L., and Foster, J.: A prototype AMSR-E global snow area and snow depth algorithm, IEEE T. Geosci. Remote, 41, 230–242, 2003.

Kustas, W. and Norman, J.: A two-source energy balance approach using directional radiometric temperature observations for sparse canopy covered surfaces, Agron. J., 92, 847–854, 2000.

Liu, Y. Y., de Jeu, R. A. M., McCabe, M., Evans, J., and van Dijk, A. I. J. M.: Global long-term passive microwave satellite-based retrievals of vegetation optical depth, Geophys. Res. Lett., 38, L18402, doi:10.1029/2011GL048684, 2011a.

Liu, Y. Y., Parinussa, R. M., Dorigo, W. A., De Jeu, R. A. M., Wagner, W., van Dijk, A. I. J. M., McCabe, M. F., and Evans, J. P.: Developing an improved soil moisture dataset by blending passive and active microwave satellite-based retrievals, Hydrol. Earth Syst. Sci., 15, 425–436, doi:10.5194/hess-15-425-2011, 2011b.

Liu, Y. Y., van Dijk, A. I. J. M., de Jeu, R. A. M., Canadell, J. G., McCabe, M. F., Evans, J. P., and Wang, G.: Recent reversal in loss of global terrestrial biomass, Nat. Clim. Change, 5, 470–474, doi:10.1038/nclimate2581, 2015.

Luojus, K. and Pulliainen, J.: Global Snow Monitoring for Climatic Research: Snow Water Equivalent (SWE) Product Guide, Finnish Meteorological Institute, Helsinki, Finland, 2010.

Mao, J., Fu, W., Shi, X., Ricciuto, D. M., Fisher, J. B., Dickinson, R. E., Wei, Y., Shem, W., Piao, S., Wang, K., Schwalm, C. R., Tian, H., Mu, M., Arain, A., Ciais, P., Cook, R., Dai, Y., Hayes, D., Hoffman, F. M., Huang, M., Huang, S., Huntzinger, D. N., Ito, A., Jain, A., King, A. W., Lei, H., Lu, C., Michalak, A. M., Parazoo, N., Peng, C., Peng, S., Poulter, B., Schaefer, K., Jafarov, E., Thornton, P. E., Wang, W., Zeng, N., Zeng, Z., Zhao, F., Zh, Q., and Zhu, Z: Disentangling climatic and anthropogenic controls on global terrestrial evapotranspiration trends, Environ. Res. Lett., 10, 094008, doi:10.1088/1748-9326/10/9/094008, 2015.

Martens, B., Miralles, D. G., Lievens, H., Fernández-Prieto, D., and Verhoest, N.: Improving terrestrial evaporation estimates over continental Australia through assimilation of SMOS soil moisture, Int. J. Appl. Earth Obs., doi:10.1016/j.jag.2015.09.012, in press, 2016.

McCabe, M. F. and Wood, E.: Scale influences on the remote estimation of evapotranspiration using multiple

satellite sensors, Remote Sens. Environ., 105, 271–285, doi:10.1016/j.rse.2006.07.006, 2006.

McCabe, M. F., Ershadi, A., Jimenez, C., Miralles, D. G., Michel, D., and Wood, E. F.: The GEWEX LandFlux project: evaluation of model evaporation using tower-based and globally-gridded forcing data, Geosci. Model Dev. Discuss., 8, 6809–6866, doi:10.5194/gmdd-8-6809-2015, 2015.

Miralles, D. G., Gash, J. H., Holmes, T. R. H., de Jeu, R. A. M., and Dolman, A. J.: Global canopy interception from satellite observations, J. Geophys. Res., 115, D16122, doi:10.1029/2009JD013530, 2010.

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

Miralles, D. G., De Jeu, R. A. M., Gash, J. H., Holmes, T. R. H., and Dolman, A. J.: Magnitude and variability of land evaporation and its components at the global scale, Hydrol. Earth Syst. Sci., 15, 967–981, doi:10.5194/hess-15-967-2011, 2011b.

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., and Fernandez-Prieto, D.: The WACMOS-ET project – Part 2: Evaluation of global terrestrial evaporation data sets, Hydrol. Earth Syst. Sci., 20, 823–842, doi:10.5194/hess-20-823-2016, 2016.

Monin, A. and Obukhov, A.: Basic laws of turbulent mixing in the surface layer of the atmosphere, Tr. Akad. Nauk SSSR Geophiz. Inst., 24, 163–187, 1954.

Monteith, J.: Evaporation and environment, Sym. Soc. Exp. Biol., 19, 205–234, 1965.

Mooney, P. A., Mulligan, F. J., and Fealya, R.: Comparison of ERA-40, ERA-Interim and NCEP/NCAR reanalysis data with observed surface air temperatures over Ireland, Int. J. Cancer, 31, 545–557, 2011.

Mu, Q., Heinsch, F. A., Zhao, M., and Running, S. W.: Development of a global evapotranspiration algorithm based on MODIS and global meteorology data, Remote Sens. Environ., 111, 519–536, 2007.

Mu, Q., Zhao, M., and Running, S.: Improvements to a MODIS global terrestrial evapotranspiration algorithm, Remote Sens. Environ., 115, 1781–1800, 2011.

Mueller, B., Seneviratne, S. I., Jimenez, C., Corti, T., Hirschi, M., Balsamo, G., Ciais, P., Dirmeyer, P., Fisher, J. B., Guo, Z., Jung, M., Maignan, F., McCabe, M. F., Reichle, R., Reichstein, M., Rodell, M., Sheffield, J., Teuling, A. J., Wang, K., Wood, E. F., and Zhang, Y.: Evaluation of global observations-based evapotranspiration datasets and IPCC AR4 simulations, Geophys. Res. Lett., 38, L06402, doi:10.1029/2010GL046230, 2011.

Mueller, B., Hirschi, M., Jimenez, C., Ciais, P., Dirmeyer, P. A., Dolman, A. J., Fisher, J. B., Jung, M., Ludwig, F., Maignan, F., Miralles, D. G., McCabe, M. F., Reichstein, M., Sheffield, J., Wang, K., Wood, E. F., Zhang, Y., and Seneviratne, S. I.: Benchmark products for land evapotranspiration: LandFlux-EVAL multi-data set synthesis, Hydrol. Earth Syst. Sci., 17, 3707–3720, doi:10.5194/hess-17-3707-2013, 2013.

Murphy, D. and Koop, T.: Review of the vapour pressures of ice and supercooled water for atmospheric applications, Q. J. Roy. Meteorol. Soc., 131, 1539–1565, 2005.

Owe, M., de Jeu, R., and Walker, J.: A methodology for surface soil moisture and vegetation optical depth retrieval using the microwave polarization difference index, IEEE T. Geosci. Remote, 39, 1643–1654, 2001.

Pinty, B., Lavergne, T., Vossbeck, M., Kaminski, T., Aussedat, O., Giering, R., Gobron, N., Taberner, M., Verstraete, M. M., and Widlowski, J.-L.: Retrieving surface parameters for climate models from Moderate Resolution Imaging Spectroradiometer (MODIS)-Multiangle Imaging Spectroradiometer (MISR) albedo products, J. Geophys. Res.-Atmos., 112, D10116, doi:10.1029/2006JD008105, 2007.

Pinty, B., Jung, M., Kaminski, T., Lavergne, T., Mund, M., Plummer, S., Thomas, E., and Widlowski, J.: Evaluation of the JRC-TIP 0.01° products over a mid-latitude deciduous forest site, Remote Sens. Environ., 115, 3567–3581, 2011a.

Pinty, B., Taberner, M., Haemmerle, V., Paradise, S., Vermote, E., Verstraete, M., Gobron, N., and Widlowski, J.-L.: Global-Scale Comparison of MISR and MODIS Land Surface Albedos, J. Climate, 24, 732–749, 2011b.

Priestley, C. and Taylor, R.: On the Assessment of surface heat flux and evaporation using large-scale parameters, Mon. Weather Rev., 100, 81–92, 1972.

Saha, S., Moorthi, S., Pan, H.-L., et al.: The NCEP climate forecast system reanalysis, B. Am. Meteorol. Soc., 91, 1015–1057, 2010.

Seemann, S. W., Borbas, E. E., Knuteson, R. O., Stephenson, G. R., and Huang, H.-L.: Development of a global infrared surface emissivity database for application to clear sky retrievals from multispectral satellite radiance measurements, J. Appl. Meteorol. Clim., 47, 108–123, 2008.

Simard, M., Pinto, N., Fisher, J., and Baccini, A.: Mapping forest canopy height globally with spaceborne lidar, J. Geophys. Res., 116, G04021, doi:10.1029/2011JG001708, 2011.

Stackhouse, P., Gupta, S., Cox, S., Mikovitz, J., Zhang, T., and Chiacchio, M.: 12-year surface radiation budget data set, GEWEX News, 14, 10–12, 2004.

Su, H., McCabe, M. F., and Wood, E. F.: Modeling evapotranspiration during SMACEX: comparing two approaches for local- and regional-scale prediction, J. Hydrometeorol., 6, 910–922, 2005.

Su, Z.: The Surface Energy Balance System (SEBS) for estimation of turbulent heat fluxes, Hydrol. Earth Syst. Sci., 6, 85–100, doi:10.5194/hess-6-85-2002, 2002.

Szczypta, C., Calvet, J.-C., Albergel, C., Balsamo, G., Boussetta, S., Carrer, D., Lafont, S., and Meurey, C.: Verification of the new ECMWF ERA-Interim reanalysis over France, Hydrol. Earth Syst. Sci., 15, 647–666, doi:10.5194/hess-15-647-2011, 2011.

Vinukollu, R. K., Sheffield, J., Wood, E. F., Bosilovich, M. G., and Mocko, D.: Multimodel analysis of energy and water fluxes: intercomparisons between operational analyses, a land surface model, and remote sensing, J. Hydrometeorol., 13, 3–26, 2011.

Wang, K. and Dickinson, R.: A review of global terrestrial evapotranspiration: Observation, modeling, climatology, and climatic variability, Rev. Geophys., 50, RG2005, doi:10.1029/2011RG000373, 2012.

Wang, K., Wan, Z., Wang, P., Sparrow, M., Liu, J., Zhou, X., and Haginoya, S.: Estimation of surface long wave radiation and broadband emissivity using Moderate Resolution Imaging Spectroradiometer (MODIS) land surface temperature/emissivity products, J. Geophys. Res., 110, D11109, doi:10.1029/2004JD005566, 2005.

Xie, P. and Arkin, P. A.: Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs, B. Am. Meteorol. Soc., 78, 2539–2558, 1997.

Zhang, K., Kimball, J. S., Nemani, R. R., and Running, S. W.: A continuous satellite-derived global record of land surface evapotranspiration from 1983 to 2006, Remote Sens. Environ., 46, W09522, doi:10.1029/2009WR008800, 2010.

Zhang, K., Kimball, J. S., Nemani, R. R., Running, S. W., Hong, Y., Gourley, J. J., and Yu, Z.: Vegetation Greening and Climate Change Promote Multidecadal Rises of Global Land Evapotranspiration, Sci. Rep., 5, 15956, doi:10.1038/srep15956, 2015.