Local tower-based merging of two land evaporation products

Carlos Jiménez[1,2], Brecht Martens[3], Diego M. Miralles[3], Joshua B. Fisher[4], Hylke E. Beck[5], and Diego Fernández-Prieto[6]

[1]Estellus, Paris, France
[2]LERMA, Paris Observatory, Paris, France
[3]Laboratory of Hydrology and Water Management, Ghent University, Ghent, Belgium
[4]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, USA
[5]Department of Civil and Environmental Engineering, Princeton University, Princeton, New Jersey, USA
[6]ESRIN, European Space Agency, Frascati, Italy

Correspondence to: Carlos Jiménez(carlos.jimenez@estellus.fr)

Abstract

An inverse-variance weighting of two terrestrial evaporation (ET) products from the WACMOS-ET project based on FLUXNET sites is presented. The two ET models, PT-JPL and GLEAM, share a common modeling framework based on the Priestley-Taylor evaporation formula-
tion, although the stress factors to convert from potential to actual evaporation are calculated differently. Both data sets are daily at 25 km resolution and use common input data when possible. The weights are based on the variance of the difference between the tower ET and the modeled ET, and are made dynamic by estimating them using a 61-day running window centered on each day. Seasonal variability in these weights is observed over
some stations, but the deviations from the 0.5 value assumed by the arithmetic mean of both products is typically small. Seasonal averaged statistics for three land cover groups confirm that the performance of the weighted product is very close to the arithmetic mean product. This implies that at a large number of stations the tower data was not informative enough to identify distinct error patterns between GLEAM and PT-JPL likely to result in
more varying weights. Different factors can be responsible for this behavior. Concerning the model estimates, some of the shared modeling assumptions and the common inputs make the ET estimates from both models rather dependent, thus their errors are expected to be correlated. Regarding the tower data, the 84 selected tower sites are mostly at temperate regions. In these regions the models ET differences are mainly reflecting the variability in
Priestley and Taylor estimates of potential ET, which translates in model ET differences smaller than in other places. The large mismatch between the model spatial resolution and the tower footprint is also likely to result again in common error patterns for the model estimates. Overall, this study suggests that merging tower observations and ET products at the time and spatial scales of this study is not straightforward, and that care should be taken
regarding the dependence of the products to be merged, the tower spatial representativeness of their measurements at the products resolution, and the nature of the error in both towers and gridded data sets.

1 Introduction

The land heat flux governs the interactions between the Earth and its atmosphere (Betts, 2009), is an essential component of the water, energy, and carbon cycles (Sorooshian et al., 2005), and thus plays a key role in the climate system (Wang and Dickinson, 2012). Terrestrial evaporation (ET) – the associated flux of water from land into the atmosphere – is also an important variable in the management of agricultural systems, forests, and hydrological resources. Hence, estimates of ET at different spatial scales, ranging from individual plants for managing irrigation, to basin scales to evaluate water resources, are required in many applications(e.g. Dunn and Mackay, 1995; Le Maitre and Versfeld, 1997; Gowda et al., 2008; Fisher et al., 2017).

In situ measurements of land heat fluxes are typically conducted during field experiments (Pauwels et al., 2008) and by more permanent flux tower networks (Baldocchi et al., 2001). Being point measurements and requiring special equipment, they cannot be applied for routine measurements covering large areas. Therefore, more readily available observations are combined with well known flux formulations (e.g., Monteith, 1965; Priestley and Taylor, 1972) to obtain local estimates. To derive global estimates, remote sensing from space can be used, but the challenge is that fluxes do not have a direct signature that can be remotely detected. Therefore, satellite remote sensing observations related to surface temperature, soil moisture, or vegetation are again combined with flux formulations to derive global estimates at different time and spatial scales (for overview see Wang and Dickinson, 2012; Zhang et al., 2016)

The combination of using different flux formulations driven by different global satellite data sets typically results in relatively large ET discrepancies, which are put in evidence when the ET products are inter-compared or evaluated with the flux networks (Jimenez et al., 2011; Mueller et al., 2011; McCabe et al., 2016). These differences are motivating efforts to derive in principle more accurate ET products by combining individual ET estimates at different time and spatial scales. These efforts range from simply averaging a number of ET products (Mueller et al., 2013) to more complex approaches, such as fusion algorithms

where the original ET products are combined to reproduce flux observations (Yao et al., 2017), or integration methodologies that seek consistency between ET products and related products of the water cycle (Aires, 2014; Munier and Pan, 2014).

Aiming at improving the predictive capability for ET, the WAter Cycle Multi-mission Observation Strategy – ET project (WACMOS-ET, *http://wacmoset.estellus.eu*) compiled a forcing data set covering the period 2005–2007, and ran four established ET models using common forcing (Michel et al., 2016; Miralles et al., 2011b). Based on a local (Michel et al., 2016) and global-scale (Miralles et al., 2011b) evaluation of these algorithms, two of the methodologies – the Priestley-Taylor Jet Propulsion Laboratory model (PT-JPL, Fisher et al., 2008), and the Global Land Evaporation Amsterdam Model (GLEAM, Miralles et al., 2011b) – were found to outperform the others. Nevertheless, substantial differences were found between both model products. As such, in this study, we investigate whether a weighted combination of GLEAM and PT-JPL estimates results in a more accurate ET product. Ideally, the weight assigned to each product during their merging should be based on an accurate description of the specific product uncertainties. However, even if some attempts to derive model uncertainty exist (Miralles et al., 2011a; Badgley et al., 2015; Loew et al., 2016), the complexity to derive estimates of ET from remote sensing data means that reliable uncertainty assessment is only attained through validation against tower flux measurements. Therefore, here we propose a flux tower-based weighting of GLEAM and PT-JPL and investigate the performance of the merged ET product over a selection of tower sites.

The paper is organized as follows: first, the ET models, the model input data sets, and the tower observations are described, followed by a description of the merging technique, and the metrics used in the analyses. The merged ET products are then presented for a selection of tower sites, and the results, and limitations of the exercise discussed. Finally, the main conclusions of the study are summarized.

2   Methods and data

2.1   ET models

The GLEAM and PT-JPL models, and the inputs required to run them at the global scale
are described in Michel et al. (2016) and Miralles et al. (2016). Only the main differences
with respect to the original WACMOS-ET runs are fully described here.

2.1.1   GLEAM

GLEAM is a simple land surface model fully dedicated to deriving evaporation. It distin-
guishes between direct soil evaporation, transpiration from short and tall vegetation, snow
sublimation, open-water evaporation, and interception loss from tall vegetation. Interception
loss is independently calculated based on the Gash (1979) analytical model forced by ob-
servations of precipitation. The remaining components of evaporation are based upon the
formulation by Priestley and Taylor (1972) for potential evaporation, constrained by multi-
plicative stress factors. For transpiration and soil evaporation, the stress factor is calculated
based on the content of water in vegetation (microwave vegetation optical depth) and the
root zone (multilayer soil model driven by observations of precipitation and updated through
assimilation of microwave surface soil moisture). For regions covered by ice and snow, sub-
limation is calculated using a Priestley and Taylor equation with specific parameters for ice
and supercooled waters. For the fraction of open water at each grid cell, the model assumes
potential evaporation.

The recent GLEAM v3 model of Martens et al. (2016) is adopted for this study and re-
places the model of Miralles et al. (2011b) previously applied for the WACMOS-ET runs.
Major differences related to the previous model are a revised formulation of the evaporative
stress, an optimized drainage algorithm, and a new soil moisture data assimilation system.
Notice that the WACMOS-ET runs were done at different time resolutions, while only daily
estimates are calculated for this study.

2.1.2   PT-JPL

The PT-JPL model by Fisher et al. (2008) is a relatively simple algorithm to derive ET. It uses the Priestley and Taylor (1972) approach to estimate potential evaporation, and then applies a series of stress factors to reduce from potential to actual evaporation. The land evaporation is partitioned first into soil evaporation, transpiration, and interception loss by distributing the net radiation to the soil and vegetation components. The potential evaporation for soil, transpiration, and interception is then calculated separately, followed by a reduction to actual evaporation by applying a series of ecophysiological stress factors. Unlike GLEAM, the stress factors are based on atmospheric moisture (vapour pressure deficit and relative humidity) and vegetation indices (normalized difference vegetation index, and soil adjusted vegetation index) to constrain the atmospheric demand for water. The partitioning between transpiration and interception loss is done using a threshold based on relative humidity, and therefore conceptually quite different from the precipitation based calculation of GLEAM. There is no independent estimation of snow sublimation, and the same algorithms are applied for snow-covered areas.

For this study, optimized vegetation products are used as inputs to the model. In WACMOS-ET, the Leaf Area Index (LAI) and Fraction of Absorbed Photosynthetic Active Radiation (FAPAR) products, derived from the Joint Research Centre Two-Stream Inversion (JRC-TIP) package (Pinty et al., 2007, 2011a, b), were converted by a simple biome-dependent calibration to a LAI/FAPAR product consistent with the Moderate Resolution Imaging Spectroradiometer (MODIS) LAI/FAPAR before being used as inputs to the model (Michel et al., 2016). Under the assumptions that the JRC-TIP FAPAR is related to the radiation absorption by the green fraction of the canopy, while the MODIS FAPAR is more related to green and non-green leaf area, a new use of the WACMOS-ET vegetation products is proposed. First, the WACMOS-ET JRC-TIP FAPAR is assumed to be close to an Enhanced Vegetation Index (EVI), and it is scaled by the factor 1.2 to become closer to the FAPAR expected by the model, as in the original PT-JPL equations. Second, the WACMOS-ET MODIS-like FAPAR is used as the Fraction of Intercepted Photosynthetic Active Radiation (FIPAR) ex-

pected by the model, which in turn is used by the model as a proxy for the fractional total vegetation cover. Using the original relationships in the model, the fractional total vegetation cover is related to a total (green and non-green) LAI, which is then used to partition the net radiation in to their soil and canopy components.

2.2 Data

2.2.1 Model inputs

The GLEAM and PT-JPL required global inputs remain unchanged with respect to Miralles et al. (2016), apart from the precipitation product, and are applied at the same resolution of 25 km. Common inputs to GLEAM and PT-JPL are the surface net radiation, coming from the NASA and GEWEX Surface Radiation Budget (SRB, Release 3.1 Stackhouse et al., 2004), and the near-surface air temperature, sourced from the ERA-Interim atmospheric reanalysis (Dee et al., 2011). PT-JPL also requires near-surface air humidity, also from ERA-Interim, and the vegetation products discussed in Section 2.1.2. On the other hand, GLEAM requires precipitation, coming from the Multi-Source Weighted-Ensemble Precipitation (MSWEP) version 1 product (Beck et al., 2017), soil moisture and vegetation optical depth from the ESA Climate Change Initiative (CCI) Soil Moisture v2.3 product (Liu et al., 2011b, a), and information on snow water equivalents, from the ESA GlobSnow product for the Northern Hemisphere (Takala et al., 2011)(Luojus and Pulliainen, 2010), and from the National Snow and Ice Data Center (NSIDC) in snow-covered regions of the Southern Hemisphere (Kelly et al., 2003).

2.2.2 Tower data

The FLUXNET 2015 synthesis data set (http://fluxnet.fluxdata.org/) is used to obtain *in situ* measurements of evaporation (referred to as tower ET), and it is processed as in Martens et al. (2016) to retain only high-quality data appropriate to evaluate the evaporation models. Starting from the original time resolution (generally 30 minutes or 1 hour), the processing involves: (1) masking measurements using the provided quality flags; (2) masking measure-

ments for rainy intervals, using the global precipitation product and the local measurements if available (eddy-covariance measurements are generally less reliable during precipitation events), and; (3) aggregating to daily values if more than 75 percent of remaining sub-hourly data exists for a given day. This processing selects 84 station for the 2002-2007 study period. Their geographical locations, and their location in an air temperature and precipitation space, are plotted in Fig. 1, with the station names, land covers and reference or Principal Investigator listed in Table 1.

Eddy-covariance measurements are subject to errors, both random and systematic. Systematic errors can arise from instrumental calibration and unmet assumptions about the meteorological conditions, while random errors are typically related to turbulence sampling errors, the assumptions of a constant footprint area, and instrumental limitations (Moncrieff et al., 1996). Estimating these errors is far from simple, and typically requires dedicated experiments (Nordbo et al., 2012; Post et al., 2015; Wang et al., 2015). Therefore reporting them is not a widespread practice and figures for the individual sites are not commonly available.

The propagation of systematic errors typically results in the lack of energy balance closure observed at many eddy-covariance sites. Techniques to correct this exist (Foken et al., 2006), and corrected fluxes are typically preferred over the original uncorrected observations. However, correcting for the energy balance closure normally implies the need for surface radiation measurements, which are not routinely measured at all stations; e.g., for the 84 stations selected here, 26 do not have corrected latent heat fluxes. For the 58 stations with corrected fluxes, the relative mean difference between raw and corrected fluxes averaged over all stations is 6.1%, with a maximum value of 16.5%. The uncorrected and corrected fluxes correlate very well in time, with a station-averaged square temporal correlation of 0.96. As these numbers do not suggest strong differences between the corrected and uncorrected measurements at our selected stations, we use the original uncorrected latent heat fluxes to avoid mixing uncorrected and corrected fluxes in the analyses, and to keep a sufficiently large database of *in situ* measurements.

Table 1. List of the FLUXNET sites used in this study together with their FLUXNET code (ID), IGBP land cover (LC) and official reference or principal investigator (PI). The CA-NS1-7 refers to seven stations closely located and run by the same group.

ID	LC	Reference/PI	ID	LC	Reference/PI	ID	LC	Reference/PI
AT-Neu	GRA	George Wohlfahrt	AU-How	SAV	Jason Beringer	BE-Bra	MF	Ivan Janssens
BE-Bra	MF	Ivan Janssens	BE-Lon	CRO	Moureaux et al. (2006)	BE-Vie	MF	Aubinet et al. (2001)
BR-Sa3	EBF	Steininger (2004)	CA-Gro	MF	McCaughey et al. (2006)	CA-Man	ENF	Dunn et al. (2007)
CA-NS1-7	ENF	B.Lamberty et al. (2004)	CA-Oas	MF	Bond-Lamberty et al. (2004)	CA-Obs	ENF	Bond-Lamberty et al. (2004)
CA-Qfo	ENF	Bergeron et al. (2007)	CA-SF1	ENF	Coursolle et al. (2012)	CA-SF2	MF	Amiro et al. (2006)
CH-Dav	ENF	Lukas Hoertnagl	CH-Fru	GRA	Zeeman et al. (2010)	CH-Oe1	GRA	Christof Ammann
CH-Oe2	CRO	Christof Ammann	CN-Cha	MF	Shijie Han	CN-Dan	GRA	Shi Peili
CN-Din	EBF	Guoyi Zhou	CN-Du2	GRA	Chen Shiping	CN-Ha2	WET	Yingnian Li
CN-HaM	GRA	Kato et al. (2006)	CN-Qia	ENF	Huimin Wang	CZ-BK1	ENF	Marian Pavelka
DE-Geb	CRO	Antje Moffat	DE-Gri	GRA	Christian Bernhofer	DE-Hai	DBF	Knohl et al. (2003)
DE-Kli	CRO	Christian Bernhofer	DE-Tha	ENF	Christian Bernhofer	DE-Lnf	DBF	Alexander Knohl
DK-Sor	DBF	Andreas Ibrom	ES-Lju	CSH	Penelope Serrano	FI-Hyy	ENF	Timo Vesala
FR-Fon	DBF	Bazot et al. (2013)	FR-Gri	CRO	Pierre Cellier	FR-LBr	CRO	Denis Loustau
FR-Pu	MF	Jean-Marc Ourcival	IT-Col	DBF	Giorgio Matteucci	IT-Lav	ENF	Damiano Gianelle
IT-MBo	GRA	Damiano Gianelle	IT-PT1	DBF	Günther Seufert	IT-Ren	ENF	Stefano Minerbi
IT-Ro1	CRO	Nicola Arriga	IT-Ro2	DBF	Nicola Arriga	JP-SMF	CRO	Ayumi Kotani
MY-PSO	EBF	Yoshiko Kosugi	NL-Loo	ENF	Eddy Moors	RU-CHE	OSH	Corradi et al. (2005)
RU-Fyo	ENF	Milyukova et al. (2002)	RU-Ha1	GRA	Dario Papale			
US-ARM	CRO	Fischer et al. (2007)	US-ARb	GRA	Margaret Torn	US-ARc	GRA	Margaret Torn
US-Blo	ENF	Goldstein et al. (2000)	US-Cop	GRA	David Bowling	US-IB2	CRO	Roser Mantamala
US-Goo	GRA	Tilden Meyers	US-Ha1	DBF	Goulden et al. (1996)	US-Los	MF	Ankur Desai
US-Ivo	WET	McEwing et al. (2015)	US-MMS	DBF	Schmid et al. (2000)	US-Me2	ENF	Campbell and Law (2005)
US-Me3	ENF	Bond-Lamberty et al. (2004)	US-Ne1	CRO	Simbahan et al. (2006)	US-Ne2	CRO	Amos et al. (2005)
US-Ne3	CRO	Verma et al. (2005)	US-Oho	DBF	Noormets et al. (2008)	US-PFa	MF	Richardson et al. (2006)
US-SRM	WSA	Scott et al. (2009)	US-Syv	MF	Ankur Desai	US-Ton	WSA	Chen et al. (2007)
US-Var	GRA	Ma et al. (2007)	US-WCr	DBF	Cook et al. (2004)	US-Wi3	DBF	Jiquan Chen
US-Wi4	MF	Jiquan Chen	US-Wi9	MF	Jiquan Chen			
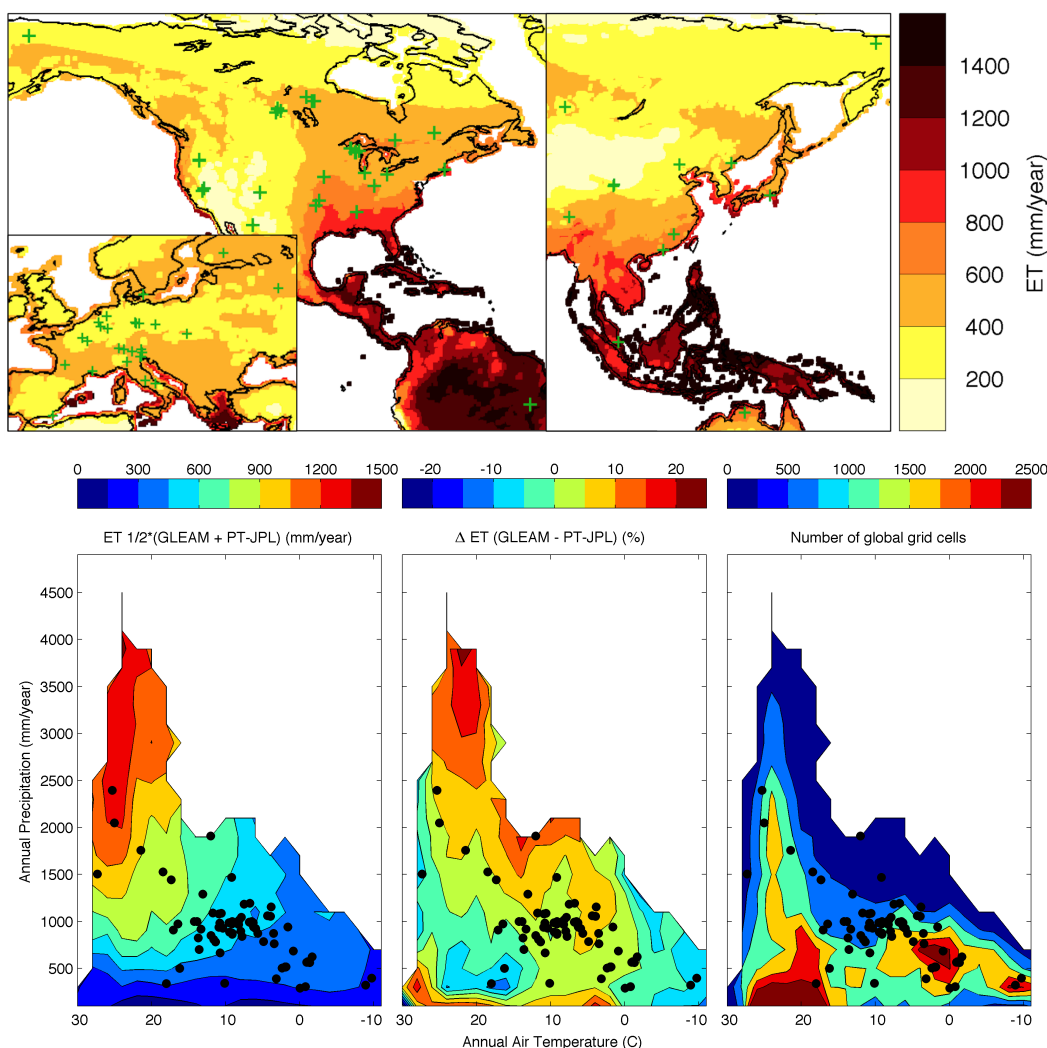
Figure 1. Distribution of tower sites used in the study. Top: geographical location (green crosses) on a map of GLEAM and PT-JPL averaged multi-annual ET. Bottom: distribution of the GLEAM and PT-JPL averaged multi-annual ET (left), the relative GLEAM and PT-JPL ET differences normalized by the previous averaged ET (middle), and the number of global grid cells (right), as function of the annual air temperature and precipitation, together with the location of the tower sites in this space (black dots).

2.2.3 Ancillary data

To help characterizing the spatial homogeneity of the grid cells where the stations are located, two data sets are considered: the MODIS Land Cover Type product MCD12Q1 at an original resolution of 500 meters, and the Terra MODIS Vegetation Continuous Fields product MOD44B at an original resolution of 250 meters. A homogeneity index (I_h) is constructed as:

$$I_h = \frac{1}{2}Fgt_{IGBP} + \frac{1}{2}(1 - \mid Fg_{bare} - Ft_{bare} \mid - \mid Fg_{herb} - Ft_{herb} \mid - \mid Fg_{forest} - Ft_{forest} \mid) \quad (1)$$

where Fgt_{IGBP} is the fraction of MCD12Q1 500 meter cells included in the 25 km model grid cell containing the tower and having the same IGBP land cover than the model cell, Ft_{bare}, Ft_{herb} and Ft_{forest} are, respectively, the bare, herbaceous, and forest fractions of the MOD44B 250 meter cell containing the tower, and Fg_{bare}, Fg_{herb} and Fg_{forest} are the same fractions but calculated for the entire 25 km model grid cell where the tower is situated. Ih takes values in the range [0,1], the larger the value the more likely the grid cell represents the landscape of the tower pixel, according to these two MODIS products.

2.3 Merging technique

A weighted combination of two products requires the definition of a set of weights, typically based on an estimation of the individual uncertainty in each of the products. The simplest strategy is to assume that both products are equally uncertain: the merged product is a Simple Average (SA-merge) of the individual products. A more elaborate strategy would be first estimating the product uncertainties, followed by a weighting of the products that takes into account this uncertainty.

The inverse-variance weighting is the usual combination equation to take into account individual product uncertainties and obtain a merged estimate bounded by the initial estimates. In the context of our analysis it can be expressed as:

$$E_{WA} = w_{GL}E_{GL} + w_{PT}E_{PT} \qquad (2)$$

$$w_{GL} = \frac{\sigma_{PT}^2}{\sigma_{GL}^2 + \sigma_{PT}^2} \qquad (3)$$

$$w_{PT} = \frac{\sigma_{GL}^2}{\sigma_{GL}^2 + \sigma_{PT}^2} \qquad (4)$$

where E_{WA} is the Weigthed Average merged product (WA-merge), E_{GL} (E_{PT}) is GLEAM (PT-JPL) ET, w_{GL} (w_{PT}) is the GLEAM (PT-JPL) weight, and σ_{GL}^2 (σ_{PT}^2) is the variance of the GLEAM (PT-JPL) error distribution. For the SA-merge, $w_{GL}=w_{PT}=0.5$. For the WA-merge, the error is defined as the difference between the model and tower-based ET, the σ_{GL} and σ_{PT} are then estimated over the time series of available errors, and the weights derived following Equations 3 and 4. Note that $w_{GL} + w_{PT} = 1$.

If the errors follow a Gaussian distribution, are unbiased, and are independent from each other, Equation 2 is the optimal linear estimator (Rodgers, 2000). However, in practice, those conditions are difficult to meet. While GLEAM and PT-JPL can be considered independent from the tower observations, they do share common inputs and formulations, likely making the errors in their estimates dependent. Bias can also be present between the satellite surface meteorology used by the evaporation models and the real meteorology at the tower, due to for instance differences in footprint. The different stress formulations can also introduce systematic errors, and consequently biases. Therefore, for this exercise the inverse-variance weighting can be seen as a simple error-based method to weigh the products, but optimality in the sense of minimizing the error variance cannot be assured.

At a given tower location, the error variance is estimated as the variance of the errors over a certain period. If estimated over the entire record, there will be one weight per station, with no seasonal variation. However, it is expected that the errors are non-stationary, hence, in order to have weights evolve in time, they will be estimated over time series of a certain number of days centered at each day of the year. The weights are then estimated daily, by running a time window centered at that day. The choice of window length is subjec-

tive: shorter time windows produce more dynamic weights, but their values are likely to be noisier given the smaller number of samples available to estimate the time series variability. A number of 30 days before and after each calendar day was found to provide a good compromise between the smoothness of weights and the number of samples required, so a 61-day running window is used to derive the daily weights.

GLEAM and PT-JPL estimate separately transpiration, soil evaporation, and the evaporation from the rain intercepted by canopies. As the tower measurements are masked for rainy intervals, only the sum of soil evaporation and transpiration is compared with the tower data and weighted. If the total ET merged product is required, the choices are either to assume that GLEAM and PT-JPL are equally uncertain in their estimation of interception loss, adding their average to the weighted soil evaporation and transpiration, or adding the GLEAM or PT-JPL individual interception if there are reasons to believe that one is overall more certain than the other.

2.4 Metrics

The main analyses are done by computing the Pearson correlation coefficient (R), the Mean Square Difference (MSD), and the Root Mean Square Difference (RMSD) according to the the expressions:

$$R = \frac{N \sum_{i=1}^{N} P_i O_i - \sum_{i=1}^{N} P_i \sum_{i=1}^{N} O_i}{\sqrt{[N \sum_{i=1}^{N} P_i^2 - (\sum_{i=1}^{N} P_i)^2]} \sqrt{N \sum_{i=1}^{N} O_i^2 - (\sum_{i=1}^{N} O_i)^2}} \quad (5)$$

$$MSD = [\frac{1}{N} \sum_{i=1}^{N} (P_i - O_i)^2] = RMSD^2 \quad (6)$$

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

where P and O are the model-derived and observed (or a second model-derived) variate, and N is the number of cases. The MSD can be decomposed into a random (MSD$_r$) and systematic (MSD$_s$) component following (Willmott, 1982) by using the expressions:

$$\text{MSD}_r = \frac{1}{N}\sum_{i=1}^{N}(\hat{P}_i - O_i)^2 = \text{RMSD}_r{}^2 \tag{7}$$

$$\text{MSD}_s = \frac{1}{N}\sum_{i=1}^{N}(P_i - \hat{P}_i)^2 = \text{RMSD}_s{}^2 \tag{8}$$

where $\hat{P}_i = a + bO_i$ is the linear least squares regression of P onto O, being a and b the regression intercept and slope, respectively. Notice that MSD = MSD$_r$ + MSD$_s$.

When improving a product and comparing with a reference, the common targets are correlation unity and zero RMSD. But as the merge product here is bounded by the original GLEAM and PT-JPL ET estimates, and the tower ET is not always within that ET range, this target cannot always be attained. A possibility to set correlation and RMSD targets is to construct the product that minimizes the RMSD with the tower ET. In the context of our merging strategy, we call this the optimum product . As normalized weights are used, the weighted estimate can be one of the original estimates (if it is weighted 1), or an intermediate value between the products (for weights between 0 and 1). In our case, if the tower ET is between the GLEAM and PT-JPL estimates, the tower ET is then the optimum estimate. But if both the GLEAM and PT-JPL estimates are below or above the tower ET, the closest model ET is the optimum estimate, and the merging cannot reproduce the tower ET observation. Note that the optimum product does not necessarily result in the product combination having the largest correlation with the tower E, although in most cases it will improve these correlations.

Statistical significance of the correlations is tested by calculating 95% confidence intervals. For the correlation differences, a Fisher Z-transformation is applied to the correlations,

and a Student t-test at a 5% significance level used to test the significance of the difference. The autocorrelation of the daily time series is taken into account by reducing the degrees of freedom using an effective sampling size (De Lannoy and Reichle, 2016; Lievens et al., 2017).

Statistics are calculated for the whole period, or separately for the boreal winter (DJF), spring (MAM), summer (JJA), and autumn (SON). Given the strong seasonality at most towers, correlations tend to be high without necessarily indicating that product and tower ET day-to-day anomalies are in close agreement. Calculating correlations after removing the mean seasonal cycle allows the study of short-term ET anomalies, but here the limited data record at most towers precludes the calculation of a robust seasonal cycle.

3 Results and discussion

3.1 Inter-product comparison

The annual GLEAM and PT-JPL ET components, together with their differences, are shown in Fig. 2. Note that for GLEAM sublimation is included in the soil evaporation. Approximately 60% of the relative differences (differences normalized by the absolute value) are within ±25%. As both models are run with a common surface radiation product, and with a similar calculation of the potential ET, this suggests that there are large differences in their respective stress factors at some places. The disagreement in the partitioning of ET into its different components can also be noticed. PT-JPL estimates transpiration to be about 1/2 of the total ET, while GLEAM estimates that around 3/4 of the total ET is transpiration. The remaining soil evaporation and interception are equally distributed for GLEAM (13% and 14%, respectively), while for PT-JPL the soil evaporation contribution is larger than the interception (28% and 20 %, respectively). Compared to Miralles et al. (2016), the GLEAM and PT-JPL contributions of each component to the total ET are comparable, with the small differences attributable to the above-described changes in the models and/or inputs (e.g., new precipitation product for GLEAM, or the revised use of the LAI/FAPAR for PT-JPL). Re-

call that, as discussed in Section 2.4, only the sum of the soil evaporation and transpiration is validated against tower fluxes.
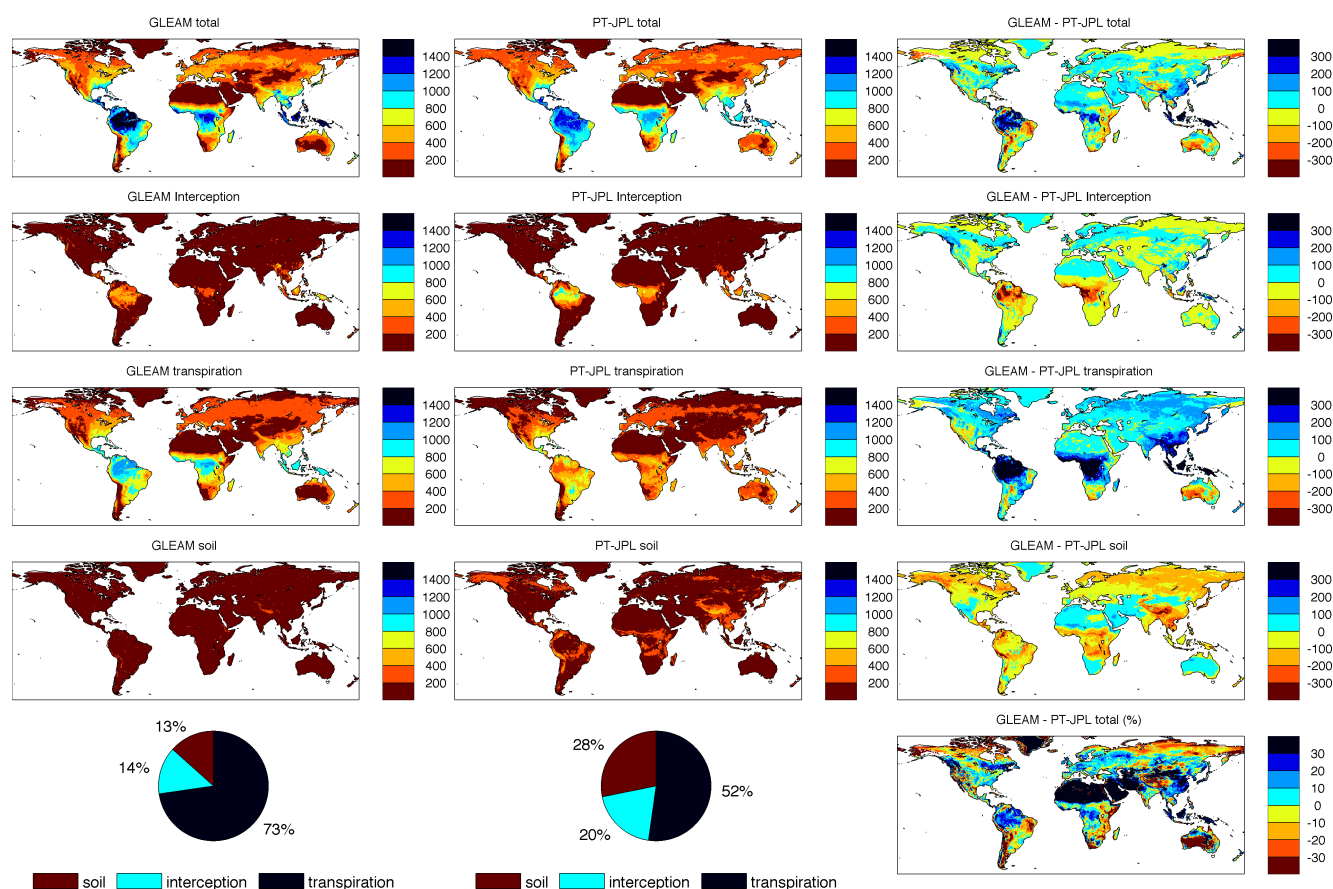
Figure 2. Summary of GLEAM and PT-JPL annual ET differences. The GLEAM (left) and PT-JPL (middle) total annual ET in mm/year, and their difference (right) is plotted in the first row. Same plots for the interception, transpiration, and soil evaporation are given in the next three rows. The last row shows the annual percentage of the ET components for GLEAM (left) and PT-JPL (middle), and the total GLEAM and PT relative annual ET differences, normalized with the GLEAM and PT-JPL averaged ET, and expressed as percentage (left).
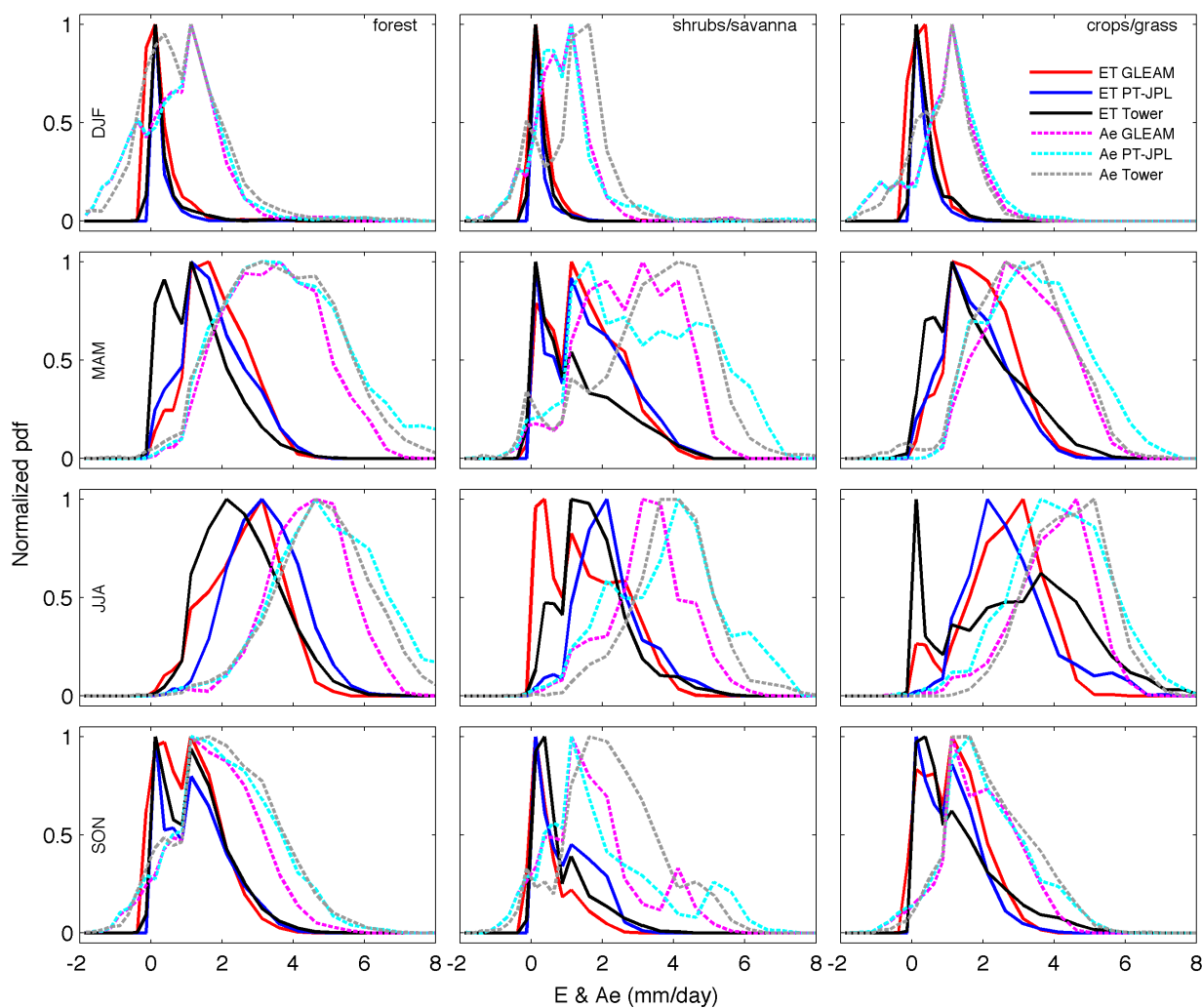
Figure 3. Normalized histograms of ET and available energy (Ae) from GLEAM, PT-JPL, and the tower observations. The histograms are calculated with the ET values at the tower locations separated first by season and land cover.

The GLEAM, PT-JPL, and tower ET are compared now at the available tower sites. The towers spatial distribution of Fig. 1 shows that approximately 75% of the towers are located in grid cells where the relative ET difference is within $\pm$ 5%, i.e., at places where the GLEAM and PT-JPL different ET stress formulations do not result in large annual ET differences. Most of those towers are located in temperate regions. The tropical rain forest and savannas, where the relative ET differences seem larger, are poorly represented in the selected tower data. Therefore, regions that would have been relevant to characterize GLEAM and PT-JPL differences are missing in the evaluation with tower data.

Seasonal distributions of ET for three vegetation classes are presented in Fig. 3. The first one groups the International Geosphere-Biosphere Programme (IGBP) forest cover stations, the second one includes the shrublands and savannas, and the third one the croplands and grasslands. They are referred to as "forest", "shrubs/savanna", and "crop/grass", respectively. Note that the stations are not evenly distributed within the three groups, with the forest group (50 stations) being more represented than the shrubs/savanna and crops/grass (10 and 24, respectively), indicating that group statistics could be more significant for forests.The surface Available Energy (Ae) is also plotted. For GLEAM and PT-JPL, Ae is the difference between the surface net radiation and the modeled ground flux. For the towers, as the surface net radiation and/or ground flux are not measured at all towers, Ae is given by the sum of the sensible and latent fluxes. Differences between GLEAM and PT-JPL distributions are visible, but overall GLEAM and PT-JPL agree better with each other than when compared with the tower ET. This is expected given the dependencies between both gridded data sets and their common footprint.

An example of good agreement is the forest group for the SON months, with the distributions of ET and Ae being quite similar for the *in situ* and modeled variables. The crops/grass group for the JJA months shows also reasonable agreement between GLEAM and PT-JPL ET distributions, but large differences with the tower ET. The tower ET has a clear bimodal distribution, which cannot be replicated by GLEAM and PT-JPL. This may be due to agricultural management practices being poorly captured by the models (e.g., irrigation), but may also reflect the large heterogeneity of croplands and their (a priori) low representa-

tiveness of the larger pixel scale. For the shrubs/savanna group and the JJA months, the three ET distributions are quite different, with GLEAM showing lower values than PT-JPL. For GLEAM, the mode of the Ae distribution is lower than for PT-JPL; as the surface net radiation is the same for GLEAM and PT-JPL, this implies that in this case GLEAM estimates larger ground fluxes than PT-JPL. The smaller GLEAM Ae can be partly responsible for some of the ET differences; however, the relatively large ET differences indicate differences in evaporative stress.
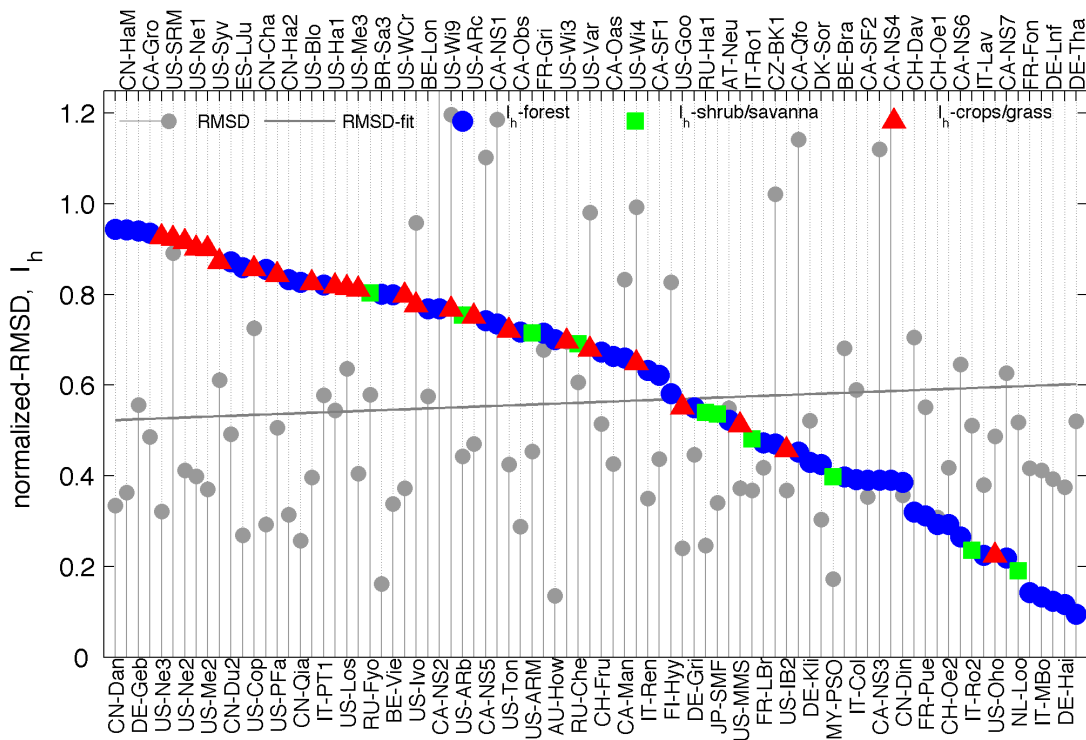
Figure 4. Homogeneity index (I_h, blue for forest stations, green for shrubs/savanna, and red for crops/grass) and RMSD of the optimum product and the towers ET, normalized by the mean annual tower ET (grey dots, with a linear fit plotted by a grey line). The towers are sorted from maximum to minimum I_h.

The very large mismatch between the model grid cells and the footprint of the tower measurements contributes to the observed differences. In order to see whether there is a clear relation between spatial homogeneity and model and tower differences, the optimum product defined in Section 2.4 and the the homogeneity index (I_h) described in Section 2.2.3 are computed for each station. The RMSD of the optimum product and the towers ET, normalized by the mean annual tower ET, is displayed in Figure 4 for all the available stations, together with the station I_h. The towers are sorted from maximum to minimum I_h, i.e., starting by the towers better representing the grid cells where the tower is located. Small and large normalized RMSD can occur at stations with comparable I_h, suggesting that spatial heterogeneity is only one of the contributing factors to the ET differences. If a linear square fit of the normalized RMSD of the sorted stations is calculated, the slope of the fit is ~0.09 mm/day, but the resulting trend is not statistically significant. Significant trends were neither found for the RMSD of the tower E and the original GLEAM and PT-JPL E. This indicates that for the constructed I_h and the stations and ET products sampled, the error related to the tower surrounding spatial homogeneity does not dominate the error budget.

3.2 Product weighting

A summary of daily weight statistics over all the sites belonging to a given land cover group is given in Fig. 5. The weights are given as the difference to 0.5, and will be referred to as positive weight deviations when the values are larger than 0.5. For instance, a 0.2 GLEAM weight deviation in the plot corresponds to a 0.7 and 0.3 absolute weights for GLEAM and PT-JPL, respectively. In this case GLEAM is positively weighted and PT-JPL negatively weighted. Likewise, a 0.2 PT-JPL weight deviation corresponds to 0.3 and 0.7 absolute weights for GLEAM and PT-JPL.

On average, the weights do not deviate much from 0.5, suggesting that, for a given land cover group, there are no systematic patterns indicating that one model agrees much better than the other when comparing with the tower data. The 25% and 75% percentiles are confined to the ±0.1 range on most days, so the variation of the weights can be considered small. Noticeable is some tendency for PT-JPL to be weighted more in winter, with GLEAM

more weighted the remaining months. Given the mostly high latitude location of the towers, the winter months correspond to situations with low or even negative surface radiation, the surface can be covered by snow, and ET values are in general low. Therefore, the weighting for these months will not change much the integrated annual values. On the contrary, the months where GLEAM seems more weighted typically correspond to larger ET values, and the corrections in the merged product for those periods will have a more visible impact on the resulting ET.
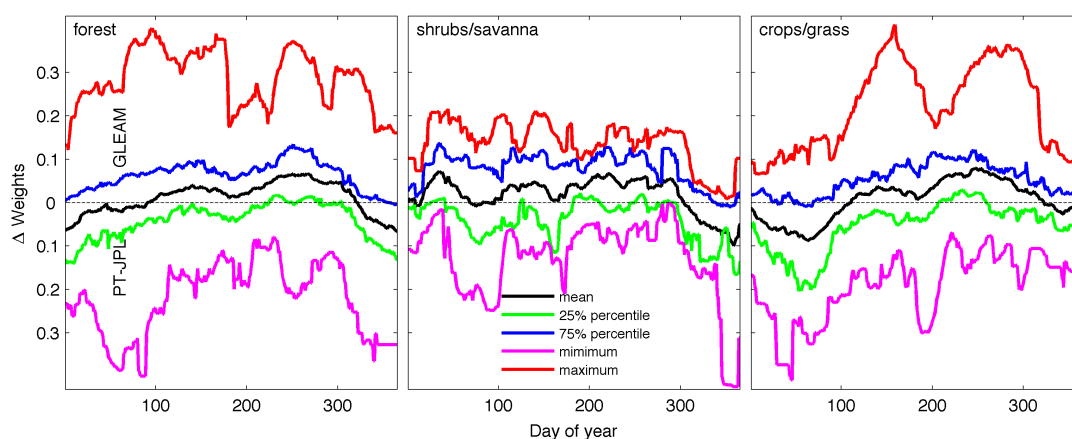
Figure 5. Daily statistics of weights over all forest (left), shrub/savanna (middle), and crop/grass (right) sites, plotted as the difference of the absolute weights and 0.5 (referred to as weight deviations, Δweight). A GLEAM Δweight of a means that GLEAM ET is weighted $0.5 + a$, and PT-JPL ET $0.5 - a$, while a PT-JPL Δweight of b means that GLEAM ET is weighted $0.5 - b$, and PT-JPL ET $0.5 + b$. Displayed are the mean, the 25% and 75% percentile, and the minimum and maximum value.

At a few sites the temporal variation of the weights is relatively large compared with the rest of the stations. At these stations GLEAM and PT-JPL behave more distinctly when compared with the tower ET, resulting in more different error variances, which translates into larger weight deviations. Examples are given in Fig. 6. At the CA-Gro site, a deciduous needleleaf forest located in Canada, PT-JPL is more weighted the first half of the year, while

more weight is attributed to GLEAM in the second half. JP-SMF, a mixed forest located in Japan, and US-Ne1, a cropland site situated in North America, show smaller and more constant weight deviations, with higher weights for GLEAM at JP-SMF, and lower at US-Ne1. US-SRM, a semi-arid grassland site, and CN-Ha2, a tundra shrubland site, show GLEAM being more weighed but especially during the first and last part of the year at CN-Ha2, and the middle of the year at US-SRM. DE-Kli, a cropland site located in Germany, show relatively close weight deviations for GLEAM and PT-JPL, with only GLEAM being slightly more weighed in spring.
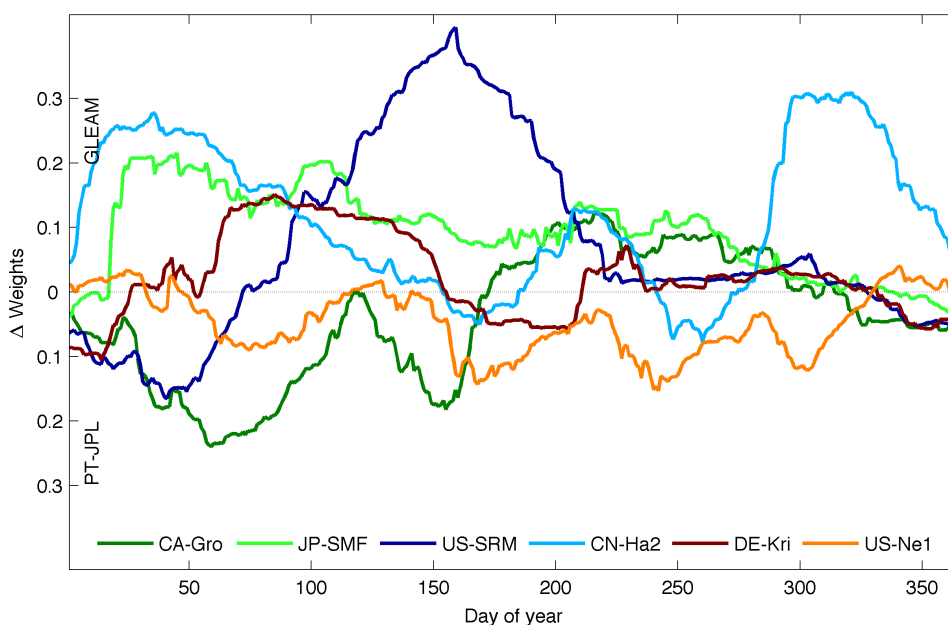
Figure 6. Example of ET Δweights at six selected stations (Δweights defined as in Figure 5).

Time series of original and WA-merged products for two of the previous stations are plotted in Fig. 7. At the semi-arid grassland US-SRM site there is a relatively large inter-annual variability, and GLEAM and PT-JPL do not agree very well with the tower ET for

three of the four years displayed (correlations of 0.31 and 0.24 for GLEAM and PT-JPL, respectively). For the first 3 years, GLEAM captures better the spring ET decrease seen by the tower, but both GLEAM and PT-JPL miss the large summer tower ET values. In this case, the larger weighting for GLEAM results in a weighted product that seems closer to the observed values, although the 0.37 correlation of the SA-merged product is not significantly higher than the 0.31 correlation of the WA-merged product. The US-Ne1 is a maize irrigated site, with expected more regular seasonal cycle and larger ET values than US-SRM. In this case PT-JPL agrees better with the observed ET than GLEAM (correlations of 0.93 and 0.84, respectively), capturing better the ET rise associated with start of the growing season, and the amplitude of the cycle for two out of the four years. PT-JPL is slightly more weighted than GLEAM in the WA-merged product, but the 0.93 correlation of the WA-merged product is similar to the PT-JPL correlation.
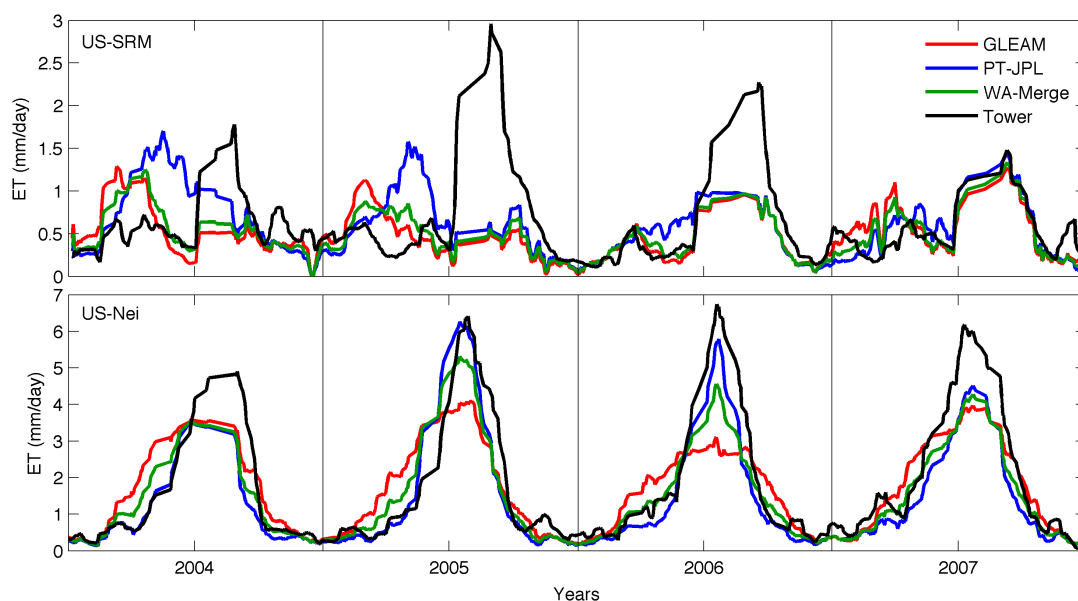
Figure 7. ET time series at the sites US-SRM (top) and US-Nei (bottom). The daily values are time smoothed with a 10-day moving averaged window to better display the more persistent temporal features.
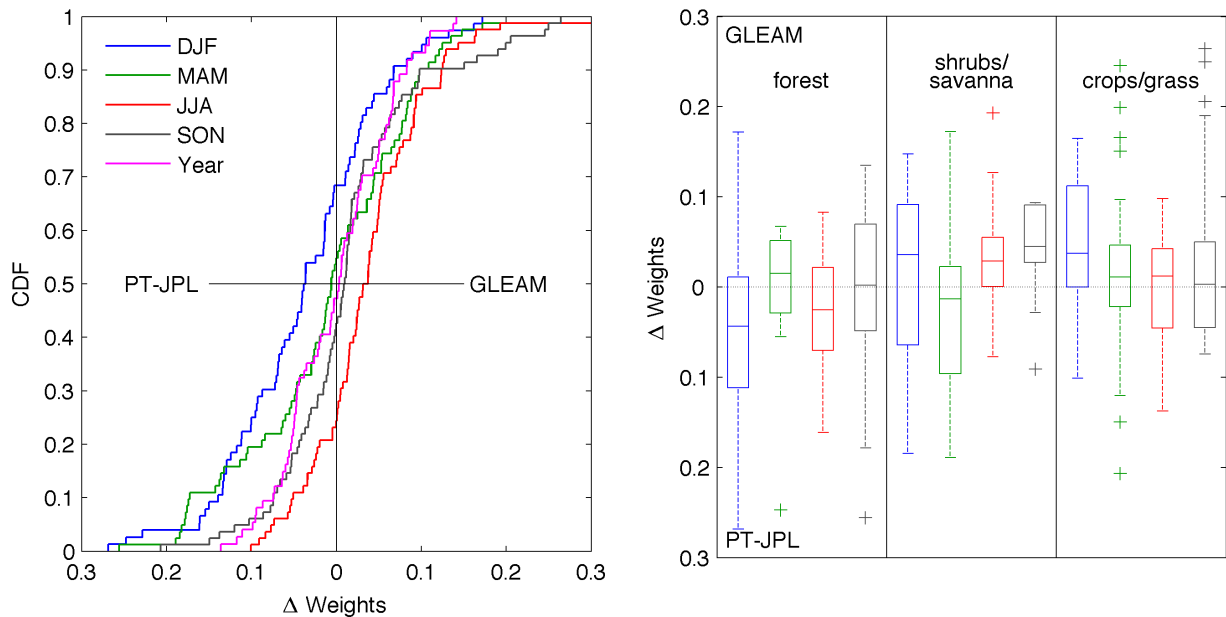
Figure 8. Left: empirical Cumulative Distribution Function (CDF) of the seasonal and annual averaged weights over all stations. Right: box plots of the seasonal averaged Δweights for the three land cover groups (Δweights defined as in Figure 5). The central mark of the box plots is the median of the group population, the box edges are the 25th ($Q1$) and 75th ($Q3$) percentiles, the whiskers extend to $Q3 + 1.5(Q3 - Q1)$ and $Q1 - 1.5(Q3 - Q1)$, and values outside the whisker are plotted individually.

More statistics of the weights are given in Fig. 8. The left-hand side panel shows the empirical Cumulative Distribution Function (CDF) of the seasonal averaged weight deviations over all stations. The differences are small, but PT-JPL has a larger number of positive weight deviations than GLEAM in winter and the opposite during summer and autumn, similar to the pattern shown in Fig. 5. The right-hand side panel displays box plots of the seasonal averaged weight deviations for the three land cover groups. The median values per season and group are well within the 0.1 range, so on average the weight deviations from the SA-merge value of 0.5 are relatively small, highlighting again the similarities between

both products. For the forest and shrubs/savanna groups the seasonal weight deviation medians can be positive for GLEAM or PT-JPL, depending on the season, while for the crops/grass group all seasonal weight deviation medians are positive for GLEAM. This may be an indication of GLEAM performing better than PT-JPL for croplands and grasslands. However, it is difficult to generalize any conclusions to the global scale given the small number of stations and their limited geographical distribution.

3.3 Merged products

The performance of the individual and merged products across the different stations is summarized by plotting seasonal group averaged correlations with the tower ET and RMSDs in Fig. 9. Given the typical small weight variations presented in Sec. 3.2, the differences in performance between the individual products are expected to be small. In order to better show these differences, Fig. 9 shows the difference of the season and land-cover averaged correlations of the optimum product ET with the correlations of GLEAM, PT-JPL, SA and WA-merge ET, expressed as a percentage of the optimum correlation. The same differences are also plotted for the RMSD, but taking the differences between the individual product RMSD and the optimum product RMSD, so the differences are also positive as for the correlations. The smaller the differences, the closer to the optimum product correlation and RMSD. Note that correlations are not significant for some stations and periods, although all correlations are averaged in Fig. 9 to have a common number of stations for the inter-product and inter-season comparisons. No significant correlations are observed at a larger number of stations in winter. For the other seasons, only a few stations do not show significant correlations, and all of them correspond to low correlation values. For the merged products, they include the two tropical stations MY-Pso and BE-Bra, with reduced seasonality and short records after removal of the rainy events, US-Var and CN-Din, at the foothills of some mountainous ranges, and US-Wi4, a red pine site with some wetlands in the surroundings of the station.

In terms of correlations the worst season is winter, reflecting the low intra-seasonal variability in this period, while the largest correlations are observed in summer. The smallest

relative differences in correlation with the optimum product are for spring and autumn, with values of around 10%. This indicates that the margin of improvement with respect to the optimum product is not very large for these seasons. The relative differences are larger for winter and summer, with values around 20%. Regarding the three land cover groups, the average correlations for the optimum product are larger for the forest and crops/grass, and a bit lower for the shrubs/savanna. In most cases the smaller correlation and RMSD differences are for the WA-merge product, with the merged products equaling the performance of the best individual product. However, running Student-t tests shows that the correlation differences between any pair of products are not statistically significant at the selected 95% confidence level. So, although GLEAM and PT-JPL ET disagree, their ET populations do not seem too distinct when compared with the tower ET.

Close GLEAM and PT-JPL weighting means close variance of their ET differences with the tower. This can be seen in Fig. 10, where box plots of the variance of the tower ET, the variance of the GLEAM and PT-JPL ET differences, the variance of the GLEAM and tower ET difference, and the PT-JPL and tower ET are shown. GLEAM and PT-JPL are much closer to each other than when individually compared with the tower ET. Also, when summarized per season and land cover, GLEAM and PT-JPL differences with the tower are quite comparable. In principle, given the different way of modeling the ET stress by GLEAM and PT-JPL, it could have been expected the tower ET to be more informative respect to the GLEAM and PT-JPL differences. But the relative closeness of GLEAM and PT-JPL, when compared with the tower ET, seems to indicate that the fact that GLEAM and PT-JPL share some model inputs, notably the surface radiation, and some common modeling assumptions, such as close estimation of the potential evaporation, produces more dominant common error patterns at the tower locations and reduce the possibility to characterize other model errors. Note also the good correspondence between average correlations per season and land cover displayed in Fig. 9 and the tower ET variability shown here. The shrubs/savanna optimum product average correlations were the lowest for all seasons, reflecting the lowest ET variability observed in these regions.
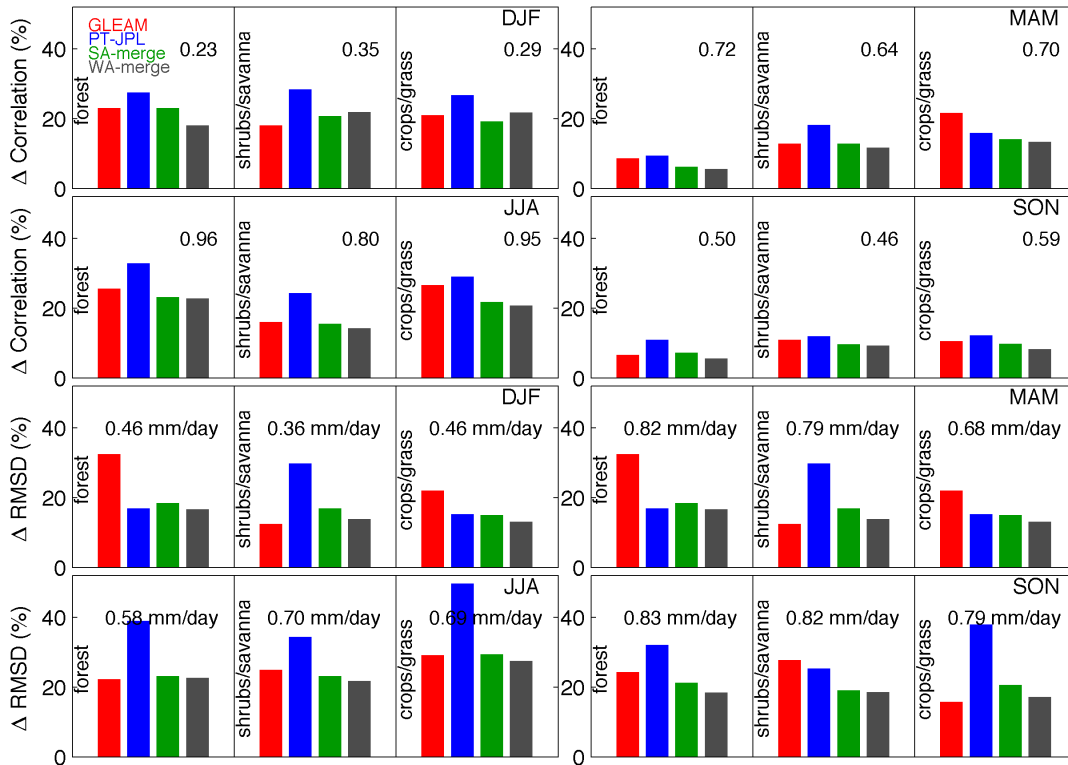
Figure 9. Difference of the season and land-cover averaged correlations of the optimum product ET with the correlation of GLEAM (red), PT-JPL (blue), SA-merge (green) and WA-merge (grey) ET, expressed as percentage of the averaged optimum product correlation (top panels). The same differences for the RMSD are also plotted, but taking the differences between the individual products and the optimum product in order to have positive differences (bottom panels). The average optimum product correlations and RMSD are given in the plots.

A factor potentially affecting GLEAM and PT-JPL merge is the statistical nature of their ET differences with the tower data. Assuming that systematic differences between models and observations come mostly from the model, a decrease of the systematic difference component when comparing with observations is typically a sign of improved model per-

formance. In the context of merging products, if the difference with the observations were mostly of random nature, we should not expect the observations to provide much guidance to combine the products. Figure 11 shows box plots of the ratio of the MSD_r over the sum of MSD_r and MSD_s (i.e, the total MSD) for GLEAM, PT-JPL, and the SA and WA-merge (see Equations 7 and 8). The ratios take values over the whole zero to one range. Analyses of the ratios for the individual towers did not show any obvious patterns in terms of biome or climate conditions. As the medians are larger than 0.5, this indicates that there is a larger number of stations where the random component is larger than the systematic for the three land cover groups. The merged products do not largely change the ratio distributions, so if we assume that a decrease of the systematic component can be indicative of a better fit to the observations, it cannot be claimed that the merge products are reducing biases with the observations.

The validity of Equation 2 for the merge product can also be discussed. Ideally Equation 2 should be applied with unbiased and independent estimates. As discussed in Section 2.4, this is not the case for GLEAM and PT-JPL ET estimates. Concerning the bias, at most stations the data record to derive a meaningful climatology is too short, so bias-correcting the model ET estimates before deriving the weights is not feasible for a large number of stations. Regarding the model ET dependences, the correlation can be taken into account by modifying Equation 2 as:

$$E_{WA} = \frac{\sigma_{PT}^2 - \rho\sigma_{GL}\sigma_{PT}}{\sigma_{GL}^2 + \sigma_{PT}^2 - \rho\sigma_{GL}\sigma_{PT}} E_{GL} + \frac{\sigma_{GL}^2 - 2\rho\sigma_{GL}\sigma_{PT}}{\sigma_{GL}^2 + \sigma_{PT}^2 - 2\rho\sigma_{GL}\sigma_{PT}} E_{PT} \qquad (9)$$

where ρ is the correlation coefficient between GLEAM and PT-JPL (Jones et al., 2008). However, negative weights are now possible. For instance, if $\sigma_{GL} > \sigma_{PT}$ and $\rho > \sigma_{PT}/\sigma_{GL}$, GLEAM will be weighted with a negative value, and even if the sum of GLEAM and PT-JPL weights is still one, the merged ET estimates will be outside the ET range defined by GLEAM and PT-JPL original ET values. This is statistically correct, but allows the values to be extrapolated outside this range. Application of Equation 9 results indeed at negative

weights over a large number of stations, but does not produce very different merged estimates, compared with the results previously discussed. The weighted average statistics (results not displayed) are very close to the merged estimates using the original formulation of Equation 2, and does not improve the significance of the results, so it seems that ignoring the dependence of GLEAM and PT-JPL estimates is not the limiting factor of the merging exercise in this particular case.
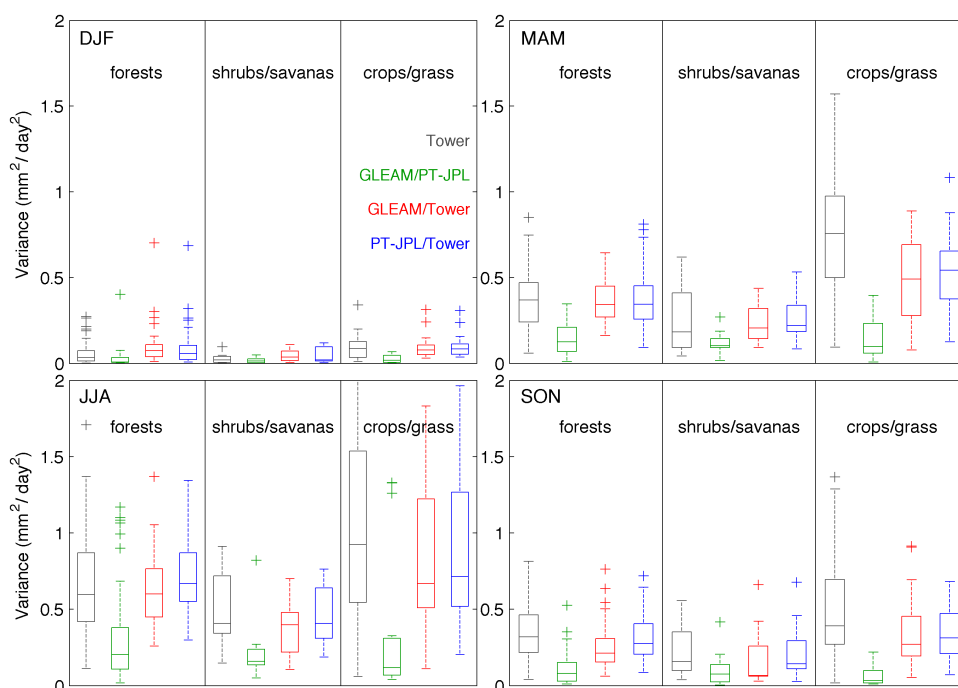
Figure 10. Box plots summarizing for the three land cover goups and four seasons the variance of the towers ET (black), the variance of difference between GLEAM and PT-JPL ET (green), and the variance of the difference between the tower ET and GLEAM (red) and PT-JPL (blue) ET.

4   Conclusions

An inverse-variance weighting of two ET daily products at 25 km resolution from the WACMOS-ET project over a selection of tower sites is tested. The two ET models, PT-JPL and GLEAM, are forced with some common input data sets, and share some common features in their modeling framework such as the Priestley-Taylor formulation to estimate the potential evaporation rate. The latter is converted into actual volumes of evaporation using model-specific formulations of evaporative stress. The weights are based on the variance of the error-distribution of the individual products, with the error defined as the difference between tower ET and modeled ET for dry conditions. To produce dynamic weights changing seasonally, they are estimated by a running window of 61 days centered at each day of the year. The assumption of both ET products being equally uncertain, i.e., assigning a constant weight of 0.5, is also tested.
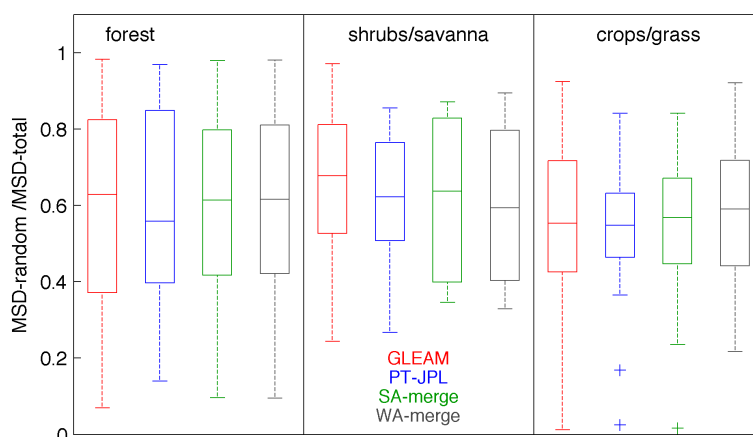
Figure 11. Box plots showing for the three land cover groups the ratio of the random MSD between the tower ET and GLEAM (red), PT-JPL (blue), SA-merge (green) and WA-merge (grey) ET over the sum of the total MSD for the same pair of ET products.

The weights over some stations showed seasonal patterns, but the differences from 0.5 were at most cases confined to ±0.2, meaning that GLEAM and PT-JPL were weighted closely to each other. Inspecting the variance of the GLEAM and PT-JPL ET differences showed smaller values than the variance of the difference between the tower ET and the individual GLEAM or PT-JPL ET. This implies that even if GLEAM and PT-JPL ET differs, they are still relatively close, compared with the tower ET of the selected sites. For a few stations where the GLEAM and PT-JPL were more distinct, their weighting resulted in a merged product better capturing some of the measured ET changes at the tower site. However, the correlations of the merged products with the tower ET were not significantly higher than the original GLEAM and PT-JPL correlations.

If a merged optimum ET product minimizing the RMSD is constructed to set a performance target, the correlation and RMSD performance improvement of the merged products with respect to this new product can be assessed. Compared with the original GLEAM and PT-JPL ET, at a large number of stations the RMSD and correlation of the merged products with the tower is closer to the optimum product, demonstrating some improvement in performance by the merged products. However, the margin of improvement with respect to the optimum product is not very large, and the simple average and weighted product statistics are very close with no statistically significant differences. According to this, for these specific ET models, spatial resolution, observation period, and available tower measurements, the tower and modeled ET differences are not informative enough to derive a merged product much better fitting the tower ET than the original GLEAM and PT-JPL products.

Limiting factors for this exercise are the restricted geographical coverage of the towers and the representativeness of the towers at the model spatial resolution. Most of the available towers for the study period are situated in temperate regions, where GLEAM and PT-JPL do not show the largest differences. In the less covered regions many towers only have data records for a limited number of years, so using a larger number of years would have given access to some of them, permitting to test the weighting at places where the tower ET may have provided more insight. The mismatch between modeled data resolution and tower footprint is always an issue. Although an apparent link between tower surrounding

spatial homogeneity and magnitude of the errors was not found, it is clear that the impact of spatial resolution mismatch is minimized if the modeled ET is available at finer spatial resolutions. Therefore, modeling ET at a finer spatial resolution should help for a posteriori tower-based merging of the estimates.

5   In this study the GLEAM and PT-JPL products are derived with common data sets for their shared inputs. These are the surface net radiation and meteorology. This is key to study inter-product differences related to the modeling components, but it eliminates the more common situation of also having inter-product differences associated to applying the models with different input data sets. For this specific observation period and available towers,

10 it seems that the unique input data sets introduced common error patterns with respect to the tower observations. The relatively close modeling framework could also be partially responsible for the common errors. The result is that, even if ET differences between GLEAM and PT-JPL are observed at the selected sites, they are not too distinct when compared with the tower observations to clearly weight GLEAM and PT-JPL very differently.

15   This study suggests that merging tower observations and ET products at the time and spatial scales of this study (daily and 25 km) is not straightforward, and that care has to be taken regarding the dependence of the products to be merged, the tower coverage, errors, and spatial representativeness of their measurements at the products resolution, and the nature of the ET product errors. For these two specific ET modelling frameworks, GLEAM

20 and PT-JPL, it is likely that more distinct error patterns would have been observed if the model inputs were more independent, which could have lead to a more useful weighting of the products. This could be the case for the presumably more different new versions of GLEAM (Martens et al., 2016) and PT-JPL (Fischer et al, manuscript in preparation), which are derived with more different input products, and which are available for a more

25 extended period, facilitating the use of a larger pool of tower data. Work on combining these new products with a similar technique, together with some other daily remote sensing ET products based on more different evaporation frameworks, is envisaged, including research on extrapolation techniques to propagate the local weights outside the tower sites with the

objective of producing a global merged data set of widely used remote sensing based ET products.

Acknowledgements. This study was funded by the European Space Agency (ESA) and conducted as part of the project WACMOS-ET-Ensemble (ESRIN Contract No. 4000117355/16/I-NB). D.G.Miralles acknowledges support from the European Research Council (ERC) under grant agreement number 715254 (DRY-2-DRY). J.B. Fischer contributed to this paper at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. California Institute of Technology. Government sponsorship acknowledged. Support to J.B Fisher was provided by NASA's SUSMAP, THP, and INCA programs, and the ECOSTRESS mission. K. Tu, from the Department of Ecosystem and Conservation Sciences, University of Montana, is acknowledged by providing guidance for the use of the vegetation products in this study. This work used eddy covariance data acquired by the FLUXNET community and in particular by the following networks: AmeriFlux (U.S. 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, USCCC. Data and logistical support for the station US-Wrc were provided by the US Forest Service Pacific Northwest Research Station. The FLUXNET eddy-covariance data processing and harmonization were carried out by the ICOS Ecosystem Thematic Center, the AmeriFlux Management Project, and the Fluxdata project of FLUXNET, with the support of CDIAC, and the OzFlux, ChinaFlux, and AsiaFlux offices.

References

Aires, F.: Combining Datasets of Satellite-Retrieved Products. Part I: Methodology and Water Budget Closure, Journal of Hydrometeorology, 15, 1677–1691, 2014.

Amiro, B., Barr, A., Black, T., Iwashita, H., Kljun, N., McCaughey, J., Morgenstern, K., Murayama, S., NESIC, Z., and ORCHANSKY, A.: Carbon, energy and water fluxes at mature and disturbed forest sites, Saskatchewan, Canada, Agricultural and Forest Meteorology, 136, 237 – 251, doi:10.1016/j.agrformet.2004.11.012, 2006.

Amos, B., Arkebauer, T. J., and Doran, J. W.: Soil surface fluxes of greenhouse gases in an irrigated maize-based agroecosystem, Soil Science Society of America Journal, 69, 387–395, doi:10.2136/sssaj2005.0387, 2005.

Aubinet, M., Chermanne, B., Vandenhaute, M., Longdoz, B., Yernaux, M., and Laitat, E.: Long term carbon dioxide exchange above a mixed forest in the Belgian Ardennes, Agricultural and Forest Meteorology, 108, 293–315, 2001.

Badgley, G., Fisher, J. B., and Jimenez, C.: On Uncertainty in Global Terrestrial Evapotranspiration Estimates from Choice of Input Forcing Datasets, Journal of Hydrometeorology, 6, 1449–1455, doi:10.1175/JHM-D-14-0040.s1, 2015.

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. T., Pilegaard, K., Schmid, H. P., Valentini, R., Verma, S., Vesala, T., Wilson, K., and Wofsy, S.: FLUXNET: A New Tool to Study the Temporal and Spatial Variability of Ecosystem-Scale Carbon Dioxide, Water Vapor, and Energy Flux Densities., Bulletin of the American Meteorological Society, 82, 2415–2434, 2001.

Bazot, S., Barthes, L., Blanot, D., and Fresneau, C.: Distribution of non-structural nitrogen and carbohydrate compounds in mature oak trees in a temperate forest at four key phenological stages, Trees, 27, 1023–1034, 2013.

Beck, H. E., van Dijk, A. I. J. M., Levizzani, V., Schellekens, J., Miralles, D. G., Martens, B., and de Roo, A.: MSWEP: 3-hourly 0.25° global gridded precipitation (1979–2015) by merging gauge, satellite, and reanalysis data, Hydrology and Earth System Sciences, 21, 589–615, 2017.

Bergeron, O., Margolis, H. A., Black, T. A., Coursolle, C., Dunn, A. L., BARR, A. G., and Wofsy, S. C.: Comparison of carbon dioxide fluxes over three boreal black spruce forests in Canada, Global Change Biology, 13, 89 – 107, doi:10.1111/j.1365-2486.2006.01281.x, 2007.

Betts, A. K.: Land-surface-atmosphere coupling in observations and models, Journal of Advances in Modeling Earth Systems, 2, 4–18, 2009.

B.Lamberty, B., Wang, C., and Gower, S. T.: Net primary production and net ecosystem production of a boreal black spruce wildfire chronosequence, Global Change Biology, 10, 473 – 487, doi:10.1111/j.1529-8817.2003.0742.x, 2004.

Bond-Lamberty, B., Wang, C. K., and Gower, S. T.: Net primary production and net ecosystem production of a boreal black spruce wildfire chronosequence, Global Change Biology, 10, 473–487, doi:10.1111/j.1529-8817.2003.0742.x, 2004.

Campbell, J. L. and Law, B. E.: Forest soil respiration across three climatically distinct chronosequences in Oregon, Biogeochemistry, 73, 109–125, 2005.

Chen, Q., Gong, P., Baldocchi, D., and Tian, Y. Q.: Estimating basal area and stem volume for individual trees from lidar data, Photogrammetric Engineering and Remote Sensing, 73, 1355–1365, doi:http://dx.doi.org/10.14358/PERS.73.12.1355, 2007.

Cook, B. D., Davis, K. J., Wang, W. G., Desai, A., Berger, B. W., Teclaw, R. M., Martin, J. G., Bolstad, P. V., Bakwin, P. S., Yi, C. X., and Heilman, W.: Carbon exchange and venting anomalies in an upland deciduous forest in northern Wisconsin, USA, Agricultural and Forest Meteorology, 126, 271–295, doi:10.1016/j.agrformet.2004.06.008, 2004.

Corradi, C., Kolle, O., Walter, K., Zimov, S. A., and Schulze, E.-D.: Carbon dioxide and methane exchange of a north-east Siberian tussock tundra, Global Change Biology, pp. 1910–1925, doi:10.1111/j.1365-2486.2005.01023.x, 2005.

Coursolle, C., Margolis, H. A., Giasson, M.-A., Bernier, P.-Y., Amiro, B., Arain, M. A., Barr, A., Black, T. A., GOULDEN, M. L., McCaughey, J., Chen, J., Dunn, A., Grant, R. F., and Lafleur, P.: Influence of stand age on the magnitude and seasonality of carbon fluxes in Canadian forests, Agricultural and Forest Meteorology, 165, 136 – 148, doi:10.1016/j.agrformet.2012.06.011, 2012.

De Lannoy, G. and Reichle, R. H.: Global assimilation of multiangle and multipolarization SMOS brightness temperature observations into the GEOS-5 catchment land surface model for soil moisture Estimation, Journal of Hydrometeorology, 17, 669–691, 2016.

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. R. Meteorol. Soc., 137, 553–597, doi:10.1002/qj.828, 2011.

Dunn, A. L., Barford, C. C., Wofsy, S. C., Goulden, M. L., and Daube, B. C.: A long-term record of carbon exchange in a boreal black spruce forest: means, responses to interannual variability, and decadal trends, Global Change Biology, 13, 577 – 590, doi:10.1111/j.1365-2486.2006.01221.x, 2007.

Dunn, S. M. and Mackay, R.: Spatial variation in evapotranspiration and the influence of land use on catchment hydrology, Journal of Hydrology, 171, 49–73, 1995.

Fischer, M. L., Billesbach, D. P., Berry, J. A., Riley, W. J., and Torn, M. S.: Spatiotemporal variations in growing season exchanges of CO2, H2O, and sensible heat in agricultural fields of the Southern Great Plains, Earth Interactions, 11, 1–21, 2007.

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

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

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

Goldstein, A. H., Hultman, N. E., Fracheboud, J. M., Bauer, M. R., Panek, J. A., Xu, M., Qi, Y., Guenther, A. B., and Baugh, W.: Effects of climate variability on the carbon dioxide, water, and sensible heat fluxes above a ponderosa pine plantation in the Sierra Nevada (CA), Agricultural and Forest Meteorology, 101, 113–129, doi:http://dx.doi.org/10.1016/S0168-1923(99)00168-9, 2000.

Goulden, M. L., Munger, J. W., Fan, S. M., Daube, B. C., and Wofsy, S. C.: Measurements of carbon sequestration by long-term eddy covariance: Methods and a critical evaluation of accuracy, Global Change Biology, 2, 169–182, doi:10.1111/j.1365-2486.1996.tb00070.x, 1996.

Gowda, P. H., Chavez, J. L., Colaizzi, P. D., Evett, S. R., Howell, T. A., and Tolk, J. A.: ET mapping for agricultural water management: present status and challenges, Irrigation Science, 26, 223–236, doi:10.1007/s00271-007-0088-6, 2008.

Jimenez, 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, Journal of Geophysical Research, 116, D02 102–27, 2011.

Jones, C. S., Finn, J. M., and Hengartner, N.: Regression with strongly correlated data, Journal of Multivariate Analysis, 99, 2136–2153, doi:doi:10.1016/j.jmva.2008.02.008, 2008.

Kato, T., Tang, Y., Gu, S., Hirota, M., Du, M., Li, Y., and Zhao, X.: Temperature and biomass influences on interannual changes in CO_2 exchange in an alpine meadow on the Qinghai-Tibetan Plateau, Global Change Biology, 12, 1285–1298, doi:10.1111/j.1365-2486.2006.01153.x, 2006.

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

Knohl, A., Schulza, E. D., Kolle, O., and Buchmann, N.: Large carbon uptake by an unmanaged 250-year-old deciduous forest in Central Germany, Agricultural and Forest Meteorology, 118, 151–167, doi:http://dx.doi.org/10.1016/S0168-1923(03)00115-1, 2003.

Le Maitre, D. C. and Versfeld, D. B.: Forest evaporation models: relationships between stand growth and evaporation, Journal of Hydrology, 193, 240–257, 1997.

Lievens, H., Martens, B., Verhoest, N. E. C., Hahn, S., Reichle, R. H., and Miralles, D. G.: Assimilation of global radar backscatter and radiometer brightness temperature observations to improve soil moisture and land evaporation estimates, Remote Sensing of Environment, 189, 194–210, 2017.

Liu, Y. Y., de Jeu, R., and McCabe, M. F.: Global long-term passive microwave satellite-based retrievals of vegetation optical depth, Geophysical Research Letters, 38, 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, Hydrology and Earth System Sciences, 15, 425–436, 2011b.

Loew, A., Peng, J., and Borsche, M.: High-resolution land surface fluxes from satellite and reanalysis data (HOLAPS v1.0): evaluation and uncertainty assessment, Geoscientific Model Development, 9, 2499–2532, 2016.

Ma, S., Baldocchi, D. D., Xu, L., and Hehn, T.: Inter-annual variability in carbon dioxide exchange of an oak/grass savanna and open grassland in California, Agricultural and Forest Meteorology, 147, 157–171, doi:10.1016/j.agrformet.2007.07.008, 2007.

Martens, B., Miralles, D. G., Lievens, H., van der Schalie, R., de Jeu, R. A. M., Férnandez-Prieto, D., Beck, H. E., Dorigo, W. A., and Verhoest, N. E. C.: GLEAM v3: satellite-based land evaporation and root-zone soil moisture, Geoscientific Model Development Discussions, 0, 1–36, 2016.

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, Geoscientific Model Development, 9, 283–305, 2016.

McCaughey, J. H., Pejam, M. R., Arain, M. A., and Cameron, D. A.: Carbon dioxide and energy fluxes from a boreal mixedwood forest ecosystem in Ontario, Canada, Agricultural and Forest Meteorology, 140, 79–96, doi:10.1016/j.agrformet.2006.08.010, 2006.

McEwing, K. R., Fisher, J. P., and Zona, D.: Environmental and vegetation controls on the spatial variability of CH4 emission from wet-sedge and tussock tundra ecosystems in the Arctic, Plant and soil, 388, 37–52, 2015.

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

Milyukova, I. M., Kolle, O., Varlagin, A. V., Vygodskaya, N. N., Schulze, E. D., and Lloyd, J.: Carbon balance of a southern taiga spruce stand in European Russia, Tellus B, 54, 429–442, doi:10.1034/j.1600-0889.2002.01387.x, 2002.

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, Hydrology and Earth System Sciences, 15, 967–981, 2011a.

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, Hydrology and Earth System Sciences, 15, 453–469, 2011b.

Miralles, D. G., Jimenez, 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 Fernández-Prieto, D.: The WACMOS-ET project - Part 2: Evaluation of global terrestrial evaporation data sets, Hydrology and Earth System Sciences, 20, 823–842, 2016.

Moncrieff, J., Malhi, Y., and Leuning, R.: The propagation of errors in long-term measurements of land-atmosphere fluxes of carbon and water, Global Change Biology, 2, 231–240, doi:10.1111/j.1365-2486.1996.tb00075.x, 1996.

Monteith, J.: Evaporation and environment, Symp. Soc. Exp. Biol, 19, 4, 1965.

Moureaux, C., Debacq, A., Bodson, B., Heinesch, B., and Aubinet, M.: Annual net ecosystem carbon exchange by a sugar beet crop, Agricultural and Forest Meteorology, 139, 25–39, 2006.

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, Geophysical Research Letters, 38, 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, Hydrology and Earth System Sciences, 17, 3707–3720, 2013.

Munier, S., Aires, F., Schlaffer, S., Prigent, C., Papa, F., Maisongrande, P., and Pan, M.: Combining data sets of satellite-retrieved products for basin-scale water balance study: 2. Evaluation on the Mississippi Basin and closure correction model, Journal of Geophysical Research,119, 12,100–12,116, doi:10.1002/2014JD021953, 2014.

Noormets, A., McNulty, S. G., DeForest, J. L., Sun, G., Li, Q., and Chen, J.: Drought during canopy development has lasting effect on annual carbon balance in a deciduous temperate forest, New Phytologist, 179, 818–828, doi:10.1111/j.1469-8137.2008.02501.x, 2008.

Nordbo, A., Järvi, L., and Vesala, T.: Revised eddy covariance flux calculation methodologies – effect on urban energy balance, Tellus B: Chemical and Physical Meteorology, 64, 18 184, 2012.

Pauwels, V. R. N., Timmermans, W., and Loew, A.: Comparison of the estimated water and energy budgets of a large winter wheat field during AgriSAR 2006 by multiple sensors and models, Journal of Hydrology, 349, 425–440, 2008.

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, Journal of Geophysical Research: Atmospheres, 112, 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.

Post, H., Hendricks Franssen, H. J., Graf, A., Schmidt, M., and Vereecken, H.: Uncertainty analysis of eddy covariance CO2 flux measurements for different EC tower distances using an extended two-tower approach, Biogeosciences, 12, 1205–1221, 2015.

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

Richardson, A. D., Hollinger, D. Y., Burba, G. G., Davis, K. J., Flanagan, L. B., Katul, G. G., Munger, J. W., Ricciuto, D. M., Stoy, P. C., and Suyker, A. E.: A multi-site analysis of random error in

tower-based measurements of carbon and energy fluxes, Agricultural and Forest Meteorology, 136, 1–18, 2006.

Rodgers, C. D.: Inverse methods for atmospheric sounding: Theory and practise. Series on Atmospheric, Oceanic and Planetary Physics, vol. 2, World Scientific Publishing, 1 edn., 2000.

Schmid, H. P., Grimmond, C. S. B., Cropley, F., Offerle, B., and Su, H. B.: Measurements of CO_2 and energy fluxes over a mixed hardwood forest in the mid-western United States, Agricultural and Forest Meteorology, 103, 357–374, doi:10.1016/S0168-1923(00)00140-4, 2000.

Scott, R. L., Jenerette, G. D., Potts, D. L., and Huxman, T. E.: Effects of seasonal drought on net carbon dioxide exchange from a woody-plant-encroached semiarid grassland, Journal of Geophysical Research-Biogeosciences, 114, doi:10.1029/2008JG000900, 2009.

Simbahan, G. C., Dobermann, A., Goovaerts, P., Ping, J. L., and Haddix, M. L.: Fine-resolution mapping of soil organic carbon based on multivariate secondary data, Geoderma, 132, 471–489, doi:10.1016/j.geoderma.2005.07.001, 2006.

Sorooshian, S., Lawford, R., and Try, P.: Water and energy cycles: Investigating the links, WMO Bulletin, 54, 58–64, 2005.

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.

Steininger, M. K.: Net carbon fluxes from forest clearance and regrowth in the Amazon, Ecological Applications, 14, 313–322, doi:10.1890/02-6007, 2004.

Takala, M., Luojus, K., Pulliainen, J., Derksen, C., Lemmetyinen, J., Kärnä, J.-P, Koskinen, J., and Bojkov, B.: Estimating northern hemisphere snow water equivalent for climate research through assimilation of space-borne radiometer data and ground-based measurements, Remote Sensing of Environment, 115, 3517–3529, 2011.

Verma, S. B., Dobermann, A., Cassman, K. G., Walters, D. T., Knops, J. M., Arkebauer, T. J., Suyker, A. E., Burba, G. G., Amos, B., Yang, H. S., Ginting, D., Hubbard, K. G., Gitelson, A. A., and Walter-Shea, E. A.: Annual carbon dioxide exchange in irrigated and rainfed maize-based agroecosystems, Agricultural and Forest Meteorology, 131, 77–96, doi:10.1016/j.agrformet.2005.05.003, 2005.

Wang, J., Zhuang, J., Wang, W., Liu, S., and Xu, Z.: Assessment of Uncertainties in Eddy Covariance Flux Measurement Based on Intensive Flux Matrix of HiWATER-MUSOEXE, IEEE Geoscience and Remote Sensing Letters, 12, 259–263, 2015.

Wang, K. and Dickinson, R. E.: A review of global terrestrial evapotranspiration: Observation, modeling, climatology, and climatic variability, Reviews of Geophysics, 50, RG2005–54, 2012.

Willmott, C. J.: Some comments on the evaluation of model performance, Bulletin of the American Meteorological Society, 63, 1309–1313, 1982.

Yao, Y., Liang, S., Li, X., Chen, J., Liu, S., Jia, K., Zhang, X., Xiao, Z., Fisher, J. B., Mu, Q., Pan, M., Liu, M., Cheng, J., Jiang, B., Xie, X., Gr nwald, T., Bernhofer, C., and ROUPSARD, O.: Improving global terrestrial evapotranspiration estimation using support vector machine by integrating three process-based algorithms, Agricultural and Forest Meteorology, 242, 55–74, 2017.

Zeeman, M. J., Hiller, R., Gilgen, A. K., Michna, P., Plüss, P., Buchmann, N., and Eugster, W.: Management, not climate, controls net CO_2 fluxes and carbon budgets of three grasslands along an elevational gradient in Switzerland, Agricultural Forest Meteorology, 50, 519–530, 2010.

Zhang, K., Kimball, J. S., and Running, S. W.: A review of remote sensing based actual evapotranspiration estimation, Wiley Interdisciplinary Reviews: Water, 3, 834–853, 2016.