Manuscript prepared for Hydrol. Earth Syst. Sci.
with version 2015/04/24 7.83 Copernicus papers of the LaTeX class copernicus.cls.
Date: 23 March 2018

Exploring the merging of the global land evaporation WACMOS-ET products based on local tower measurements

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 three terrestrial evaporation (ET) products from the WACMOS-ET project based on FLUXNET sites is presented. The three ET models, GLEAM, PT-JPL, and PM-MOD, are run daily and at a resolution of 25 km for the 2002-2007 period, and use common input data when possible. The local weights are based on the variance of the difference between the tower ET and the modelled ET, are made dynamic by estimating them using a 31-day running window centered on each day, and are extrapolated from the tower locations to the global landscape by regressing them on the main model inputs. Seasonal variability in the local weights is observed over some stations, but the deviations from the 1/3 value assumed by the arithmetic mean of the three products is small at many stations. The global weights show seasonal and geographical patterns, which can be related to deficiencies in model parameterization and inputs, but also to errors in the local weights derivation and the weights extrapolation. The latter was confirmed by tests showing that the tower data set, mostly located at temperate regions, has limitations to represent different biome and climate conditions. 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, the nature of the error in both towers and gridded data sets, and how all these factors impact the weights extrapolation from the tower locations to the global landscape.

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).

Point-based measurements of land heat fluxes are typically conducted during field experiments (Pauwels et al., 2008) and by more permanent lysimeters (Hirschi et al., 2017) and 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. These efforts range from simply averaging a number of ET products (Mueller et al., 2013) to more complex approaches, such as weighted averages (Hobeichi et al., 2018), 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). ET products based on a direct regression of tower ET on a set of explanatory variables also exist(Jung et al., 2011).

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 to explore ET estimation from process-based algorithms (Michel et al., 2016; Miralles et al., 2011). Three of the models – the Priestley-Taylor Jet Propulsion Laboratory model (PT-JPL, Fisher et al., 2008), the Global Land Evaporation Amsterdam Model (GLEAM, Miralles et al., 2011), and the Penman–Monteith algorithm from the MODerate resolution Imaging Spectroradiometer (MODIS) evaporation product (PM-MOD, Mu et al., 2011) – were run at global scale, and substantial differences were found between the three model products. As such we pose the question: can a product combining the GLEAM, PT-JPL, and PM-MOD estimates result in a more accurate ET estimate? To start research in this direction, we will investigate a weighted combination of the three model ET estimates. 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 quality assessment is only attained through validation against tower flux measurements. Therefore, here we propose a flux tower-based weighting of GLEAM, PT-JPL, and PM-MOD and investigate the performance of the resulting merger over a selection of tower sites, followed by an exploration of the possibility to use the methodology for a global merged ET product.

The paper is organized as follows: first, the ET models, a description of the merging technique, and the metrics used in the analyses are presented. The model input data sets, the tower observations, and ancillary data used in the analyses are then described. This is followed by a presentation of the merged products at the local and global scales, and a discussion of the limitations and quality of the products. Finally, the main conclusions of the study are summarized.

2 Methods

2.1 ET models

The GLEAM, PT-JPL, and PM-MOD models, and the inputs required to run them at the global scale are extensevly 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 distinguishes 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 observations of precipitation. The remaining components of evaporation are based upon the formulation by Priestley and Taylor (1972) for potential evaporation, constrained by multiplicative 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, sublimation 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 replaces the model of Miralles et al. (2011) 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. Note that the WACMOS-ET runs were done at 3-hourly and daily 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) expected 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.

2.1.3 PM-MOD

The PM-MOD is based on the Monteith (1965) adaptation of Penman (1948), and the version applied here follows the implementation of Mu et al. (2011). It estimates ET as the sum of interception loss, transpiration, and soil evaporation. Aerodynamic and surface resistances for each component of evaporation are based on extending biome-specific conductance parameters to the canopy scale using vegetation phenology and meteorological data. The surface resistance schemes uses LAI, with further constrains based on air temperature and vapour pressure deficit, avoiding the more typical use of soil moisture and wind speed to parameterize the resistances. Different to GLEAM and PT-JPL, which do not use tower-based calibration, some of the resistance parameters require a biome-based calibration derived from tower measurements. As for PT-JPL, there are no specific parameterization for snow-covered areas.

The WACMOS-ET LAI/FAPAR products are used with PM-MOD as in Michel et al. (2016), i.e., the model is run with the vegetation products rescaled by a biome-dependent calibration to make them consistent with the expected MODIS values. As the biome-based calibration of PM-MOD was derived with MODIS products, any errors introduced by this simple rescaling can propagate to the PM-MOD estimates and can be responsible for some ET patterns differing from the official use of the Mu et al. (2011)algorithm for the MODIS ET product.

2.2 Merging technique

2.2.1 Tower weighting

A weighted combination of 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 the 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:

$$E_{WA} = w_{GL}E_{GL} + w_{PT}E_{PT} + +w_{PM}E_{PM} \tag{1}$$

$$w_{m=GL,PT,PM} = \sigma_m^{-2}(\sigma_{GL}^{-2} + \sigma_{PT}^{-2} + \sigma_{PM}^{-2})^{-1} \tag{2}$$

where E_{WA} is the Weigthed Average merged product (WA-merge), E_{GL}, E_{PT}, and E_{PM} are GLEAM, PT-JP, and PM-MOD ET, w_{GL}, w_{PT}, and w_{PM} are their respective weights, and σ_{GL}^2,

σ^2_{PT}, and σ^2_{PT} are the variances of their respective error distributions. For the SA-merge, weights are equal and w_{GL}=w_{PT}=w_{PM}=1/3. For the WA-merge, the error is defined as the difference between the model and tower-based ET, the σ_{GL}, σ_{PT}, and σ_{PM} are estimated over the time series of available errors, and the weights derived following Equation 2. Notice that $w_{GL} + w_{PT} + w_{PM} = 1$.

If the errors follow a Gaussian distribution, are unbiased, and are independent from each other, Equation 1 is the optimal linear estimator (Rodgers, 2000). However, in practice, those conditions are difficult to meet. While GLEAM, PT-JPL, and PM-MOD can be considered independent from the tower observations, they do share common inputs, likely making the errors in their estimates dependent. Bias can also be present between the surface meteorological products used by the ET models, such as the surface radiation, or the near-surface air temperature and humidity, and the real meteorological conditions at the tower, due to for instance differences in footprint. The different ET 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 centred at that day. The choice of window length is subjective: 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 15 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 31-day running window was found to provide the daily weights.

GLEAM, PT-JPL, and PM-MOD estimate separately transpiration, soil evaporation, and the evaporation from the rain intercepted by canopies. Tower measurements will be masked for rainy intervals (see Section 3.2), so the interception loss of the modelled ET cannot be evaluated. Therefore, only the sum of soil evaporation and transpiration is compared with the tower data and weighted. To derive the total ET merged product, an estimate of interception loss should also be provided, either by (1) assuming that GLEAM, PT-JPL, and PM-MOD interception loss are equally uncertain and adding their average to the weighted soil evaporation and transpiration, or; (2) by adding just one of the individual model interception losses, if there are reasons to believe that the selected one is less uncertain. Here we adopt the first approach, so the total ET product is the sum of the weighted soil evaporation and transpiration, together with the average of the three products interception losses.

2.2.2 Weights extrapolation

In order to produce a global weighted product, an extrapolation of the weights from the tower space (i.e., the 84 pixels where the towers are located) to the pixels of the remaining continental land is needed. The approach tested here is to predict the weights outside the tower space by non-linearly regressing the weights on the main model inputs. For the regression, we use a neural network (NN). NNs are broadly used given their capability to approximate non-linear functions, so the NN is in principle a suitable tool to extrapolate the weights. Nevertheless, it is clear that the weights can never be perfectly predicted. Only the systematic component of the error can potentially be captured by the NN prediction, and there is no warranty that all the systematic errors are dependent on the model inputs.

A standard multi-layer perceptrons of one hidden layer, with their initial weights randomly initialized by the Nguyen-Widrow algorithm (Nguyen and Widrow, 1990), and the final weights assigned by a Marquardt-Levenberg backpropagation algorithm (Hagan and Menhaj, 1994), is used for the regression. To prevent over-fitting to the training data set, a cross-validation technique is applied to monitor the evolution of the training error function. Regarding the predictors, we use the surface net radiation, the near-surface air temperature, the relative humidity, the soil moisture, the vegetation optical depth, and the project LAI and FAPAR as inputs to the NN (see Section 3.1), with the GLEAM, PT-JPL, and PM-MOD weights being the outputs to be predicted by the NN.

2.3 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:

$$\mathrm{R} = \frac{N\sum_{i=1}^{N} P_i O_i - \sum_{i=1}^{N} P_i \sum_{i=1}^{N} O_i}{\sqrt{[N\sum_{i=1}^{N} {P_i}^2 - (\sum_{i=1}^{N} P_i)^2]}\sqrt{N\sum_{i=1}^{N} {O_i}^2 - (\sum_{i=1}^{N} O_i)^2}} \tag{3}$$

$$\mathrm{MSD} = [\frac{1}{N}\sum_{i=1}^{N}(P_i - O_i)^2] = \mathrm{RMSD}^2 \tag{4}$$

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:

$$\mathrm{MSD}_r = \frac{1}{N}\sum_{i=1}^{N}(\hat{P}_i - O_i)^2 = {\mathrm{RMSD}_r}^2 \tag{5}$$

$$\mathrm{MSD}_s = \frac{1}{N}\sum_{i=1}^{N}(P_i - \hat{P}_i)^2 = {\mathrm{RMSD}_s}^2 \tag{6}$$

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$_s$.

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 Data

3.1 Model inputs

The GLEAM, PT-JPL, and PM-MOD 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 the models 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 and PM-MOD also requires near-surface air humidity, also from ERA-Interim, and the vegetation products discussed in Sections 2.1.2 and 2.1.3. On the other hand, GLEAM requires precipitation, coming from the Multi-Source Weighted-Ensemble Precipitation (MSWEP) version 1 product (Beck et al., 2017b), 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), and from the National Snow and Ice Data Center (NSIDC) in snow-covered regions of the Southern Hemisphere (Kelly et al., 2003).

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			

3.2 Tower data

The FLUXNET 2015 synthesis data set (http://fluxnet.fluxdata.org/) is used to obtain point-based 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 measurements for rainy intervals, only leaving observations if both the global precipitation product and the local measurements (if available) do not indicate precipitation (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 selected 97 stations, but this number was further reduced to 84 by removing stations too close to water bodies, or clearly not representing the overall land cover of the 25 km spatial scale of the gridded ET estimates. The geographical locations of the 84 stations, 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. Note that nearly all stations are in Europe and US.

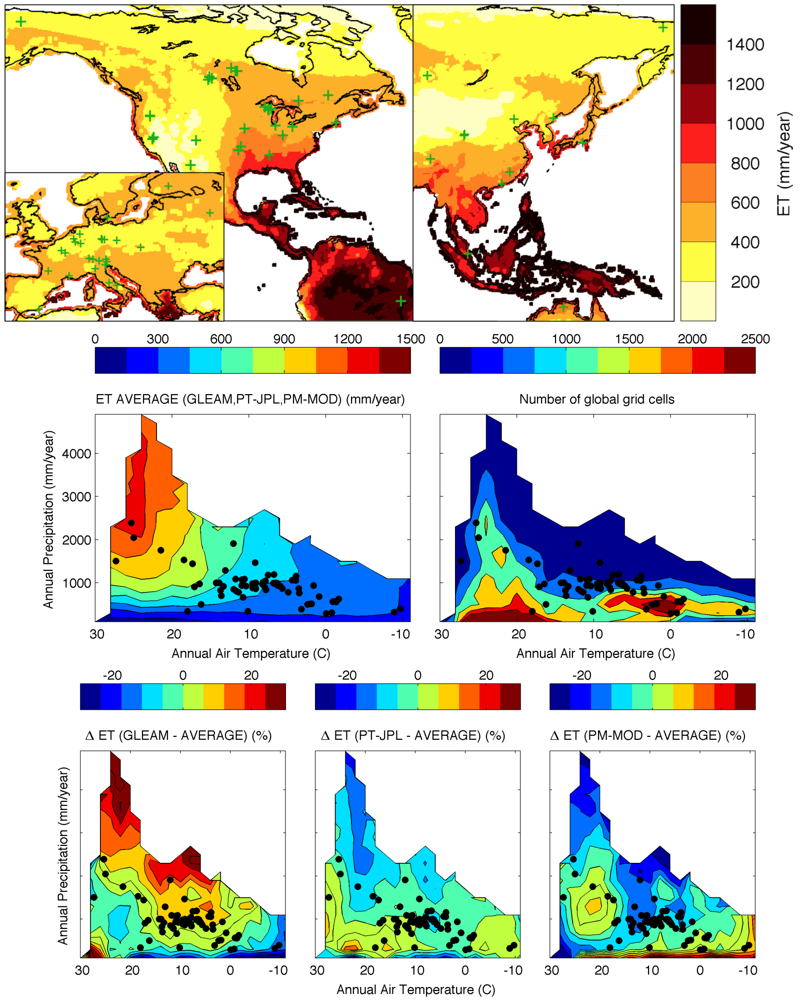

Figure 1. Distribution of tower sites used in the study. Top: geographical location (green crosses) on a map of GLEAM, PT-JPL, and PM-MOD averaged multi-annual ET. Middle: distribution of the averaged multi-annual ET (left), and the number of global grid cells (right), as function of the annual air temperature and precipitation, together with the location of the tower sites in this space (black dots). Bottom: the relative GLEAM (left), PT-JPL (middle), and PM-MOD (right) ET differences normalized by the previous averaged ET.

Eddy-covariance measurements are subject to errors, both random and systematic, and any merging technique using them as reference is likely to be impacted by those errors. 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 (Wilson et al., 2002; Foken, 2008). Methods to correct the energy unbalance exist, being the most frequently adopted the Bowen ratio approach (Twine et al., 2000) and the energy balance residual approach (Amiro, 2009). Corrected fluxes are typically preferred over the original uncorrected observations, but the correction implies the need for surface radiation and soil heat flux measurements, which are not routinely measured at all stations. At the sites where they exist, the FLUXNET 2015 data set offers a test product containing an energy balance corrected version of the heat fluxes based on the assumption that the measured Bowen ratio is correct. For the 84 stations selected here, 26 do not have Bowen Ratio Corrected (BRC) fluxes. For the remaining 58 stations, the relative mean difference between the original and BRC latent heat fluxes averaged over all stations is 6.1%, with a maximum value of 16.5%. If the correlation coefficient between original and BRC fluxes is calculated at each station and then averaged over all stations, we obtain 0.96, showing that they correlate well in time. Also, if over the 58 stations with BRC fluxes we calculate the normalized weights given by Equation 2 with the original and BRC fluxes, they display a 0.91 average correlation over all stations and models, with an average RMSD of 0.035. As these numbers do not suggest strong differences between using the uncorrected and BRC measurements over our selected stations, we use the original uncorrected fluxes for all stations to avoid mixing original and BRC fluxes.

Not all stations completely cover the 2002-2006 period, with 6, 14, 24, 9, and 31 stations having 2, 3, 4, 5, and 6 years of data, respectively. At stations where inter-annual variability is large the weights may not be representative of the overall climate conditions at the tower if only a relatively short number of years exist. Limiting the study to stations with a relatively large number of years could have been used to minimize the impact of this, but this severely reduced the number of towers, so this filtering has not been applied. For instance, if we only derive weights if at least 4 years of data are available, half of the towers would have been removed. Notice also that due to the masking of the tower data at very few occasions the 31 daily estimates are present in the running window applied to derive the weights, and at least 10 daily values in the running window are required to derive a daily weight. Most stations have weights for nearly all days, but at 8 stations there are larger gaps. The worst case is the tropical BR-Sa3 station, where the frequent rainy episodes complicate the derivation of the weights.

3.3 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:

$$\mathrm{I}_h = \frac{1}{2}Fgt_{IGBP} + \frac{1}{2}(1- \mid Fg_{bare} - Ft_{bare} \mid - \mid Fg_{herb} - Ft_{herb} \mid - \mid Fg_{forest} - Ft_{forest} \mid) \tag{7}$$

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. The first term is the mismatch between the land cover at the tower and at the grid cell level, and the remaining terms are the net mismatch in land cover types across the two resolutions. 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.

To evaluate the merged products, we use river run-off from a compilation of monthly data using different sources, as described in Beck et al. (2015). We also use annual precipitation estimates from Fick and Hijmans (2017), here denoted as WorldClim, as the MSWEP product is used by GLEAM, and therefore not independent from the ET products. Both MSWEP and WordlClim correlate higher than other precipitation products compared in (Beck et al., 2017a), and are therefore preferred here to study differences with the river run-off.

4 Inter-product comparison

The annual GLEAM, PT-JPL, and PM-MOD total ET, together with their absolute and relative differences, are shown in Fig. 2. Differences of the same order can be observed when other products are inter-compared (Jimenez et al., 2011). Using different surface radiation products can already be largely responsible for the differences, but as the models here are run with a common surface radiation product, the observed differences are mainly introduced by the different ways to model ET. The disagreement also extends to the the models partitioning of ET into its different components, as shown in Miralles et al. (2016). We recall here that, as discussed in Section 2.3, only the sum of the soil evaporation and transpiration is validated against tower fluxes.

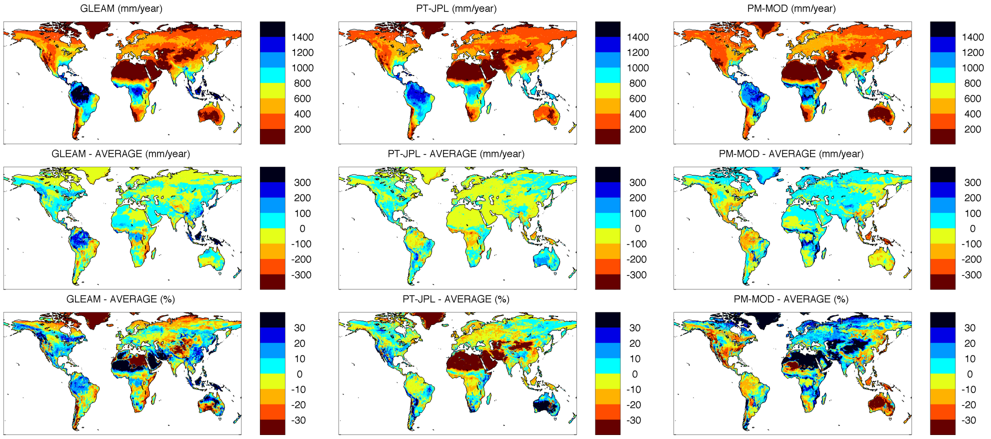

Figure 2. Summary of GLEAM, PT-JPL, and PM-MOD annual ET differences. Top: The GLEAM (left), PT-JPL (middle), and PM-MOD (right) total annual ET in mm/year. Middle: differences with the models averaged ET, in mm/year. Bottom: same differences but normalized with the models averaged ET, and expressed as a percentage.

The GLEAM, PT-JPL, PM-MOD and tower ET are compared now at the available tower sites. If we calculate the relative ET differences between each pair of products at the tower sites, approximately 60% (GLEAM and PT-JPL), 30% (GLEAM and PM-MOD), and 70% (PT-JPL and PM-MOD) of the towers are located in grid cells where the relative ET difference is within ±10%. If we look at the towers spatial distribution of Fig. 1, we can see that most of the towers are located in temperate regions. The tropical rain forest and savannas, where the relative ET differences seem larger, are less represented in the selected tower data. Therefore, some regions that would have been relevant to characterize the model ET 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. 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 the models, Ae is the difference between the surface net radiation and the modelled 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. Clear differences between GLEAM, PT-JPL, PM-MOD and the tower distributions are visible. Overall GLEAM and PT-JPL agree better with each other than with PM-MOD, which may be related to the common modelling framework of Priestley-Taylor for GLEAM and PT-JPL, compared with the more different Penman–Monteith approach for PM-MOD.

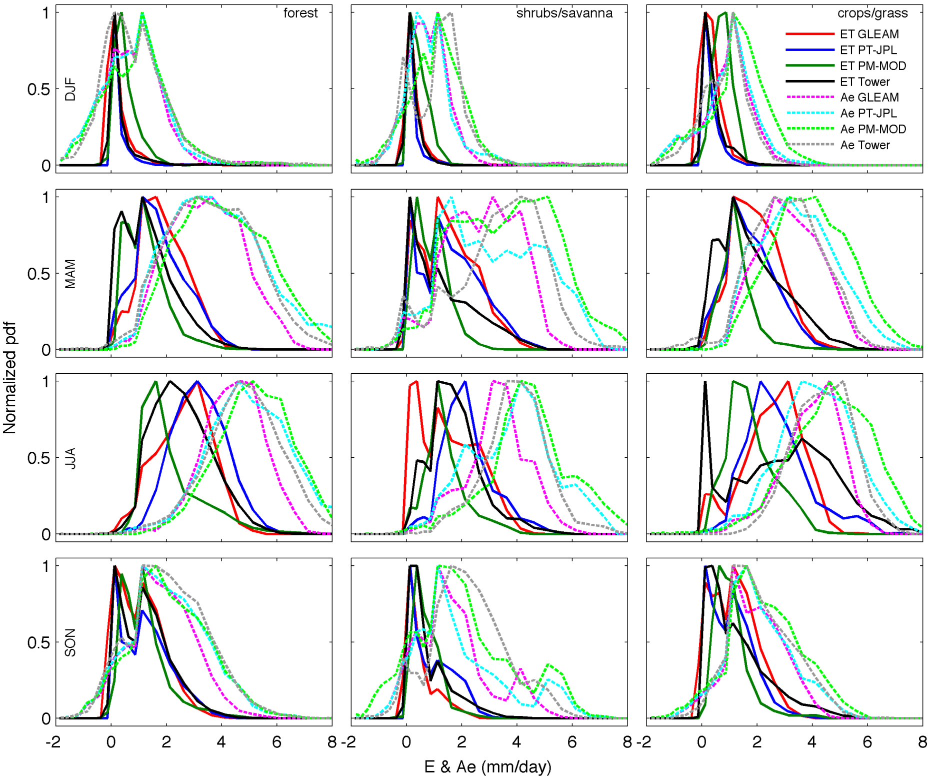

Figure 3. Normalized histograms of ET and available energy (Ae) from GLEAM, PT-JPL, PM-MOD, and the tower observations. The histograms are calculated with the ET values at the tower locations separated first by season and land cover.

An example of good agreement is the forest group for the SON months, with the distributions of 355 ET and Ae being quite similar for the observed and modelled variables. The crops/grass group for the JJA months shows also reasonable agreement between the GLEAM and PT-JPL ET distributions, but larger differences with PM-MOD and the tower ET. The tower ET has a clear bimodal distribution, which cannot be replicated by the modelled ET. This may be due to agricultural management practices being poorly captured by the models (e.g., irrigation), but may also reflect the large het- 360 erogeneity of croplands and their (a priori) low representativeness of the larger pixel scale. For the shrubs/savanna group and the JJA months, the four ET distributions are quite different, with the Ae distributions also showing differences. For these cases it is difficult to identify whether tower and model ET differences are due to biases in the surface radiation, or discrepancies in the ET formulations.

5 Local product merging

5.1 Local weights

A summary of daily weight statistics over all the sites belonging to a given land cover group is given in Fig. 4. The SA-merge product equally weights all products with a value of 1/3, and this line is added to the plots to highlight changes with respect to this value. On average, the weights do not deviate much from 1/3, suggesting that, for a given land cover group, there are no clear systematic patterns indicating that one model agrees with the tower data much better than the others. The relative weight of each model can change along with season, suggesting that model agreement with tower data is not uniform along the year. The 25% and 75% percentiles can have large departures for some seasons and covers. For the forest class, noticeable is some tendency for PT-JPL to be weighted more in DJF, while the same is true for PM-MOD in MAM. For the shrub/savanna class, again PT-PJPL in winter, with GLEAM more weighted in JJA and SON. For the crop/grass class, the weight differences between the models are smaller.

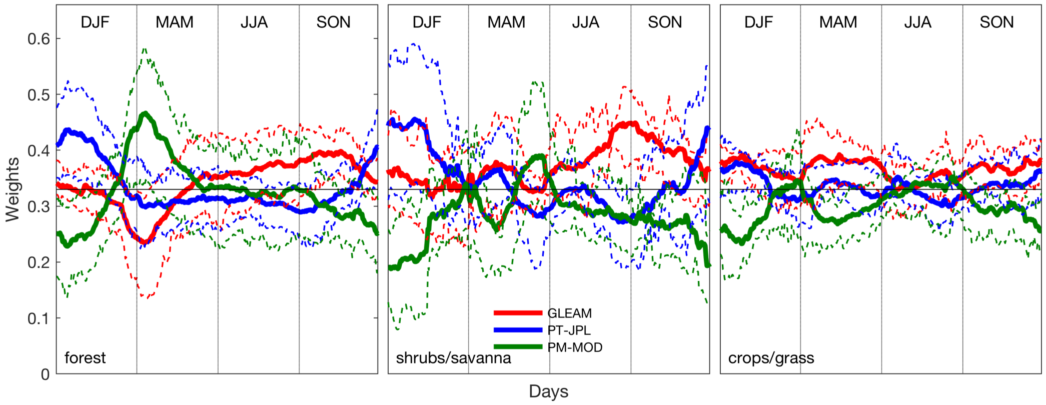

Figure 4. Daily statistics of weights over all forest (left), shrub/savanna (middle), and crop/grass (right) sites. Displayed are the mean (thick solid line), and the 25% and 75% percentile (thin dashed lines) for GLEAM (red), PT-JPL (blue), and PM-MOD (green).

Example of weights at three individual stations are given in Fig. 5. At the FR-Pue site, a Mediterranean forest located in France, the weights are not very different for the first half of the year, while for the second part GLEAM is the most weighted product. At the US-SRM site, a semi-arid grassland site in North America, for the MAM period PM-MOD is much more weighted than GLEAM and PT-JPL, while for the other periods the weight differences are smaller. The last site, the US-Ne1 cropland station situated in North America , is an example of very close weights for all models, a situation that can be observed at quite some other stations.

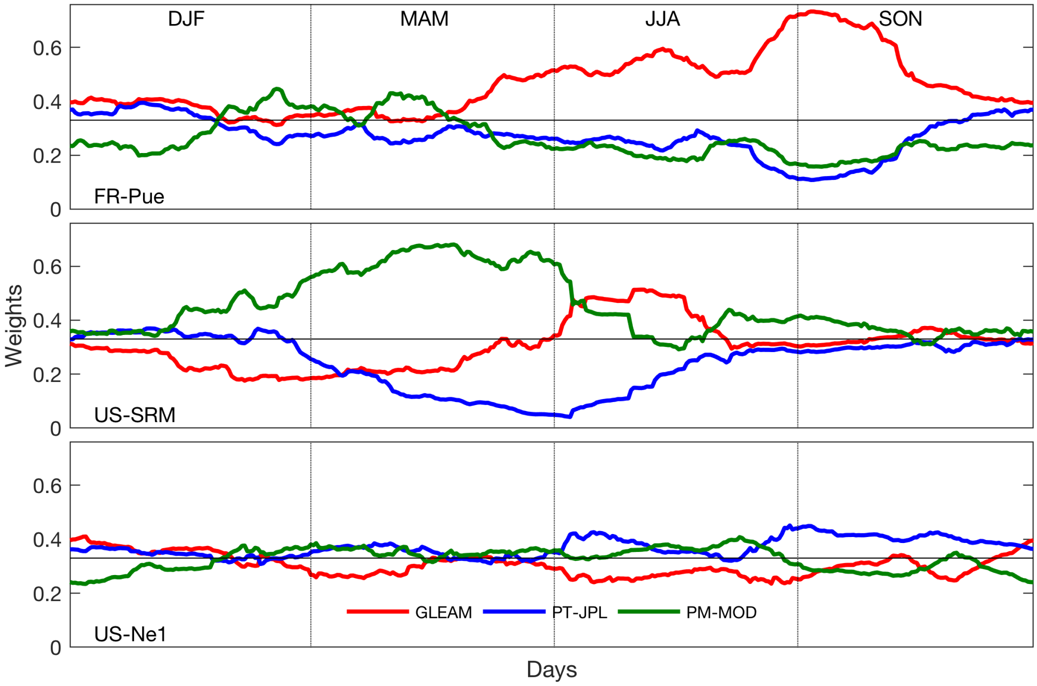

Figure 5. Example of GLEAM (red), PT-JPL (blue), and PM-MOD (green) weights at the FR-Pue (top), US-SRM (middle), and US-Ne1 (bottom) stations.

5.2 Merge products

To illustrate the merged products, time series of the original and merged products for the three sites of Fig. 5 are plotted in Fig. 6. Only 2 years are displayed to help readability. The FR-Pue site shows large inter-annual variability related to the alternance of cold and warm seasons and the availability of soil moisture (Rambal et al., 2004). All products disagree with the tower ET in 2006 for a large part of the year, while in 2007 GLEAM agrees well with the tower. The largest weight for GLEAM helps getting the WA-merge product closer to the tower, compared with the SA-merge product. The US-SRM site also shows a relatively large ET seasonal variability, with the ET tightly linked to the precipitation and associated increased in soil moisture (Scott et al., 2009). GLEAM and PT-JPL capture better than PM-MOD the sudden increase in ET values at the beginning of summer related to the rainfall coming from the North American monsoon. The merged products capture well the summer ET rise, but fail to replicate the following largest ET values as all original ET estimates are below the tower ET. This is the consequence of the merged product values always being bound by the original ET estimates, and differs from other merging approaches where the ET tower is directly regress on a set of explanatory variables (Jung et al., 2011) or ET products (Yao et al., 2017). The differences between the SA-merge and WA-merge products is smaller than at FR-Pue, consequence of the closer weights at US-SRM during the months with larger ET. The US-Ne1 is a rainfed maizesoyabean irrigated site, with a expected more regular seasonal cycle and larger ET values than the two previous sites (Verma et al., 2005). The original products have have more similar values, not capturing well the ET rise associated with start of the growing season. The closer values results in closer weights and very close SA-merge and WA-merge products.

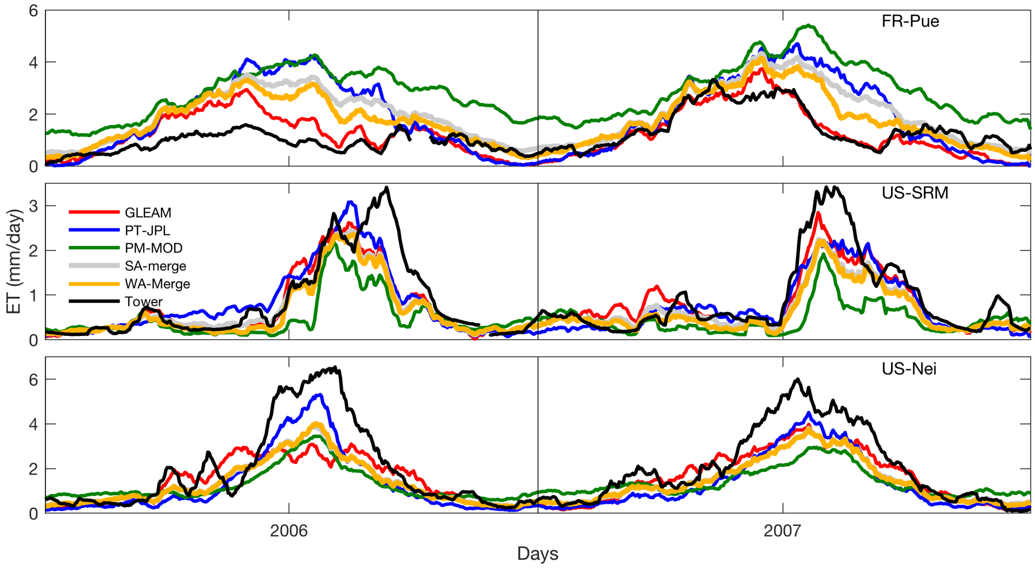

Figure 6. 2006-2007 time series of the different ET products and the tower ET at the sites FR-Pue (top), US-SRM (middle), and US-Nei (bottom). The daily values are time smoothed with a 10-day moving averaged window to better display the more persistent temporal features.

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. 7. Given the typical small weight variations presented in Sec. 5.1, the differences in performance between the SA-merge and WA-merge products are expected to be small. Note that correlations are not significant for some stations and periods, although all correlations are averaged to have a common number of stations for the inter-product and inter-season comparisons. No significant correlations are observed at a large 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.

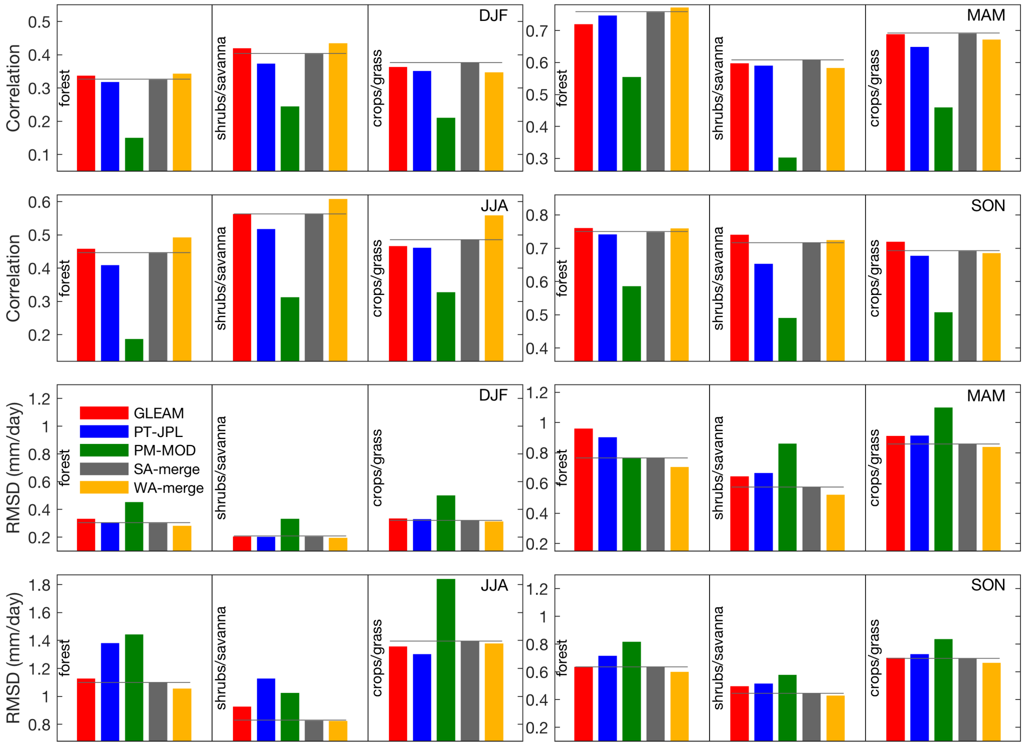

Figure 7. Season and land-cover averaged ET correlations (top two rows) and RMSD (bottom two rows) of the tower and the different products. To highlight differences with the SA-merge product, a grey line has been added to its bar. Note that the axis are not identical, but they cover similar ranges (0.5 for the correlation, 1.2 mm/day for the RMSD).

In terms of correlations the worst season is DJF, reflecting the low intra-seasonal variability in this period, while the largest correlations are observed in MAM and SON where typical vegetation greening and browning results in larger ET variability. Only in JJA the WA-merge product clearly improves the correlation of the SA-merge and original products for the three land covers. For the other seasons, the differences are smaller and sometimes the WA-merge product does not show the largest correlation with the tower ET. Concerning the RMSDs, the impact of the SA-merge product seems larger. Apart from the crop/grass cover in JJA, the SA-merge product has the smallest RMSD of all products for all seasons and land covers. However, as it was the case for the correlations, the differences in RMSD between the different products and the tower are not very large. If student-t tests are run to test the significance of the correlation differences, only for PM-MOD the other products correlate significantly higher at a large number of stations. For GLEAM, PT-JPL, and the merged products only at a very few stations the correlation increase is significant. This suggests that although the individual ET products show differences, their ET populations do not seem too distinct when compared with the tower ET, and the overall performance of the SA-merge and WA-merge across all sites do not largely differ.

6 Global product merging

6.1 Global weights

Although the performance of the local weights shown in Fig. 7 does not suggest large differences for the SA-merge and WA-merge products, they have been extrapolated by the NN described in Section 2.2.2. The resulting global weights are presented seasonally averaged in Fig. 8. As the SA-merge product equally weights all products with a value of 1/3, positive (negative) departures of the weights from that value are displayed in red (blue) to highlight the weight differences. Overall, GLEAM is the product that contributes the most to the WA-merge product, but all products have weights larger than 1/3 at some regions and seasons, suggesting that the SA-merge benefits from the merge of the three models. Some geographical patterns are visible. For instance, over the equatorial forest GLEAM and PT-JPL are more weighted than PM-MOD, a feature that persists along the year. In other regions, such as the European continent, there is a seasonal dependence of the weights, with PM-MOD less weighted than GLEAM and PT-JPL in DJF, but more weighted in JJA.

The reasons of this seasonal weight patterns are difficult to pinpoint. Errors in the weights prediction can certainly not be excluded, but some of the patterns could in principle be related to deficiencies in either the model inputs or model parameterizations. For instance, evaporation in winter at the northern latitudes is low. Later in spring when the plants start to green, the ET differences between the tower and the products can change substantially depending on whether the greening is captured by the model, resulting in very different weights. The PT-JPL and PM-MOD weights over that regions in DJF and MAM reverse the sign of their departure from 1/3, with PT-JPL (PM-MOD) more weighted in DJF (MAM). In this particular case, we could speculate that the PM-MOD vegetation inputs or parameterizations capture this vegetation development better than PM-MOD.

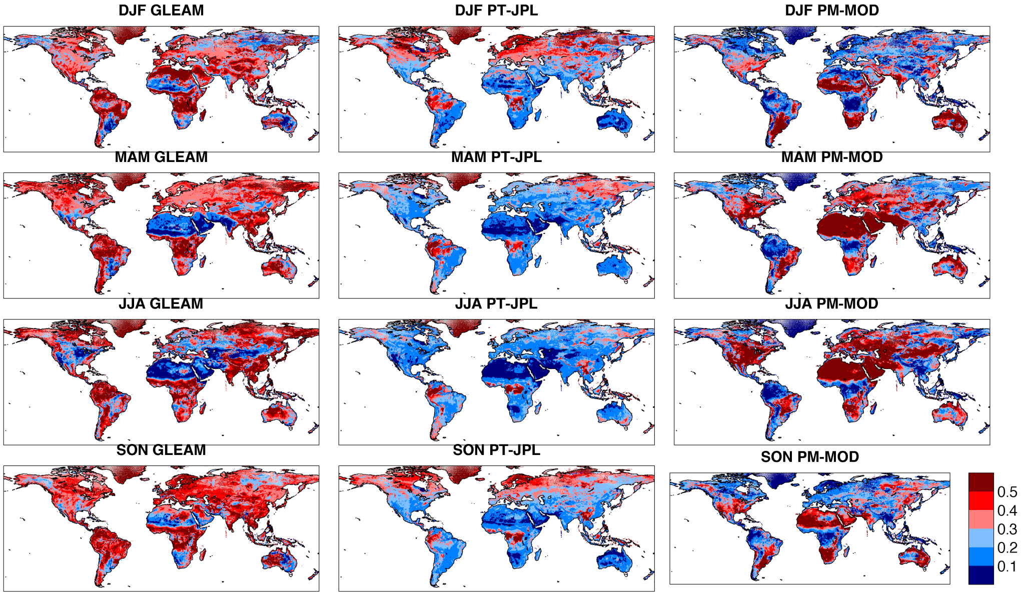

Figure 8. Seasonally averaged global weights for GLEAM (left), PT-JPL (middle), and PM-MOD (right).

6.2 Merged products

The seasonally averaged ET differences between the SA-merge and WA-merge product, normalized by the seasonal SA-merge ET are plotted in Fig. 9. Given the relatively small weight departures from the 1/3 value shown in Fig. 8, large differences between the WA-merge and SA-merge product cannot be expected. The large differences over very dry areas or the winter northern latitudes are related to the very low ET absolute value. For the remaining land, most of the relative differences are within the ±15% range. Some geographical structures are visible. For instance, many regions in North America display smaller ET for the WA-product, while the reverse is true for the equatorial regions of South America and Africa. In Europe and Asia the SA-merge and WA-merge product differences change more with season and region. Although exceptions can be found, overall there seems to be a tendency for regions with large ET to have larger values in the WA-merge product, while the reverse is true for regions with lower ET, with the WA-merge product having lower values. Independent of whether this reflects a more accurate ET estimation, this seems to show a larger dynamic range for the WA-product compared with the SA-product.

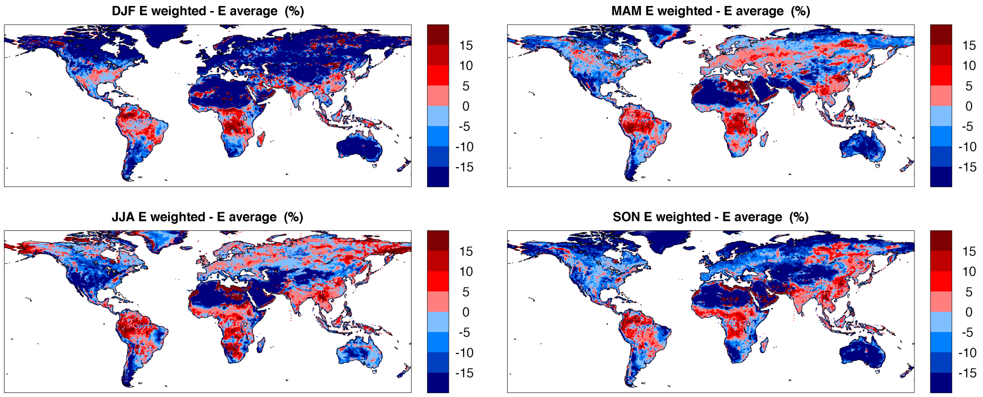
DJF E weighted - E average (%)
MAM E weighted - E average (%)
JJA E weighted - E average (%)
SON E weighted - E average (%)
15
10
5
0
-5
-10
-15

Figure 9. Seasonally averaged normalized ET differences between the SA-merge and WA-merge product, expressed as a percentage of the seasonally averaged SA-merge product ET.

7 Discussion

7.1 Inverse-variance weighting

The inverse-variance weighting is based on the differences between the model and tower ET. Factors potentially affecting the products merging are the spatial resolution mismatch between the tower and model estimates, the statistical nature of their differences, or their independence with the tower data.

The very large mismatch between the model grid cells and the footprint of the tower measurements contributes to the observed differences. The RMSD of the SA-merge product and the towers ET, normalized by the mean annual tower ET, is displayed in Fig. 10 for all the available stations, together with the station I_h described in Section 2.3. 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 close to zero. Also, significant trends were not found for the RMSD of the tower E and the original GLEAM, PT-JPL, and PM-MOD ET. 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.

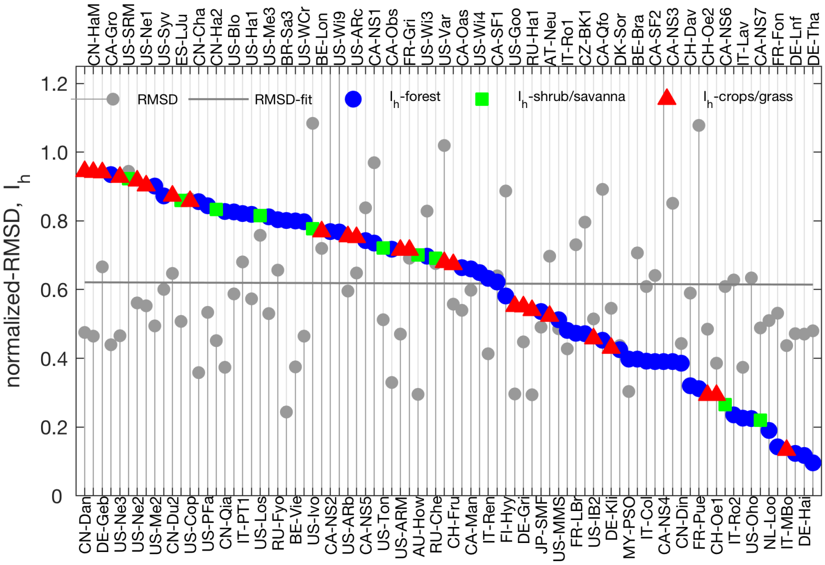

Figure 10. Homogeneity index (I_h, blue for forest stations, green for shrubs/savanna, and red for crops/grass) and RMSD of the SA-merge product and the towers ET, normalized by the mean annual tower ET (grey dots, with a linear fit plotted by a grey line). The towers are sorted from maximum to minimum I_h.

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 performance. 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. Fig. 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 the different products (see Equations 5 and 6). The ratios take values over the whole zero to one range, with only PM-MOD showing a distinctly lower median value. The medians for the other products are larger than 0.5, indicating that there is a large 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 merged products are reducing biases with the observations.

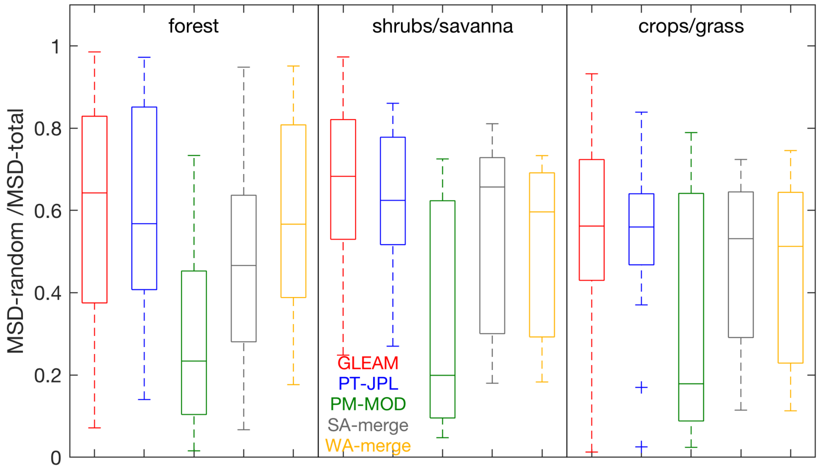

Figure 11. Box plots showing for the three land cover groups the ratio of the random MSD between the tower ET and GLEAM (red), PT-JPL (blue), PM-MOD (green), SA-merge (grey) and WA-merge (yellow) ET over the sum of the total MSD for the same pair of ET products.

The validity of Equation 1 for the WA-merge product can also be discussed. Ideally Equation 1 should be applied with unbiased and independent estimates. As discussed in Section 2.3, this is not the case for the GLEAM, PT-JPL, and PM-MOD 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 models ET dependence, it can be taken into account by modifying Equation 1 to include the correlation between the model estimates (e.g. Jones et al., 2008; Hobeichi et al., 2018). However, negative weights are now possible, and even if the sum of the weights is still one, the merged ET estimates can be outside the ET range defined by the original products. This is statistically correct, but allows the values to be extrapolated outside this range. This was tested over our stations, resulting indeed in negative weights over a large number of stations, but without producing 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 1, and does not improve the significance of the results, so it seems that ignoring the dependence between the model estimates is not the limiting factor of the merging exercise in this particular case.

7.2 Weights extrapolation

The number of stations used in this merging exercise is certainly limited in terms of covering different biome and climate conditions, so the validity of the tower data set to produce weights outside the tower space can be questioned. A firs test is presented in Fig. 12, where the correlation and RMSD between the station weights and the weights predicted by the NN is presented for two situations: (1) when all stations are included in the tower data set, i.e., the standard configuration used to produce the global WA-merge product, and; (2) when the station where the prediction will be checked is removed from the data set, i.e, the weights prediction over that station are derived using a NN that did not include that station in the training phase. Fig. 12 clearly shows that the correlation between the predicted weights and the original weights at the stations degrades notably when the station is removed from the prediction data set, implying that the global extrapolation of the weights can be quite uncertain at some regions and seasons. Cases where the correlation is large when predicted with the standard data set, but poor with the one-station-removed data set, are indicative of stations with particular conditions that are not well represented. This happens for stations such as US-Wi4 (forest with a snowy winter and warm humid summer) and CN-Dan (grasslands with a polar tundra climate). But there are also stations with very poor correlation for the all-station and one-station-removed data sets, signalling that a link between the model inputs and the error with the towers could not be found. This is the case for stations such as IT-Col (deciduous broadleaf forest with temperate climate) or MY-Pso (tropical forest), and indicates that the extrapolation of weights to areas with similar conditions will be very uncertain, even if those conditions are represented in the tower data set. Concerning the RMSD, it also degrades for the one-station-removed, although for a large number of stations the RMSD remains below 0.1, which means a relative RMSD of around 30% for the 1/3 weight value of the SA-merge product.

A further test to check the representativeness of the tower data set is conducted by globally extrapolating the weights with each of the previous 84 NNs trained without one station, and then checking the variability of the predicted weights. For the conditions well represented in the data set, it is expected that removing one station in the training will change little the weights extrapolation. But for regions not well covered the prediction problem is not well constrained, and slightly different data sets are likely to result in quite different weights. This is illustrated in Fig. 13. The displayed weight variability is calculated by estimating for each global cell the annual standard deviation of the GLEAM, PT-JPL, and PM-MOD weights, normalizing by the annual model weights, and averaging over the three models. The smallest variability in the weights coincides with the regions where the database is more representative, namely US, Central Europe, and some parts of China, possibly indicating a bias in the tower data set linked to the specific location of the towers selected. The variability in tropical regions, where only 3 stations are part of the database, is in general larger than for the previous regions. The largest variability occurs over the very dry regions, a regime poorly represented in the tower data set as shown in Fig. 1. While a poor extrapolation of weights is not critical over very dry regions, given their low E values, uncertain weights over the very humid regions is more of a concern due to their typically large E values and their significance in the total figures of global E.

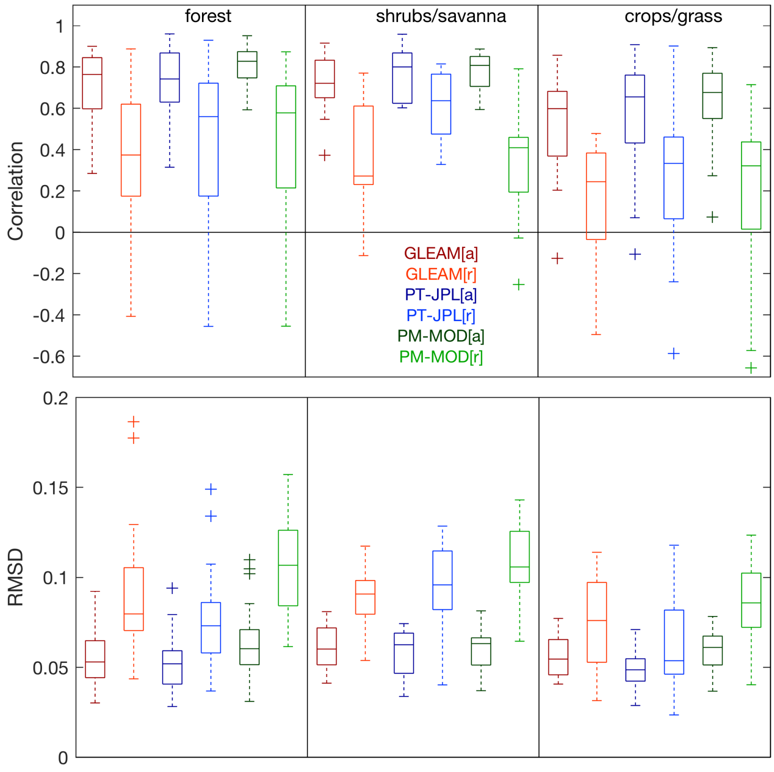

Figure 12. Box plot showing for the three land cover groups the correlation (top) and RMSD (bottom) between the station local weights and the weights predicted by the NN when all stations are included in the tower data set ("a" in legend, dark colours), and when the station where the prediction will be checked is removed from the data set ("r" in legend, light colours). See the text for details.

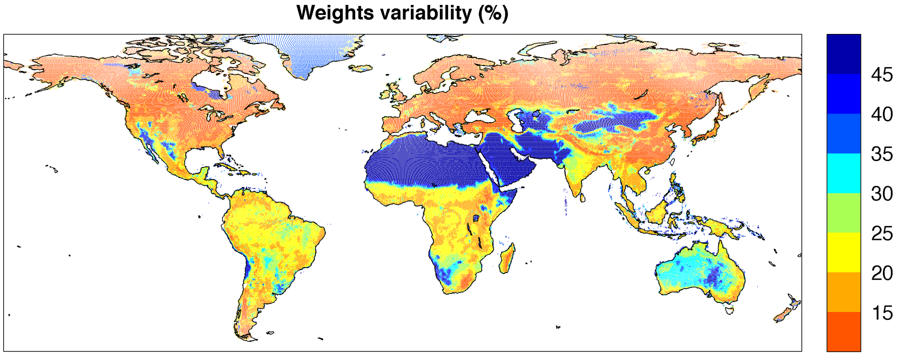

Figure 13. Relative annual variability of the global weights extrapolated by 84 different NNs. See the text for details.

7.3 Quality of the merged products

Evaluations of ET products are typically conducted by comparing at point scale with tower fluxes, or at much larger spatial scales by deriving spatially integrated estimates from related data sets, such as precipitation (P) and river run-off (Q). As the towers are used to derive the merge products, the alternative for an independent assessment of the merged products is to conduct catchments mass balance analyses, similar to those presented in Miralles et al. (2011).

The mass balance of a catchment implies that the space and time integration of P-Q equals ET integrated over the same space and time. Here, the 2002-2007 ET estimates from GLEAM, PT-JPL, PM-MOD, and the merged products are averaged in time to produce an annual map, followed by the spatial integration producing an ET annual value per basin. The basin P-Q estimate is calculated for the Q and two P products (MSWEP and WordlClim) described in Section3.3, and only if the P-Q data record is available for a minimum of 3 years in the 2002-2007 period, to assure some common period between ET and P-Q. To reduce noise in the basin-integrated ET estimates, only basins with a catchment area containing at least 3x3 cells of the 25 km resolution gridded estimates are included in the comparison. This leaves 685 basins, with $\sim$75 % of the basins situated in the Northern hemisphere, showing a similar geographical bias as the tower data set. There are further divided into three groups of 243, 295, and 147 basins based on an aridity index (AI, basin potential ET over the basin P) taking values in the intervals AI<1, 1<AI<2, and AI>2.

Scatter plots showing the correspondence between P-Q and ET are given in Fig. 14. Linear fits for the three AI classes are plotted, and the slope of the fits, the correlation, and the RMSD given in the plot. Overall, the statistics of the the water balance comparison using MSWEP or WordlClim as P are close. From the original products, PM-MOD shows the worst agreement with P-Q. GLEAM agrees better than PT-JPL for the driest and wettest basins, specially for the driest ones (AI<1, correlations of 0.93 (MSWEP) 0.88 (WorldClim) for GLEAM, and 0.74 (MSWEP) and 0.69 (WorldClim) for PT-JPL), while PT-JPL agrees slightly better than GLEAM for the 1<AI<2, but this time with much closer correlations. The SA-merge product shows close statistics to GLEAM and PT-JPL, so adding the PM-MOD product does not largely degrade the skill to close the water catchment budget. Regarding a comparison between the SA-merge and WA-merge, RSMDs and correlation are very close. Larger differences are observed only for the slope of the linear fits, where the WA-merge shows slopes closer to the expected 1:1 rate of change than the remaining products. This suggests some skill of the WA-merge product in terms of better closing the water catchment budget for this specific period and selected basins, although the differences with the SA-merge are not large.

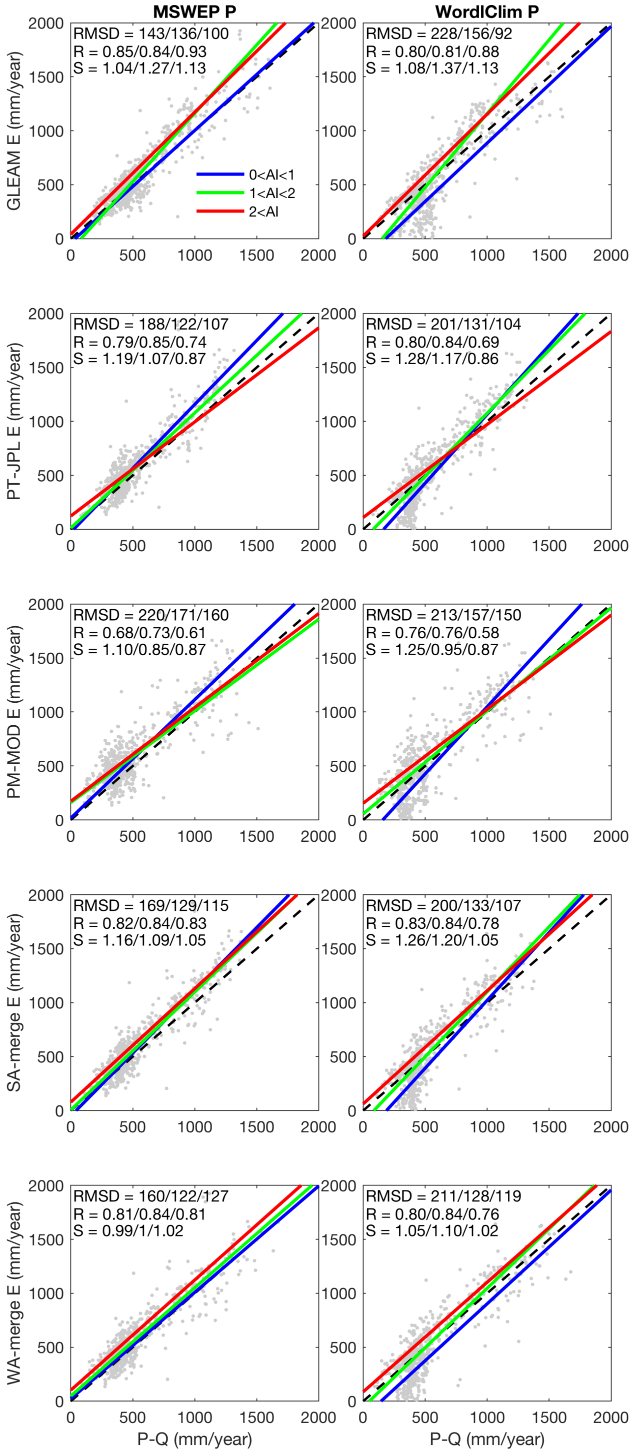

Figure 14. Scatter plots of P-Q and ET from the different products. Linear fits for three AI classes are plotted, together the slope of the fits, the correlation, and the RMSD. From left to right, the statistics are given for AI<1 (blue line), 1<AI<2 (green), and AI>2 (red), i.e., from wet to dry basins.

8 Conclusions

An inverse-variance weighting of the three global ET products run during the WACMOS-ET project (WA-merge) is presented. To test the merge, three ET models, GLEAM, PT-JPL, and PM-MOD, are forced with some common daily inputs at a resolution of 25 km for the period 2002-2007. GLEAM and PT-JPL share some common features in their modeling framework, such as the Priestley-Taylor formulation to estimate the potential evaporation rate, followed by a conversion into actual volumes of evaporation using model-specific formulations of evaporative stress, while PM-MOD uses a more different modelling approach based on a Penman-Monteith formulation. 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 modelled ET for non-rainy conditions. To produce dynamic weights changing seasonally, they are estimated by a running window of 31 days centred at each day of the year. The local weight estimation is followed by a regression of the weights on the main model inputs in order to derive weights outside the tower space and produce a global merged product. The assumption of the three ET products being equally uncertain, i.e., a simple average assigning a constant weight of 1/3 to each product (SA-merge), is also tested.

The local weights over some stations show seasonal patterns, but the differences from 1/3 were not very large at many stations. The closest weights are observed over the cropland and grassland stations. Stations where the weights are close are characterized by a variance of the inter-differences of the original products smaller than the variance of the difference between the tower ET and the original products. This implies that even if GLEAM, PT-JPL, and PM-MOD are shown to differ, they are still relatively close, compared with the tower ET, at a large number of the selected sites. For stations where the original products are more distinct, their weighting results in a WA-merge product more different from the SA-merge product. When averaged over all stations, seasonal correlations between the tower ET and the SA-merge and WA-merge products only show clear larger values for the boreal summer. For the other seasons the correlations are more comparable. Results are more positive regarding RMSDs, where apart from the cropland/grassland land cover in the boreal summer, the SA-merge product has the smallest RMSD of all products for all seasons and land covers.

The globally extrapolated weights show seasonal and regional departures from the 1/3 value of the SA-merge, with overall GLEAM being the product that contributes the most to the WA-merged product. The weight differences can be related to deficiencies in model parameterizations or inputs, but can also come from errors in the weight prediction. The weight patterns result in seasonal differences of the global SA-merge and WA-merge product. At a large number of regions these differences are confined to the ± 15 relative range, indicating that the SA-merge and WA-merge are relatively close, as it was also observed for the local weights. If the merged products are compared with ET inferred from the difference between basin-integrated precipitation and river run-off, correlations and RMSDs of the annual ET values are close for the SA-merge and WA-merge products. Only the slopes of the linear fit are closer to one for the WA-merge product, compared with the other products, suggesting some skill of the WA-merge product in terms of better closing the water catchment budget for this specific period and selected basins.

The tower data set is certainly limited in terms of the biome and climate conditions represented, with most of the 84 selected stations located at temperate regions. This is apparent when the weights prediction is tested over individual stations with the prediction data set not including the concerned station. Correlations and RMSDs between the station original and predicted weights can largely degrade at some stations, compared with the prediction using the whole tower data set, pointing to the limitations of the data set to extrapolate the weights outside the tower locations. This is further confirmed by reproducing the global weights with the tower data sets missing one station and observing that the largest weight variability happens over the regions where the towers are less represented. Tn these less covered regions many towers only have data records for a limited number of years, so longer ET product data records need to be used to give access to some of these stations, extending the tower data set to some of the less represented regions.

Another limiting factors for the merging exercise is the mismatch between the towers and the products spatial resolution. For our selected towers test conducted to search for an apparent link between tower surrounding spatial homogeneity and magnitude of the errors were not successful. Nevertheless, it is clear that the impact of spatial resolution mismatch is minimized if the modelled ET is available at finer spatial resolutions. Therefore, modeling ET at a finer spatial resolution should help for future tower-based merging of the estimates.

In this study the GLEAM, PT-JPL, and PM-MOD products are derived with common data sets for their shared inputs. This is key to study inter-product differences related to the modeling components, one of the initial objectives of the WACMOS-ET project, but it eliminates the more common situation of also having inter-product differences associated to applying the models with different inputs. For this specific observation period and available towers, it is likely that the unique input data sets introduced common error patterns with respect to the tower observations. The relatively close modeling framework for GLEAM and PT-JPL could also be partially responsible for the common error patterns observed between these two products. The result is that, even if ET differences between the individual products are observed, they are not too distinct when compared with the tower observations to clearly weight the ET products very differently at many sites. Therefore, it is expected that tower observations could be more informative if merging more independent ET products.

Overall, 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. As previously commented, it is likely that more distinct error patterns are observed if the model inputs are more independent, leading to a more informative weighting of the products and more added value to a weighted product, compared with just the simple product average. For this to happen, efforts from the ET developers to publicly made available these products for more extended periods, and if possible at finer spatial resolutions, are required to more effectively used a larger pool of tower data and continue work in this direction.

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.: Measuring boreal forest evapotranspiration using the energy balance residual, Journal of Hydrology, 366, 112–118, 2009.

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, 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.

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., De Roo Journal of, A., and van Dijk, A. I. J. M.: Global maps of streamflow characteristics based on observations from several thousand catchments, journals.ametsoc.org, 16, 1878–1501, 2015.

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&amp;deg; global gridded precipitation (1979&amp;ndash;2015) by merging gauge, satellite, and reanalysis data, Hydrology and Earth System Sciences, 21, 589–615, 2017a.

Beck, H. E., Vergopolan, N., Pan, M., Levizzani, V., van Dijk, A. I. J. M., Weedon, G. P., Brocca, L., Pappenberger, F., Huffman, G. J., and Wood, E. F.: Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling, Hydrology and Earth System Sciences, 21, 6201–6217, 2017b.

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 . . . , 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.

Fick, S. E. and Hijmans, R. J.: WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas, International Journal of Climatology, 37, 4302–4315, 2017.

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.: The energy balance closure problem: an overview, Ecological Applications, 18, 1351–1367, 2008.

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.

Hagan, M. T. and Menhaj, M.: Training feedforward networks with the Marquardt algorithm, IEEE Trans. Neural Networks, 5, 989–993, 1994.

Hirschi, M., Michel, D., Lehner, I., and Seneviratne, S. I.: A site-level comparison of lysimeter and eddy covariance flux measurements of evapotranspiration, Hydrology and Earth System Sciences, 21, 1809–1825, 2017.

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

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.

Jung, M., Reichstein, M., Margolis, H. A., Cescatti, A., Richardson, A. D., Arain, M. A., Arneth, A., Bernhofer, C., BONAL, D., Chen, J., Gianelle, D., Gobron, N., Kiely, G., Kutsch, W., Lasslop, G., Law, B. E., Lindroth, A., Merbold, L., Montagnani, L., Moors, E. J., Papale, D., Sottocornola, M., Vaccari, F., and Williams, C.: Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations, Journal of Geophysical Research, 116, G00J07–16, 2011.

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 . . . , 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.

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 &ndash; 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., 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, 2011.

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, http://dx.doi.org/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.

Mu, Q., Zhao, M., and Running, S. W.: Improvements to a MODIS global terrestrial evapotranspiration algorithm, Remote Sensing of Environment, 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, Geophysical Research Letters, 38, n/a–n/a, 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. F. A. S. S. C. P. F. P. P. M. 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, pp. 1–17, 2014.

Nguyen, D. and Widrow, B.: Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptative weights, in: Proceedings of the 1990 International Joint Conference on Neural Networks, pp. 21–26, 1990.

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 CO<sub>2</sub> 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.

Rambal, S., Joffre, R., Ourcival, J. M., Cavender-Bares, J., and Rocheteau, A.: The growth respiration component in eddy CO2 flux from a Quercus ilex mediterranean forest, Global Change Biology, 10, 1460–1469, 2004.

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 CO2 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, G04 004, 2009.

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 spaceborne radiometer data and ground-based measurements, Remote Sensing of Environment, 115, 3517–3529, 2011.

Twine, T. E., Kustas, W. P., Norman, J. M., Forest, D. C. A., , and 2000: Correcting eddy-covariance flux underestimates over a grassland, Elsevier, 103, 279–300, 2000.

Verma, S. B., Dobermann, A., Forest, K. C. A., , and 2005: Annual carbon dioxide exchange in irrigated and rainfed maize-based agroecosystems, Elsevier, 131, 77–96, 2005.

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.

Wilson, K., Goldstein, A., Falge, E., Forest, M. A. A., , and 2002: Energy balance closure at FLUXNET sites, Elsevier, 113, 223–243, 2002.

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.