Biogeosciences

Water, Energy, and Carbon with Artificial Neural Networks (WECANN): a statistically based estimate of global surface turbulent fluxes and gross primary productivity using solar-induced fluorescence

Seyed Hamed Alemohammad[1,2], Bin Fang[1,2], Alexandra G. Konings[3], Filipe Aires[1,4], Julia K. Green[1,2], Jana Kolassa[5,6], Diego Miralles[7], Catherine Prigent[1,6], and Pierre Gentine[1,2,8]

[1]Department of Earth and Environmental Engineering, Columbia University, New York, 10027, USA
[2]Columbia Water Center, Columbia University, New York, 10027, USA
[3]Department of Earth System Science, Stanford University, Stanford, 94305, USA
[4]Observatoire de Paris, Paris, 75014, France
[5]Universities Space Research Association/NPP, Columbia, MD, 21046, USA
[6]Global Modeling and Assimilation Office, NASA Goddard Spaceflight Center, Greenbelt, MD, 20771, USA
[7]Laboratory of Hydrology and Water Management, Ghent University, Ghent, 9000, Belgium
[8]Earth Institute, Columbia University, New York, 10027, USA

Correspondence to: Pierre Gentine (pg2328@columbia.edu)

Abstract. A new global estimate of surface turbulent fluxes, latent heat flux (LE) and sensible heat flux (H), and gross primary production (GPP) is developed using a machine learning approach informed by novel remotely sensed solar-induced fluorescence (SIF) and other radiative and meteorological variables. This is the first study to jointly retrieve LE, H, and GPP using SIF observations. The approach uses an artificial neural network (ANN) with a target dataset generated from three independent data sources, weighted based on a triple collocation (TC) algorithm. The new retrieval, named Water, Energy, and Carbon with Artificial Neural Networks (WECANN), provides estimates of LE, H, and GPP from 2007 to 2015 at $1° \times 1°$ spatial resolution and at monthly time resolution. The quality of ANN training is assessed using the target data, and the WECANN retrievals are evaluated using eddy covariance tower estimates from the FLUXNET network across various climates and conditions. When compared to eddy covariance estimates, WECANN typically outperforms other products, particularly for sensible and latent heat fluxes. Analyzing WECANN retrievals across three extreme drought and heat wave events demonstrates the capability of the retrievals to capture the extent of these events.

Uncertainty estimates of the retrievals are analyzed and the interannual variability in average global and regional fluxes shows the impact of distinct climatic events – such as the 2015 El Niño – on surface turbulent fluxes and GPP.

1 Introduction

Turbulent fluxes from the land surface to the atmosphere, particularly sensible heat flux (H) and latent heat flux (LE), and plant carbon uptake characterized by gross primary production (GPP) are key to understanding ecosystem response to climate and the feedback on the overlying atmosphere, as well as constraining the global carbon, water, and energy cycles. In recent years, there has been substantial effort towards estimating these variables from remote sensing observations on a global scale (see, e.g., Fisher et al., 2008; Jiang and Ryu, 2016; Jiménez et al., 2009, 2011; Jung et al., 2009; Miralles et al., 2011a; Mu et al., 2007; Mueller et al., 2011). Two typical approaches have been used to estimate these from remote sensing information. The first approach uses physically based or semiempirical models (e.g., the Priestley–Taylor or

Penmann–Monteith equation in the case of evapotranspiration (ET), or a light-use efficiency model in the case of GPP) informed by remote sensing information (e.g., vegetation indices, infrared temperature, microwave soil moisture), often in combination with reanalysis meteorological forcing data (Fisher et al., 2008; Miralles et al., 2011a; Mu et al., 2007; Zhang et al., 2016b; Zhao et al., 2005; Zhao and Running, 2010). These approaches are sensitive to the assumptions and imperfections of the underlying flux models. The second approach, uses machine learning (e.g., a model tree ensemble) to determine LE, H, and GPP from meteorological drivers and optical remote sensing data (Tramontana et al., 2016). Like all supervised machine learning models, this approach relies on a training dataset to determine the nonlinear statistical relationships. In this case, in situ turbulent flux and GPP estimates from eddy covariance towers are used (Beer et al., 2010; Jung et al., 2011). Such an approach relies implicitly on an assumption that a long temporal record of these variables at a small number of sites captures the full range of behavior and sensitivities of terrestrial ecosystems around the globe. In addition, extreme and therefore rare events may be difficult to capture based on the limited data availability.

Alternatively, one can use a machine learning approach, such as an artificial neural network (ANN), trained on globally representative but imperfect estimates of the fluxes (such as those from models) to parameterize the nonlinear statistical relationships between remote sensing observations and surface fluxes. This approach has been successfully used for global soil moisture retrieval (Aires et al., 2012; Kolassa et al., 2013, 2016; Rodriìguez-Fernández et al., 2015) and surface heat flux retrieval (Jiménez et al., 2009). Such ANNs require a target dataset for training. Climate model simulations of the relevant geophysical variable are usually used as the training dataset to facilitate subsequent data assimilation efforts (Aires et al., 2012; Kolassa et al., 2013, 2016). However, the downside of this approach is that the resulting fluxes estimated by the ANN often exhibit some of the same biases as the simulations used to train the network (Rodriìguez-Fernández et al., 2015), even if improvements can be achieved such as a more realistic seasonal cycle as it is informed by the seasonal cycle of the remote sensing data (Jiménez et al., 2009).

Previous studies show a strong relationship between the rate of photosynthesis and solar-induced fluorescence (SIF) observations and indicate that the plant fluorescence measurements can be a useful proxy for photosynthesis estimation (Flexas et al., 2002; Govindjee et al., 1981; Havaux and Lannoye, 1983; van Kooten and Snel, 1990; Krause and Weis, 1991; McFarlane et al., 1980; Toivonen and Vidaver, 1988; van der Tol et al., 2009). Recently, satellite observations of SIF have become available, opening new possibilities for the global monitoring of photosynthesis (Frankenberg et al., 2011, 2012, 2014; Guanter et al., 2012; Joiner et al., 2013; Schimel et al., 2015; Xu et al., 2015).

SIF observations from the Global Ozone Monitoring Experiment–2 (GOME-2) instrument are shown to better track the seasonal cycle of GPP compared to typical high-resolution optically based vegetation index estimates (Guanter et al., 2012, 2014; Joiner et al., 2014; Walther et al., 2016). SIF has also been shown to be a pertinent indicator of vegetation water stress (Lee et al., 2013). Moreover, a near-linear relationship between monthly SIF retrievals and GPP is found for different vegetation types, which suggests that SIF estimates can strongly constrain GPP retrievals (Frankenberg et al., 2011).

Recently, a new SIF product was developed from observations of the GOME-2 satellite using a new retrieval algorithm that disentangles three components from multispectral observations (Joiner et al., 2013). SIF retrievals are shown not to be strongly affected by cloud contamination and seasonal variabilities in aerosol optical depth (Frankenberg et al., 2014). More recently, remotely sensed SIF retrievals have been used to successfully provide estimates of GPP in cropland and grassland ecosystems (Guanter et al., 2014; Zhang et al., 2016a). SIF retrievals are also integrated with photosynthesis estimates from the National Center for Atmospheric Research Community Land Model version 4 (NCAR CLM4), which result in significant improvement of the photosynthesis simulation (Lee et al., 2015). As GPP relates to plant transpiration through stomata regulation (Damour et al., 2010; DeLucia and Heckathorn, 1989; Dewar, 2002), and transpiration water fluxes dominate continental ET (Jasechko et al., 2013), the use of remotely sensed SIF has the potential to also better constrain estimates of the continental water (LE) and energy (H) cycles, in addition to the carbon (GPP) cycle.

In this study, we develop an ANN approach to retrieve monthly estimates of LE, H, and GPP on a global scale. The network uses remotely sensed SIF estimates in addition to other data including precipitation, temperature, soil moisture, snow cover, and net radiation as inputs (predictor). To our knowledge, this is the first study that uses remotely sensed SIF estimates on a global scale to retrieve LE and H surface turbulent fluxes along with GPP.

Moreover, to reduce any errors, we introduce a Bayesian perspective to generate the target dataset for the ANN. Multiple estimates of LE, H, and GPP are selected according to an a priori probability that reflects the quality and information content of the dataset at the particular pixel of interest (details are provided in Sect. 3.2). This approach enables us to generate a robust target dataset for remote sensing observations along with a statistical algorithm for the retrieval, while bypassing the need for a land surface model and radiative transfer scheme. We use the triplet of GLEAM, ECMWF, and FLUXNET-MTE (Multi-Tree Ensemble) for training of LE and H and the triplet of MODIS-GPP, ECMWF, and FLUXNET-MTE for GPP training.

This new global product is named WECANN (Water, Energy, and Carbon Cycle with Artificial Neural Networks).

WECANN monthly estimates for the period 2007–2015 are provided on a $1° \times 1°$ resolution grid and with units of $W\,m^{-2}$ for LE and H and $gC\,m^{-2}\,day^{-1}$ for GPP. The spatial coverage of WECANN is presented in Fig. S1 in the Supplement. It includes all the land areas, except for Greenland, Antarctica, and any $1° \times 1°$ pixel that permanently has more than 75 % water, snow, or ice. To estimate the fraction of water, snow, and ice in each pixel, we used the $0.05° \times 0.05°$ MODIS-based Land Cover Type product (MCD12C1 v051) (NASA LP DAAC, 2016).

2 Data

The inputs of WECANN include six remotely sensed variables introduced in Sect. 2.2 and Table 2: SIF, net radiation, air temperature, soil moisture, precipitation, and snow water equivalent (SWE). These are used to retrieve LE, H, and GPP. Different observation- and/or model-based datasets are used as the training dataset and are explained in Sect. 2.1 and summarized in Table 1. All the data presented here are projected and gridded on a $1° \times 1°$ geographic grid and averaged at monthly temporal resolution. Finally, independent datasets used for evaluation of the ANN retrievals are presented in Sect. 2.3.

2.1 Training datasets

Four products are introduced in this section, and a triplet of them is used for training of each of the LE, H, and GPP (Sect. 3.2). For LE and H, training is performed based on GLEAM, FLUXNET-MTE, and ECMWF ERA HTESSEL (Hydrology Tiled ECMWF Scheme for Surface Exchanges over Land). For GPP, training is performed on FLUXNET-MTE, ECMWF ERA HTESSEL, and MODIS-GPP. Table 1 summarizes the characteristics of the training datasets used here.

2.1.1 GLEAM

The Global Land Evaporation Amsterdam Model (GLEAM) is a set of algorithms to estimate terrestrial evapotranspiration using satellite observations (Martens et al., 2017; Miralles et al., 2011a). GLEAM is a physically based model composed of (1) a rainfall interception scheme, driven by rainfall and vegetation cover observations; (2) a potential evaporation scheme, calculated from the Priestley and Taylor (1972) equation and driven by satellite observations; and (3) a stress factor attenuating potential evaporation, based on a semiempirical relationship between microwave vegetation optical depth (VOD) observations and root zone soil moisture estimates (based on a running water balance for rainfall and assimilating satellite soil moisture). The data are provided at a $0.25° \times 0.25°$ spatial resolution and daily temporal resolution and start in 1980. GLEAM data have been used for studying land–atmosphere interactions and the global wa-

ter cycle (Guillod et al., 2014, 2015, Miralles et al., 2011a, 2014a, b). In this study, we use LE and H estimates from the latest version v3.0a (Martens et al., 2017).

2.1.2 FLUXNET-MTE

The FLUXNET-MTE provides global surface fluxes at $0.5° \times 0.5°$ spatial resolution derived from empirical upscaling of eddy covariance measurements from the FLUXNET global network (Baldocchi et al., 2001). The MTE method used is an ensemble learning algorithm that enables the learning of a diverse sequence of different model trees by perturbing the base learning algorithm (Jung et al., 2009, 2010, 2011). The data cover the period from January 1982 to December 2012 and can be used for benchmarking land surface models and assessment of biosphere gas exchange. We use LE, H, and GPP estimates from FLUXNET-MTE.

2.1.3 ECMWF ERA HTESSEL

The ECMWF Reanalysis (ERA) is a global 3-D variational data assimilation (3D-Var) product that uses HTESSEL in the forecast system. HTESSEL has a surface runoff component and accounts for a global nonuniform soil texture unlike the old TESSEL model (Balsamo et al., 2009). This is an offline model simulation, and HTESSEL is driven by meteorological forcing output from the forecast runs. Photosynthesis in the model is computed independently (i.e., with its own canopy conductance) from LE, so that the carbon cycle does not interact with the water cycle at the stomata level, adding errors. We use LE, H, and GPP estimates from ERA HTESSEL provided on a $0.25° \times 0.25°$ geographic grid with daily temporal resolution.

2.1.4 MODIS-GPP

The MODIS sensor is onboard the sun-synchronous NASA satellites Terra (10:30 LT overpasses) and Aqua (13:30 LT overpasses). It provides 44 global data products (Justice et al., 2002) from 36 spectral bands, including visible, infrared, and thermal infrared spectrums to monitor and understand Earth's surface: atmosphere, land, and ocean processes. The MODIS GPP/NPP project (MOD17) provides gross and net primary production estimates covering the whole land surface and is useful for analyzing the global carbon cycle and monitoring environmental change. The MOD17 algorithm is based on a light-use efficiency approach proposed by Monteith and Moss (1977), which states that GPP is proportional to the product of incoming photosynthetically active radiation (PAR), fraction of absorbed PAR (fAPAR), and efficiency of radiation absorption in photosynthesis. We use the monthly MOD17A2 GPP product (Running et al., 2004; Zhao et al., 2005; Zhao and Running, 2010). MOD17A2 is available from 2000 until 2015 and was provided at a $0.05° \times 0.05°$ spatial resolution.

Table 1. Characteristics of products used for training of ANN.

Product	Output variables used for training	Temporal coverage	Spatial coverage	Temporal resolution	Spatial resolution	Reference
GLEAM	LE, H	1980–2015	Global	Daily	$0.25° \times 0.25°$	Martens et al. (2017)
ECMWF ERA HTESSEL	LE, H, GPP	2008–2015	Global	Daily	$0.25° \times 0.25°$	Balsamo et al. (2009)
FLUXNET-MTE	LE, H, GPP	1982–2012	Global	Monthly	$0.5° \times 0.5°$	Jung et al. (2009)
MODIS-GPP	GPP	2000–2015	Global	Monthly	$0.5° \times 0.5°$	Running et al. (2004)

Table 2. Characteristics of observations used as input in the WECANN product.

Variable	Product name and version	Temporal coverage	Spatial coverage	Temporal resolution	Spatial resolution	Reference
SIF	GOME-2 Fluorescence v26	2007–present	Global	Daily	$0.5° \times 0.5°$	Joiner et al. (2013)
Net radiation	CERES L3 SYN 1deg	2002–present	Global	Monthly	$1° \times 1°$	Wielicki et al. (1996)
Air temperature	AIRS3STD v6.0	2002–present	Global	Daily	$1° \times 1°$	Aumann et al. (2003)
Soil moisture	ESA-CCI v2.3	1978–2015	Global	Daily	$0.25° \times 0.25°$	Liu et al. (2012)
Precipitation	GPCP 1DD v1.2	1996–2015	Global	Daily	$1° \times 1°$	Huffman et al. (2001)
Snow water Equivalent	GlobSnow L3A v2	1979–present	Global	Daily	$25\,\mathrm{km} \times 25\,\mathrm{km}$	Luojus et al. (2013)

2.2 Input datasets

Six sets of observations are used as input to the WECANN retrieval algorithm. These are selected in a way that provides necessary physical constraints on the estimates from the ANN. Table 2 lists the characteristics of each of the datasets, and they are briefly introduced in the following.

2.2.1 Solar-induced fluorescence

The GOME-2 instrument is an optical spectrometer onboard the Meteorological Operational Satellite Program (MetOp-A and MetOp-B) satellites, which were launched by the ESA. GOME-2 was designed to monitor atmospheric ozone profiles as well as other trace gases and water vapor content. It senses Earth backscatter radiance and solar irradiance at a $40 \times 40\,\mathrm{km}$ spatial resolution (prior to July 2013 the spatial resolution was $40 \times 80\,\mathrm{km}$). Recently, the retrieval of SIF using GOME-2 observations in the 650–800 nm spectrum has been investigated (Joiner et al., 2013, 2016). We use version 26 of the daily SIF product that uses the MetOp-A GOME-2 channel 4 with a ~ 0.5 nm spectral resolution and wavelengths between 734 and 758 nm. SIF estimates are provided on a geographic grid with $0.5° \times 0.5°$ grid spacing.

2.2.2 Net radiation

Net radiation is the main control of the rates of sensible and latent heat in wet environments and is closely related to PAR. The Clouds and Earth's Radiant Energy System (CERES) is a suite of instruments that measure radiometric properties of solar-reflected and Earth-emitted radiation from the top

of the atmosphere to Earth's surface, from three broadband channels at 0.3–100 µm. The CERES sensors are onboard the Earth Observing System (EOS), which includes Terra, Aqua, and Tropical Rainfall Measuring Mission (TRMM) (Kato et al., 2013; Loeb et al., 2009). We use the net radiation estimates, which are provided on a $1° \times 1°$ geographic grid with monthly time resolution, from the Synoptic Radiative Fluxes and Clouds (SYN) product of CERES.

2.2.3 Air temperature

The Atmospheric Infrared Sounder (AIRS) is a high-spectral-resolution spectrometer onboard the NASA Aqua satellite launched in 2002. It provides hyperspectral (visible and thermal infrared) observations for monitoring process changes in the Earth's atmosphere and land surface, as well as for improving weather prediction. The AIRS instrument was designed to obtain atmospheric temperature and humidity profiles of every 1 km layer of the atmosphere. The accuracy of AIRS temperature observations is typically better than $1\,°\mathrm{C}$ in the lower troposphere under clear sky conditions (Aumann et al., 2003). We use daily temperature estimates from the lowest layer of the AIRS level 3 standard product that is provided on a $0.5° \times 0.5°$ geographic grid.

2.2.4 Surface soil moisture

The ESA Climate Change Initiative (CCI) program soil moisture (ESA CCI SM) is a multi-decadal (1980–2015) global satellite-observed surface soil moisture product. It merges observations from passive sensors (e.g., Scanning Multichannel Microwave Radiometer (SMMR), Special Sen-

sor Microwave/Imager (SSM/I), AMSR-E) and active ones (e.g., the European Remote Sensing, ERS; Advanced Scatterometer, ASCAT), based on a triple collocation (TC) error characterization (Dorigo, et al., 2017; Liu et al., 2011, 2012; Wagner et al., 2012). Here, we use daily data from the latest version, v2.3. ESA CCI SM is provided on a $0.25° \times 0.25°$ geographic grid.

2.2.5 Precipitation

The Global Precipitation Climatology Project (GPCP) provides global daily precipitation estimates at $1° \times 1°$ spatial resolution from October 1996 to near present (Huffman et al., 2001). Global precipitation estimates from infrared and microwave instruments are combined with monthly gauge measurements to produce the daily estimates. In this study, v1.2 of the One-Degree Daily (1DD) product of GPCP is used and daily estimates are aggregated to monthly scales. Several studies have evaluated the GPCP 1DD product on global or regional scales, and results show that it has high accuracy and good agreement with independent in situ measurements and other global precipitation estimates (Gebremichael et al., 2005; Joshi et al., 2012; McPhee et al., 2005; Rubel et al., 2002).

2.2.6 Snow water equivalent

The GlobSnow project is developed by ESA and provides long-term snow-related variables: snow water equivalent (SWE) and areal snow extent (SE). It combines microwave-based retrievals of snow information (including Nimbus-7 SMMR, DMSP F8/F11/F13/F17 SSM/I(S) observations) and ground-based station data through a data assimilation process and provides the SWE and SE products at different temporal resolutions: daily, weekly, and monthly (Pulliainen, 2006). Here, we use v2 of the daily L3A SWE product, which is posted on a $25\,km \times 25\,km$ Equal-Area Scalable Earth Grid (EASE).

2.3 Evaluation datasets

2.3.1 Eddy covariance tower estimates

FLUXNET is a network of regional tower sites that measure turbulent flux exchanges (water vapor, energy fluxes, and carbon dioxide) between ecosystems and the atmosphere (Baldocchi et al., 2001). FLUXNET comprises over 750 sites covering five continents. Measurements from the FLUXNET towers provide valuable information for evaluating satellite-based retrievals of surface fluxes. In this study, FLUXNET measurements from the FLUXNET 2015, the La Thuile Synthesis dataset, and the Large-scale Biosphere–Atmosphere (LBA) experiment in Brazil are used for evaluation (details are provided in Sect. 4.2).

FLUXNET 2015 tier 1 and tier 2 data were retrieved from http://fluxnet.fluxdata.org/data/fluxnet2015-dataset/. The data have been systematically quality controlled with a standard format throughout the dataset (http://fluxnet.fluxdata.org/data/fluxnet2015-dataset/data-processing/, Pastorello et al., 2014) and gap-filled using ERA meteorological forcing downscaling.

From the LBA experiment in Brazil, we use data from sites in Rondônia at the edge of a deforested region (BR-Ji1 and BR-Ji2) and near São Paulo (BR-Sp1). As the data did not span recent years, we instead use a climatology of the fluxes for comparison from 1999 to 2003. We note that, of course, the interannual variability in the region (such as El Niño and La Niña) could alter the seasonality and magnitude of the fluxes in the region.

We also use data from the La Thuile Synthesis Dataset (http://fluxnet.fluxdata.org/data/la-thuile-dataset/) covering 24 sites. These data are part of the free and fair use version of the dataset.

A total of 85 sites from the three datasets are selected for evaluation of WECANN retrievals spanning a large climatic and biome gradient (Fig. S2 and Table S1). For AmeriFlux towers, if measurements from both the FLUXNET 2015 dataset and the La Thuile dataset were available, we used the FLUXNET 2015 data. We have only selected sites that had at least 24 months of continuous measurements during 2007–2015. Any site that would have fallen outside of the WECANN land mask (Fig. S1) is excluded (several sites in coastal regions).

2.3.2 Basin-scale ET

We use estimates of an independent water budget closure model across five major basins to evaluate WECANN retrievals on regional scales (Aires, 2014; Munier et al., 2014) . ET estimates from the budget closure approach satisfy a water budget closure with no residual; therefore, they can be used as a reference to evaluate WECANN ET estimates on a basin scale. These basins include the Amazon ($4\,680\,000\,km^2$), Colorado ($618\,715\,km^2$), Congo ($3\,475\,000\,km^2$), Mississippi ($2\,964\,255\,km^2$), and Orinoco ($836\,000\,km^2$). Details of the water budget estimate are provided in Munier and Aires (2017), but in summary they combine estimates of precipitation, evaporation, water storage, and runoff to define a best estimate of the different fluxes and changes in storage, constrained by the water budget over the basin. Their analysis is carried out from 2002 through 2010.

3 Methodology

3.1 Artificial neural network setup

We developed an ANN retrieval algorithm to estimate the surface fluxes (LE and H) and GPP based on our six sets of input observations: SIF, net radiation, air temperature, soil moisture, precipitation, and SWE (as described in Sect. 2.2). The ANN used here is a feedforward network consisting of
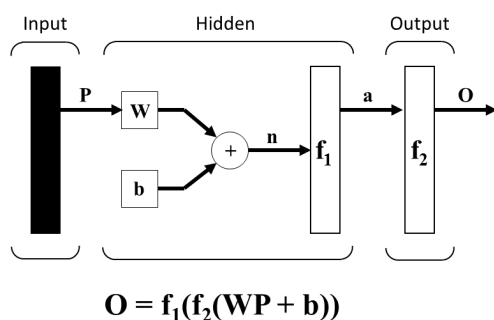
$$O = f_1(f_2(\mathbf{WP} + b))$$

Figure 1. Architecture of the ANN layers. Input layer provides the matrix $\mathbf{P}$ of the inputs to the hidden layer. The hidden layer has a matrix $\mathbf{W}$ of weights and b of biases for the neurons and the f_1 transfer function. The output of the hidden layer ($a = f_1(\mathbf{WP} + b)$) is an input to the output layer that applies the transfer function f_2 to the estimates and generates final outputs O.

three layers: (1) an input layer that directly connects to the input data, (2) one hidden layer, and (3) an output layer that produces the three output estimates. The number of neurons in the input and output layer is determined by the number of input and output variables, whereas for the hidden layer it has to be chosen according to the complexity of the problem (see below). The neuron output from each layer is fed to neurons in the subsequent layer through weighted connections. Each neuron output is the weighted sum of its inputs plus a bias, which is then subjected to a transfer function. In this study, we chose a tangent sigmoid transfer function for neurons in the hidden layer and a linear transfer function in the output layer. The change of the transfer function for the hidden layer (log sigmoid or tangent sigmoid) did not produce any significant changes in the retrievals (not shown); thus, we used the more common method. A schematic of the ANN architecture is provided in Fig. 1.

The training step of the ANN aims at estimating the weights for each of the neuron connections, such that the mismatch between the ANN outputs and target estimates is minimized. For this, we used the mean squared error (MSE) as the cost function and a backpropagation algorithm to adjust the ANN weights. During the training, the network implicitly learns the coupling of the LE, H, and GPP by using one set of neurons (with their respective weights and biases) to estimate the three variables. This is an advantage of using a machine learning technique that eliminates the need to define physical relationships between different variables.

For the purpose of training, the target data are divided into three subsets: training, validation, and testing constituting 60, 20, and 20 % of the target data, respectively. In each iteration, the training subset is used to estimates the weights in the network, and the convergence of the training towards the target data is checked using the validation subset. When overfitting of the network weights to the training data occurs, the validation estimates start diverging from the target data and the

training is stopped (early stopping). The weights from the last iteration before the occurrence of the divergence represent the final solution. The testing subset is used to assess the ANN performance after the training phase.

As an additional measure to avoid overfitting, we repeated the training for several ANNs with an increasing number of neurons in the hidden layer (1 to 15). For one to five neurons, the R^2 value between the target data and the ANN estimates increased with an increasing number of neurons. For more than five neurons, little change in the skill was observed when increasing the number of hidden layer neurons (Fig. S3). Thus, an ANN with five hidden layer neurons represents the simplest ANN that can converge to a solution and model the nonlinear relationship between the satellite inputs and the surface flux estimates.

To train the ANN, we used LE, H, and GPP estimates from the years 2008–2010. The target dataset was generated through a TC-based merging of triplets of the flux estimates introduced in Sect. 2.1 (details are discussed in Sect. 3.2). After completion of the training, the performance of the ANN and its ability to generalize was evaluated using the LE, H, and GPP target data from 2011. Finally, WECANN retrievals are evaluated against other global products and eddy covariance tower data. Results of these comparisons are presented in Sect. 4.

3.2 Target dataset: a Bayesian prior using triple collocation

One of the key issues in the design of an ANN to retrieve any geophysical variable is defining a good target dataset. One practice has been to use outputs from a land surface model as the target (Aires et al., 2005; Jiménez et al., 2013; Kolassa et al., 2013; Rodriìguez-Fernández et al., 2015). However, all observations and models contain random errors and biases. Therefore, the retrieval based on the ANN exhibits some of the biases of the original target dataset even if the ANN is able to make corrections to its original target data (e.g., correction of an imperfect seasonal cycle, as demonstrated by Jiménez et al., 2009). To address this issue, we use three datasets, which are sufficiently independent so that the training can learn from each dataset and benefit from all of them, synergistically. We implement a pseudo-Bayesian training by probabilistically weighting the occurrence of each training dataset by its likelihood and define a target dataset. The three datasets are listed in Table 1 for each variable.

To define this prior distribution, we use the TC technique. TC is a method to estimate the RMSE (and, if desired, correlation coefficients) of three spatially and temporally collocated measurements by assuming a linear error model between the measurements (McColl et al., 2014; Stoffelen, 1998). This methodology has been widely used in error estimation of land and ocean parameters, such as wind speed, sea surface temperature, soil moisture, evaporation, precipitation, fAPAR, and in the rescaling of measurement systems

to reference system for data assimilation purposes (Alemo-hammad et al., 2015; D'Odorico et al., 2014; Gruber et al., 2016; Hain et al., 2011; Lei et al., 2015; Miralles et al., 2010, 2011b; Parinussa et al., 2011), as well as in validating categorical variables such as the soil freeze–thaw state (McColl et al., 2016). The relationship between each measurement and the true value is assumed to follow a linear model:

$$X_i = \alpha_i + \beta_i t + \varepsilon_i \quad i = 1, 2, 3, \tag{1}$$

in which X_i is the measurement from the collocated system i (e.g., remote sensing observation, model output), t is the true value, and α_i and β_i are the intercept and slope of the linear model, respectively. ε_i is the random error in measurement i and TC estimates the variance of this random variable in each measurement. By further assuming that the errors from the three measurements are uncorrelated $\left(\text{Cov}\left(\varepsilon_i, \varepsilon_j\right) = 0, \text{ for } i \neq j\right)$ and the errors are uncorrelated with the truth ($\text{Cov}\left(\varepsilon_i, t\right) = 0$), the RMSE of each measurement error can be calculated as (McColl et al., 2014)

$$\begin{bmatrix} \sigma_{\varepsilon_1} \\ \sigma_{\varepsilon_2} \\ \sigma_{\varepsilon_3} \end{bmatrix} = \begin{bmatrix} \sqrt{Q_{11} - \dfrac{Q_{12}Q_{13}}{Q_{23}}} \\ \sqrt{Q_{22} - \dfrac{Q_{12}Q_{23}}{Q_{13}}} \\ \sqrt{Q_{33} - \dfrac{Q_{13}Q_{23}}{Q_{12}}} \end{bmatrix}, \tag{2}$$

in which Q_{ij} is the (ith or jth) element of the covariance matrix between the three measurements. Since the triplet of datasets used for training each of the fluxes (see Table 1) is derived through different semiempirical approaches with different sources of errors, the assumption of uncorrelated errors is more likely to be met. In the following, we will calculate the standard deviation of the random error component of Eq. (1) using TC for each of the surface fluxes and use them as TC-based errors of each product.

The TC-estimated errors for the surface fluxes and GPP are shown in Figs. S4–S6. The white regions represent missing retrievals or discarded negative estimates due to an insufficient data record. For LE, high TC errors are found in the Amazon rainforest and tropical Africa for GLEAM, in the Amazon rainforest and the Sahel for ECMWF, on the Indian peninsula for FLUXNET-MTE, and in US Great Plains for ECMWF and FLUXNET-MTE. For H, in addition to the aforementioned regions, high TC errors are also found in Southeast Asia for GLEAM and ECMWF and in northern Canada for FLUXNET-MTE. For GPP, MODIS and ECMWF have the highest errors in the Amazon rainforest, ECMWF and FLUXNET-MTE have relatively higher errors in US Great Plains, and all three products have similar errors in tropical Africa.

There are several likely causes for these errors. For the FLUXNET-MTE data, the regions that are not covered by (many) FLUXNET eddy covariance stations may result in larger uncertainties, and those regions for which interception

is a large component of the LE flux as well (Michel et al., 2016). For the GLEAM and ECMWF data, thick vegetation generally induces biases compared to the satellite observations, especially in tropical regions (Anber et al., 2015).

Finally, we use the TC-based RMSE estimates at each pixel to compute the a priori probability (P_i) of selecting a particular dataset in each pixel, if that pixel is used as part of the training dataset:

$$P_i = \frac{\frac{1}{\sigma_{\varepsilon_i}^2}}{\sum_{i=1}^{3} \frac{1}{\sigma_{\varepsilon_i}^2}}, \tag{3}$$

in which P_i is the probability of selecting dataset i when sampling from three measurements. We assume that these probabilities are time independent as we are limited by the currently available duration of the input data; however, future versions will explore the use of seasonally varying probabilities.

4   Results and discussion

4.1   Global magnitude of and variability in LE, H, and GPP

In this section, we present and compare the retrievals of LE, H, and GPP for the year 2011, which was not included in the training step of WECANN. Thus, it is used here to evaluate the ANN fit to the target values.

Figure 2 illustrates the annual global average and scatter plots of retrievals vs. target estimates. The spatial patterns of the WECANN retrievals are similar to expectations. The average global values in 2011 are $38.33 \, \text{W m}^{-2}$ for LE, $39.44 \, \text{W m}^{-2}$ for H, and $2.34 \, \text{gC m}^{-2} \, \text{day}^{-1}$ (or $123.16 \, \text{PgC yr}^{-1}$) for GPP. LE has the best R^2 (0.95) compared to H ($R^2 = 0.89$) and GPP ($R^2 = 0.90$). The root mean squared difference (RMSD) of each of the retrievals with respect to the target estimates is as following: for LE, RMSD $= 11.06 \, \text{W m}^{-2}$; for H, RMSD $= 13.13 \, \text{W m}^{-2}$; and for GPP, RMSD $= 1.22 \, \text{gC m}^{-2} \, \text{day}^{-1}$.

The seasonal variability in and spatial pattern of the retrievals from 2011 are shown in Figs. 3–5. LE does not exhibit any variability over deserts such as the Sahara and Arabian Peninsula, as expected (Fig. 3). Wet tropical forests exhibit subtle seasonal variability in LE. These spatial variabilities in the seasonal cycle reflect changes in the radiation, temperature, water availability during the dry season, soil nutrients, soil type conditions, and leaf flushing (Anber et al., 2015; Morton et al., 2014, 2016; Restrepo-Coupe et al., 2013; da Rocha et al., 2009; Saleska et al., 2016). In contrast, seasonal variability dominated by radiation availability is noticeable in wet midlatitude regions for both the Northern Hemisphere (NH) and Southern Hemisphere (SH), i.e., East Asia, the eastern US, and the north and east Australian coasts with over $60 \, \text{W m}^{-2}$ difference between winter and
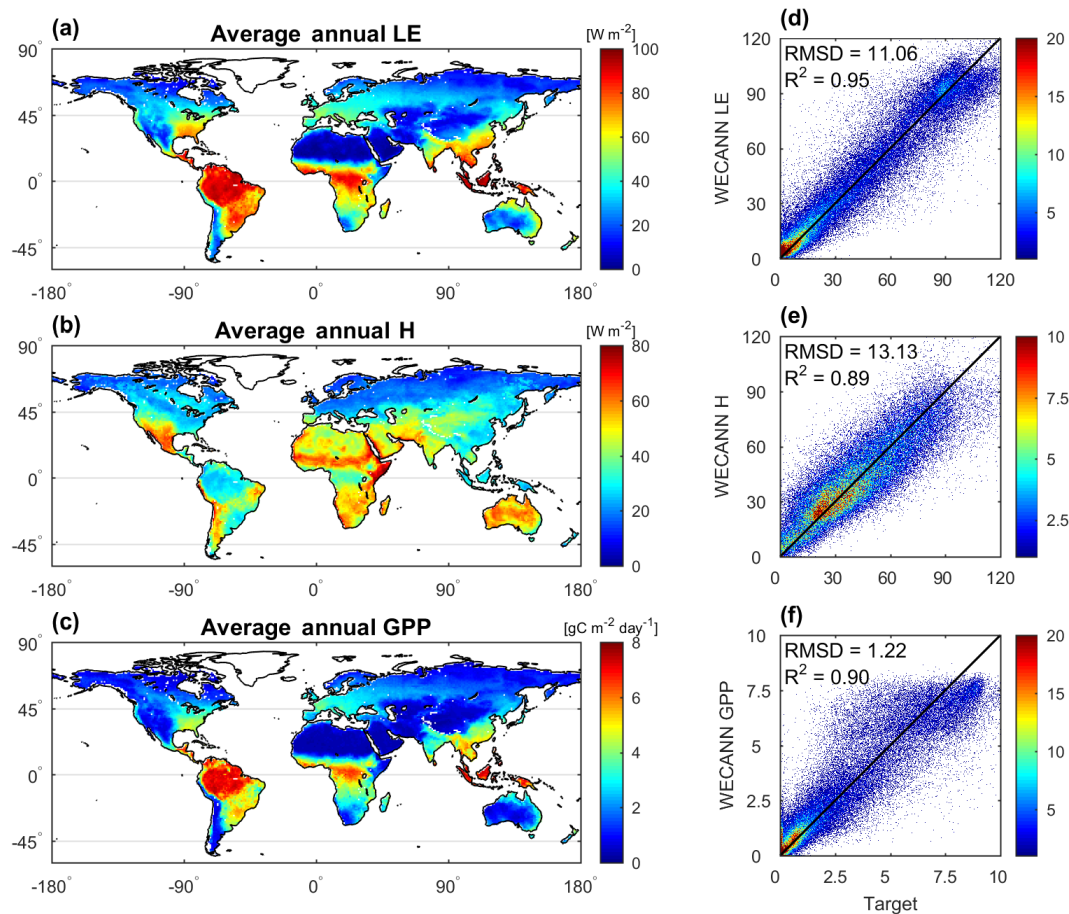
Figure 2. Left column: annual average retrievals in 2011 for **(a)** LE, **(b)** H, and **(c)** GPP. Right column: density scatter plot between estimates of ANN and target data for **(d)** LE, **(e)** H, and **(f)** GPP during the validation period (2011). The density of scatter points is represented by the shading color. The diagonal black line depicts the 1 : 1 relationship.

summer months. One exceptional case is South Asia, where LE does not significantly rise in spring, likely due to the effects of the monsoonal climate. In eastern South America, the ET estimates are relatively high compared to GPP estimates. This difference can be caused by either low water use efficiency or significant rain reevaporation and soil evaporation. Moreover, the SIF relationship with GPP likely changes in C4 plants. However, we did not impose the C4–C3 delimitation in the ANN as it would be highly dependent on the quality of the classification map used. We note that all training products used here include C3–C4 delimitation and therefore the C3–C4 delimitation is implicit in the training dataset; therefore, it can be learned by the network.

Seasonal variabilities in H (Fig. 4) are distributed in an opposite pattern to LE, as expected. Deserts and dry regions, i.e., the Sahara, southwestern US, and Western Australia demonstrate much more seasonal variability than the rest of the world. Given the strong water limitations there, the available energy converted into H becomes dictated by the seasonal cycle of solar radiation. In contrast, tropical rainforests

(Amazon, Congo, Indonesian) exhibit limited seasonal variability. In midlatitude energy-limited regions (central and eastern Europe, the eastern US), H also reflects the course of available energy, and in more water-limited regimes (e.g., the western US and Mediterranean Europe), it reflects the interplay between soil dryness and available energy, with a peak between spring and summer for dry regions.

The seasonal variability in GPP (Fig. 5) in northern latitudes follows the availability of radiation in wet regions, with a peak in summer and another in spring for dry regions, corresponding to both soil water availability and high incoming radiation. A clear east–west transition conditioned by water availability is observed in the continental US. In the tropics and subtropics, the response is diverse. The Amazon rainforest exhibits high GPP throughout the year with a peak between September and February in the wetter part of the basin, following the dry season, consistent with the observations at eddy covariance towers near Manaus and Santarém (Restrepo-Coupe et al., 2013; da Rocha et al., 2009). Compared to LE, substantial geographical variability is observed
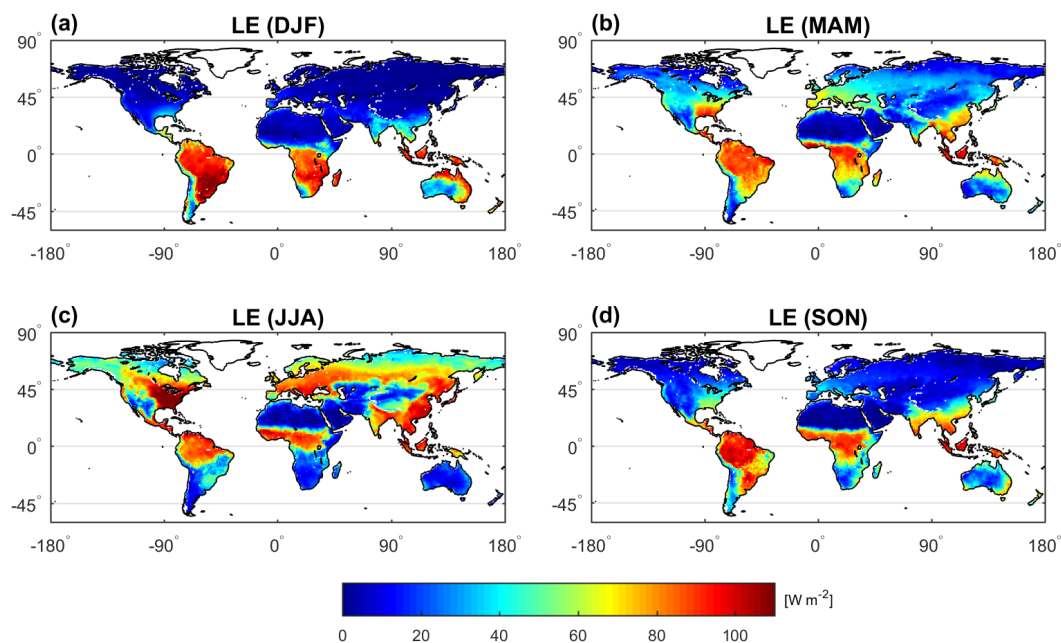
Figure 3. Global patterns of seasonal average LE from WECANN in 2011: **(a)** December–February, **(b)** March–May, **(c)** June–August, and **(d)** September–November.
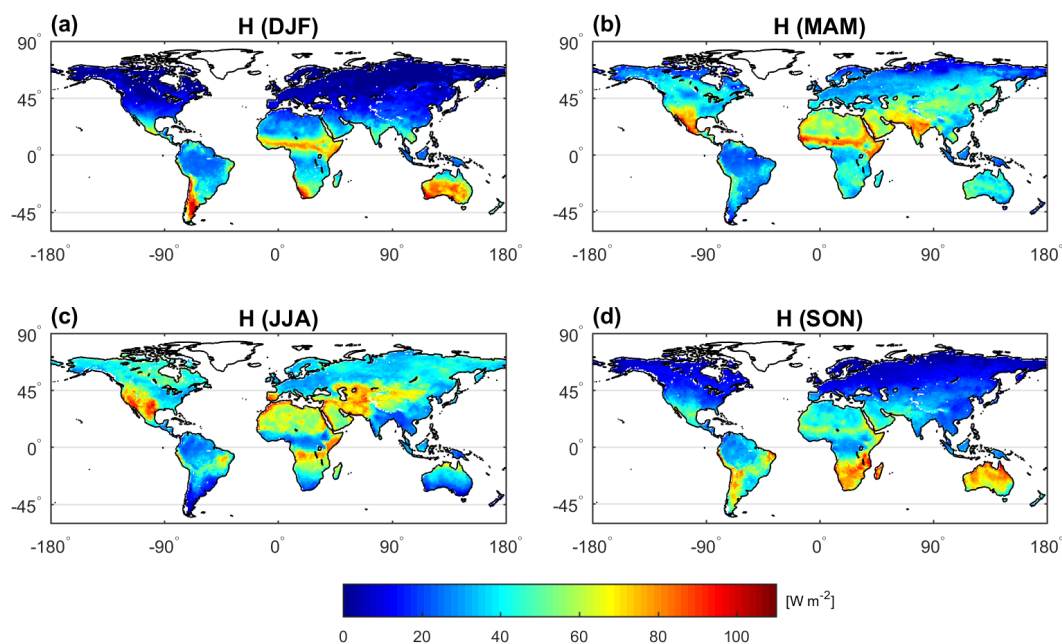

Figure 4. Similar to Fig. 3 but for H instead of LE.

in the Amazon because of the strong variabilities in soil type, green-up, biodiversity, and rooting depth. In the drier part of the basin, water availability controls the seasonal cycle of photosynthesis, and the peak in GPP is observed in the wet season (DJFMA). In the Congo rainforest, GPP exhibits four seasons, with two wet and two dry ones, with a substantial decrease in GPP during those dry spells. In Indonesia,

GPP is steadier throughout the year, exhibiting high values year round. Monsoonal climates over India, Southeast Asia, northern Australia, and Central–North America are well captured with rapid rise in GPP following water availability. The highest GPP values are observed in rainforests and the agricultural US Great Plains, in JJA for the latter. Northern latitude regions mainly exhibit substantial GPP in the summer
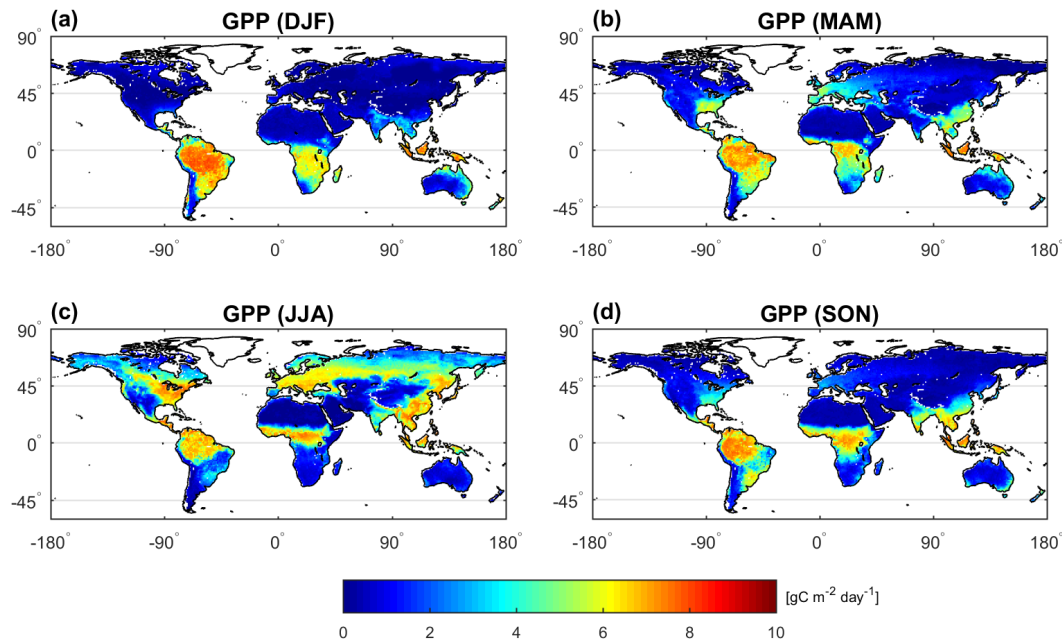
Figure 5. Similar to Fig. 3 but for GPP instead of LE.

and late spring and small values throughout the rest of the year.

4.2 Evaluation with FLUXNET data

Direct validation of the WECANN retrievals is challenged by the fact that no global, error-free estimates of LE, H, and GPP are available. Remote sensing or model products such as those used for training have their own errors. When three datasets with uncorrelated errors (commonly assumed to be true if the sources of error in each dataset have no common physical origin) are available, TC provides a valuable technique to evaluate large-scale datasets in the absence of a known truth. However, WECANN's use of different training datasets will cause the presence of some correlated errors between WECANN retrievals and any of the datasets used for the training. Instead, we evaluate the retrievals by comparing them to data from a set of FLUXNET eddy covariance towers. WECANN uses three training datasets, one of which (FLUXNET-MTE) is based on upscaling FLUXNET eddy covariance tower estimates. This might cause some dependence between WECANN retrievals and the tower estimates. However, WECANN learns from all three training datasets collectively and uses remote sensing observations as input. Therefore, this dependence is negligible. In situ estimates from eddy covariance towers with a footprint of a few hundred meters to kilometers may not be representative of the entire $1° \times 1°$ pixel and are known to have problems with energy closure (Foken et al., 2010). However, in the comparison against tower data the impact of large-scale climate variability and seasonality can still be seen even on different

spatial scales. For instance, the phenology has a strong impact on the seasonal cycle of the LE, H, and GPP and in the following examples; it is clearly highlighted when comparing different products to flux tower estimates.

A summary of statistics across 85 FLUXNET sites is provided in Tables S2–S4. Overall, WECANN performs better than other alternative global products. In particular, WECANN has the highest correlation for 76 % of sites for LE, 58 % of sites for H, and 55 % of sites for GPP. This high R^2 reflects the capacity of WECANN to correctly capture the seasonal cycle and interannual variability, as it is largely imposed by the remote sensing observations rather than by the statistical retrieval (Jiménez et al., 2009). One of the reasons for this is the presence of the SIF information in the ANN retrieval, which is directly related to GPP and plant transpiration (Frankenberg et al., 2011). The RMSE of WECANN is lower than all other products at 71 % of sites for LE, 46 % of sites for H, and 51 % of the sites for GPP. The bias is also reduced compared to other retrievals, even if some variability can be seen from site to site.

Figure 6 shows a summary of the correlation coefficients presented in Tables S2–S4 for each group of plant functional types (PFTs). Each class has between 6 and 22 sites. WECANN has the best mean within each PFT class, and the smallest variability in most of the classes for all three variables.

Figure 7 shows the comparison of monthly WECANN retrievals and three other global products' estimates with the tower estimates across five select sites that span a range of climatic and vegetation coverage conditions. At the Oklahoma agricultural site (US-ARM), H and LE are reproduced
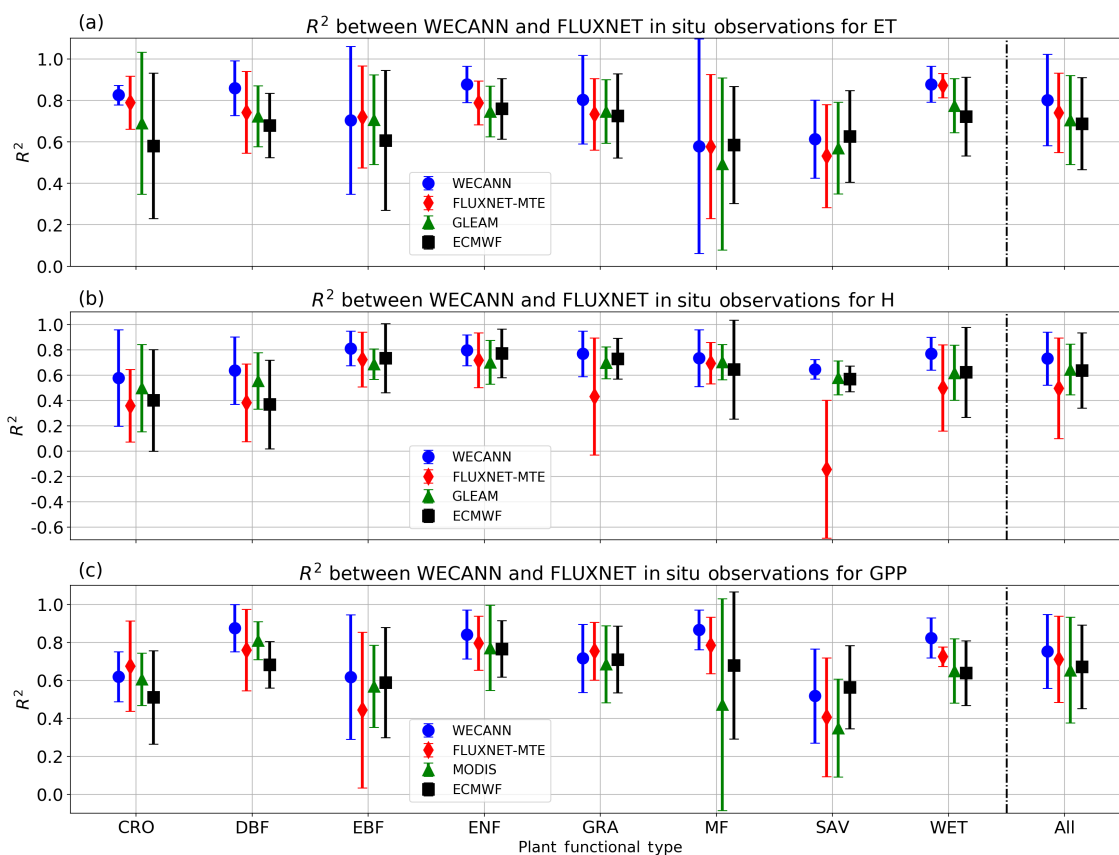
Figure 6. Correlation coefficient (R^2) between WECANN retrievals and FLUXNET tower estimates categorized across different plant functional types for **(a)** LE, **(b)** H, and **(c)** GPP. Markers show mean, and whiskers show 1-standard-deviation intervals. (CRO: croplands, DBF: deciduous broadleaf forests, EBF: evergreen broadleaf forests, ENF: evergreen needleleaf forests, GRA: grasslands, MF: mixed forests, SAV: savannas, and WET: permanent wetlands).

well, yet dry year H is underestimated (Fig. 7a). The GPP reported at the site very rapidly decays at the end of the spring, whereas the region is highly agricultural with sustained agriculture in the summer. The difference between the reported GPP and WECANN retrievals might again be due to the difference in the footprint of the two estimates.

At the Brasschaat, Belgium, site (BE-Bra) (Fig. 7b), LE is very well captured by WECANN, which captures the seasonal cycle well, yet misses some of the interannual variability. WECANN outperforms the other retrievals of LE and GPP and captures the GPP seasonal cycle very well compared to other products, which display a too-early GPP rise and overestimate summer GPP. Again, the SIF data provide independent useful data compared to other environmental information (radiation, temperature, vegetation indices) used by the other retrieval schemes. All retrievals strongly underestimate the reported eddy covariance H. At this humid site though, the magnitude of the measured H is often higher or on the same order in the summer as LE. Given the high degree of urbanization around the site, it is most likely a reflection of the footprint of the eddy covariance and the fact that it

observes urbanized surfaces with high H. Indeed, the surface energy budget is not locally balanced and turbulent fluxes are higher than the observed net radiation minus ground heat flux.

At the cold Finland site (FI-Hyy), WECANN captures the seasonal cycle of GPP and LE very well, as well as H to a lesser extent. WECANN reproduces the seasonality, amplitude, and interannual variability better compared to other retrievals (Fig. 7c). It also reflects the difficulties of retrieving fluxes in snow-dominated regions. SIF has the great advantage that it is not directly sensitive to snow compared to vegetation indices, for instance, which incorrectly attribute snowmelt and changes in observed ground color to photosynthesis onset (Jeong et al., 2017).

At the monsoonal grassland site of Santa Rita, AZ, WECANN correctly captures the complex dynamics of H and LE at the site, with some rain periods preceding the monsoon period (Fig. 7d). However, WECANN slightly underestimates LE and overestimates GPP. In fact, all products overestimate GPP in the dry and cold seasons. The landscape in the region is highly heterogeneous, with denser vegetation in
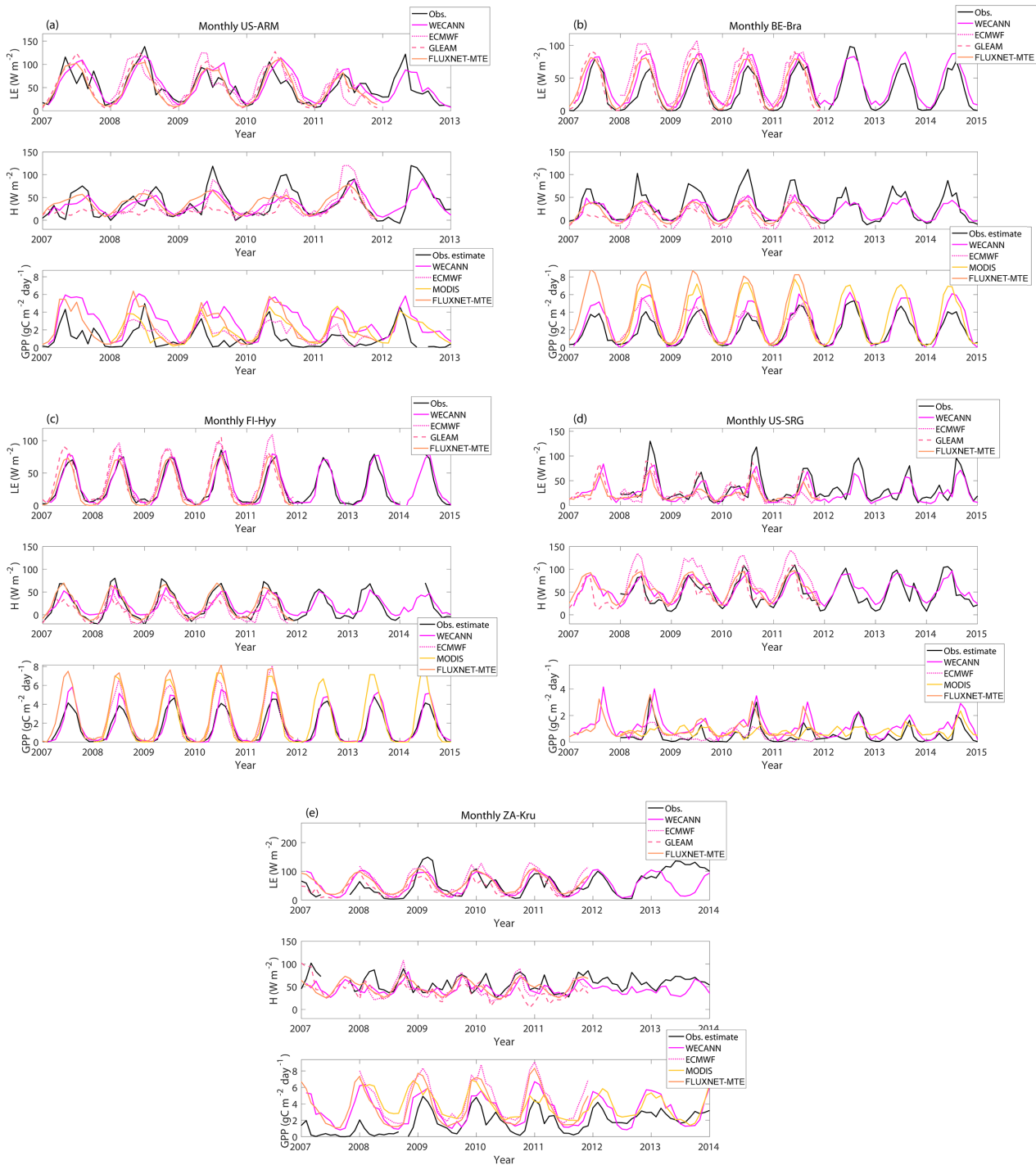
Figure 7. Comparison of the retrievals with eddy covariance observations of LE, H, and GPP across five sites: **(a)** US-ARM site, USA; **(b)** AT-Neu site, Austria; **(c)** BE-Bra site, Belgium; **(d)** FI-Hyy site, Finland; **(e)** US-SRG, USA; and **(f)** ZA-Kru, South Africa.

riparian zones, away from the tower location, which may explain the lower GPP value at the site compared to estimates of the larger-scale values.

Finally, at the South African Mediterranean site, ZA-Kru, WECANN reproduces some of the dynamics of the observed

H, yet is typically smoother (Fig. 7e). It reasonably captures the LE dynamics, except for the suspect cold season increase reported at the tower in 2013 (like other products). All products overestimate the reported GPP, though WECANN

is closest to the observations and captures the seasonal dynamics better compared to other products.

Overall, across the different sites, the WECANN retrieval performs better than other products, especially in terms of the seasonality of LE, H, and GPP. Several factors contribute to the improved retrieval of WECANN compared to other products, even at those smaller footprint sites. First, the SIF measurements that are directly correlated with GPP provide a better constraint on estimating LE, H, and GPP. The ANN approach in WECANN also uses a novel training technique based on probabilistically merging different datasets to remove outliers from its target dataset. Therefore, WECANN retrievals learn collectively from the different datasets (and remote sensing observations) and are closer to the truth than each of the individual target datasets.

4.3 Comparison against other products based on remote sensing

In this section, we compare the WECANN-based estimates to other datasets used in the training to better understand how WECANN differs from those training data. Figure 8 shows the comparisons for LE and indicates that our product has a relatively similar R^2 with the three products ($R^2 = 0.96$ with FLUXNET-MTE and ECMWF and $R^2 = 0.94$ with GLEAM). However, the scatter plot with FLUXNET-MTE is more concentrated and aligned along the 1 : 1 line, further emphasizing the consistency between the two datasets (RMSD of 6.42 W m^{-2} for FLUXNET MTE versus 8.47 and 9.72 W m^{-2} for GLEAM and ECMWF, respectively). Differences in spatial patterns shown in Fig. 8a–c reflect that WECANN exhibits smaller spatial differences with FLUXNET-MTE than GLEAM or ECMWF and such differences exhibit a narrower range between -10 and 10 W m^{-2}. FLUXNET-MTE overestimates LE compared to WECANN in transitional tropical and subtropical regions and particularly over India, which are regions with few eddy covariance towers. GLEAM exhibits substantial differences with our product, particularly in regions dominated by seasonal water stress such as Brazilian savannas, the Horn of Africa, Central America, India, and the subtropical humid part of Africa south of the Congo. In the Sahel, GLEAM LE is higher than our estimate and FLUXNET-MTE. The LE estimate of ECMWF is nearly always higher than our estimate, with much higher values in the Congo, the Amazon, southern Brazil, and northern Canada. In Europe, where the ECMWF estimate should be best because of the frequent weather operational forecast checks and model adjustment in the region, the estimates are more similar. The differences and similarities of WECANN retrievals with the three target datasets are consistent with the error estimates from TC. For example, Fig. S4 shows that FLUXNET-MTE has the smallest error in LE estimates globally compared to GLEAM and ECMWF, other than across India. WECANN retrievals also have better agreement with FLUXNET-MTE.

The differences in H estimates are more complex (Fig. 9). First, the R^2 values between WECANN and the other datasets are slightly lower than for LE. ECMWF and FLUXNET-MTE yield a higher R^2 with WECANN (0.92) while GLEAM has an R^2 of 0.87. GLEAM exhibits lower H in most of the NH, especially in seasonally dry regions, potentially due to its simple formulation of ground heat flux (G). H estimates are relatively higher over the Amazon and Congo but lower over Indonesia for GLEAM. In the southern Sahara and northern Sahel as well as in eastern Asia and Canada, GLEAM has lower H compared to WECANN and FLUXNET-MTE. ECMWF exhibits higher values in seasonal dry regions such as the western US, Brazilian savannas, southern Congo, and the Sahel compared to WECANN and smaller values in the Amazon, Indonesia, over desert areas of the Sahara and Arabian Peninsula, and Southeast Asia. The GLEAM and ECMWF H difference maps show many similar patterns: the Sahara, eastern Europe, and East Asia are underestimated, while southern Africa and the eastern part of the Amazon are overestimated. Similarly the errors patterns estimated from TC (Fig. S5) are consistent with the comparison of WECANN and the target datasets. Figure S5 shows that ECMWF has higher errors in the Sahel, southern Congo, and Brazilian savanna, and GLEAM has higher errors in the Amazon, East Asia, and Central Africa.

The comparison between the GPP estimates shows significant differences (Fig. 10). WECANN compares the best against FLUXNET-MTE ($R^2 = 0.93$), with MODIS ($R^2 = 0.91$) and ECMWF ($R^2 = 0.90$) following. While all three products have a similar R^2, their spatial differences are distinct. In the Amazon, ECMWF and FLUXNET-MTE have larger GPP estimates compared to WECANN, while MODIS estimates are much smaller. In cold northern latitude regions of Siberia and northern Canada, all three products have a higher GPP than WECANN. In Congo, MODIS and FLUXNET-MTE have higher GPP, while ECMWF has a lower one. In central and the southwestern US, all three products tend to yield lower GPP. Comparison of these findings with the error estimates from TC (Fig. S6) shows that FLUXNET-MTE has the lowest errors globally, while ECMWF has the largest errors in the Amazon.

Finally, we compare annual anomalies of WECANN retrievals and the three training datasets globally and in different climatic zones. The zones are defined as polar (90–60° N), NH midlatitude (60–10° N), tropics (10–15° S), and SH midlatitude (15–60° S). Results are presented in Fig. S7. WECANN anomalies are derived from the mean values between 2007 and 2015. However, not all the training products are available for this period. Therefore, their anomalies are calculated from their respective temporal domain, which is 2007–2011 for FLUXNET-MTE, GLEAM, and MODIS and 2008–2011 for ECMWF. Anomalies from the four products have similar patterns in general, while their absolute values differ. Anomalies in GPP have better agreement across different products compared to LE and H. Evaluation of the
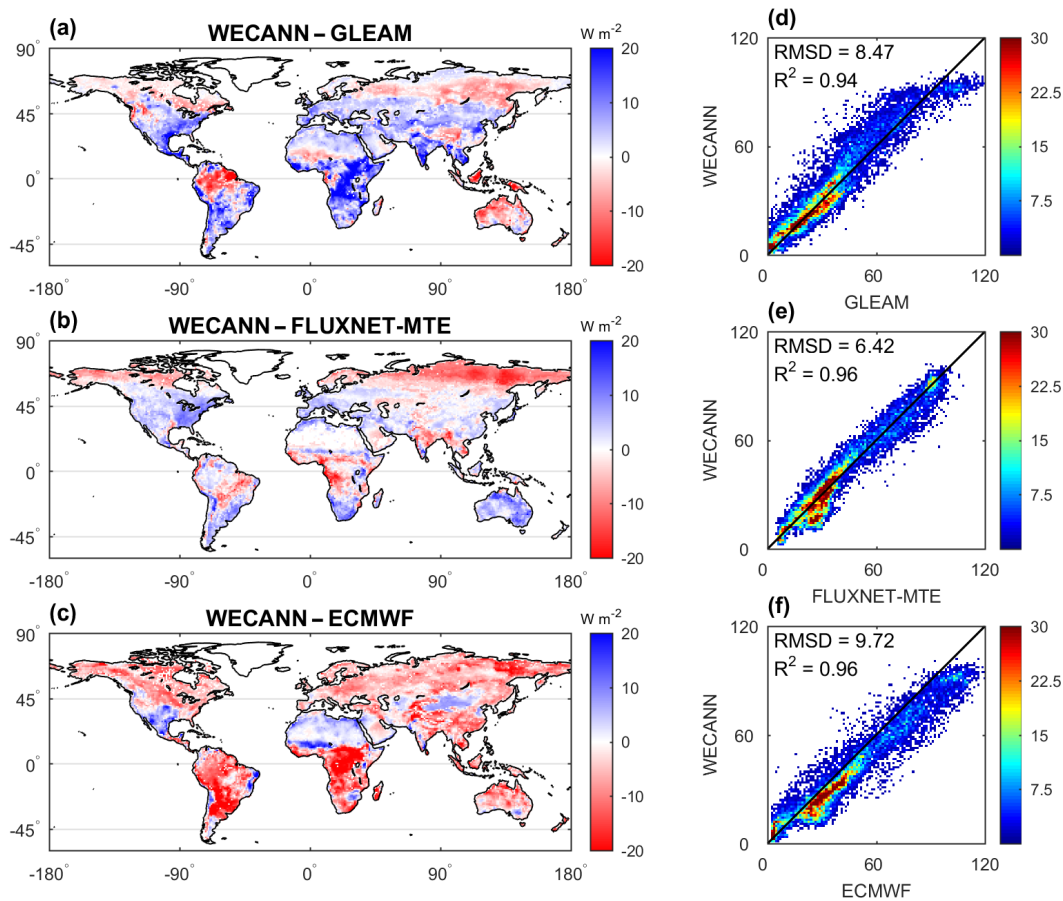
Figure 8. Difference between annual mean LE retrieved by WECANN and the three target datasets **(a–c)**. Scatter plots of LE retrieved from WECANN vs. from each of the target datasets **(d–f)**. Data used are from 2011.

discrepancies between these anomalies is beyond the scope of this paper and will be characterized in detail in a future study.

4.4 Extreme event assessment

In order to further assess WECANN on regional scales, we analyze its capacity to capture extreme events. We thus selected three major heat wave and drought events that occurred during the temporal coverage of the WECANN product. These events are the 2010 Russia heat wave, 2011 Texas drought, and 2012 US Corn Belt drought. Figure 11 shows the percentage of average monthly anomalies with respect to mean values, for LE, H, and GPP, in each of the three cases. The patterns reveal significant anomalies in all fluxes, which is consistent with reported patterns. In summer 2010, a historical heat wave occurred over western Russia and resulted in an all-time maximum temperature record in many locations (Dole et al., 2011). The extent of reduction in LE and increase in H derived from WECANN retrievals is consistent with estimates reported in the literature (Lau and Kim, 2012), with a 10–15 % increase in H and a 15–20 % reduc-

tion in LE. In early 2011, drought conditions developed in the southern US, particularly in the states of Texas and Louisiana (Luo and Zhang, 2012). By April, most of Texas, Oklahoma, Louisiana, and Arkansas was classified in the D4 drought condition (exceptional drought), and the situation continued throughout the summer and fall of 2011 as reported by the US Drought Monitor (Svoboda et al., 2002). As Fig. 11 reveals, the same spatial pattern is pronounced in the monthly anomalies derived from WECANN retrievals, emphasizing massive reduction in LE and GPP accompanied by high H over the region.

Finally, an intense drought in the central US, particularly in the Corn Belt, occurred in 2012 and reduced maize yields by about 25 % and increased prices by 17–24 % (Boyer et al., 2013; USDA, 2013). By mid-September 2012 almost two-thirds of the continental US was covered by drought, and different parts of the US Corn Belt were categorized as either D3 (extreme drought) or D4 (exceptional drought) conditions. Figure 11 shows similar patterns in LE, H, and GPP, with a significant positive anomaly in H (~ 20 %) and re-
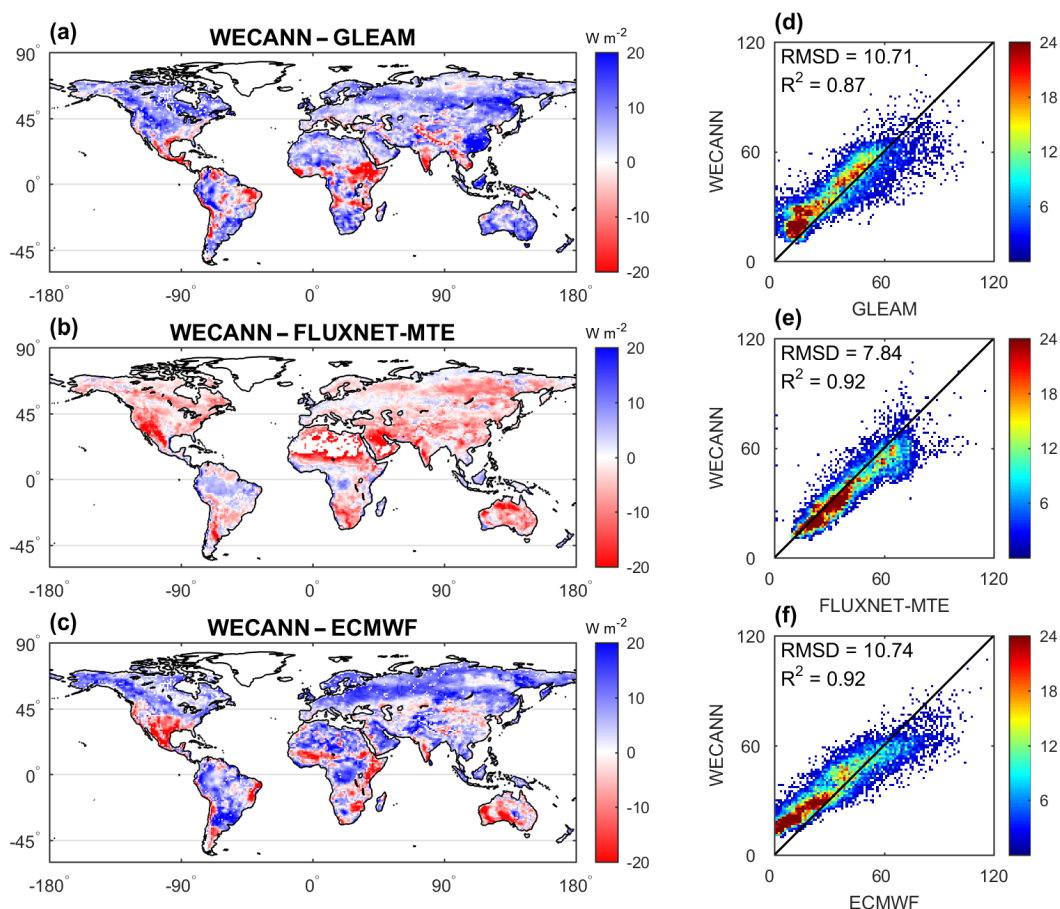
Figure 9. Similar to Fig. 8 but for H instead of LE.

ductions in LE and GPP ($\sim$ 20 %), consistent with crop yield decrease.

4.5 Basin-scale evaluation

We also assess the accuracy of WECANN ET retrievals using ET estimates from the independent water budget closure model introduced in Sect. 2.3.2. The analysis is carried out for the years 2007 to 2010, which overlap between WECANN retrievals and the water budget closure study. Figure 12 shows the relative absolute difference in ET estimates from WECANN compared to the ET estimates from the water budget study for each of the five basins (mean value and 1 standard deviation across 48 months). Mean absolute differences vary between a low of 5 % in the Amazon and a larger 24 % value in Colorado, while the other three basins have mean differences of 9, 17, and 20 %. While the differences vary between a low and moderate range, it should be noted that the coarse spatial resolution of the WECANN product causes a difference in the spatial averaging to get the basin-level estimates of ET. Moreover, in the budget closure estimates only a single runoff (at the outlet of the considered basin) is used over the entire basin; therefore, large hetero-

geneous basins such as the Colorado and Mississippi have large uncertainties associated with them, as runoff does not correctly constrain the flux distribution over the entire basin. It is over those basins that the WECANN retrieval compares less favorably with these large-scale estimates. Downscaled version of those estimates would further help in the evaluation of ET products.

4.6 Uncertainty analysis of WECANN retrievals

One of the advantages of a statistical retrieval algorithm, in particular of ANNs, is that the run time is extremely fast after the training step. This enables us to characterize the uncertainty of the retrievals by propagating the uncertainties in the input variables through the network. For this purpose, we set up a 10 000-bootstrap experiment and run the WECANN retrieval by adding error to input variables. The errors are normally distributed with a mean of zero and a standard deviation that depends on the input variable. For SIF, air temperature, and soil moisture, we use the error estimates or standard deviations reported in their associated products. These errors vary spatially and temporally and we used the associated value for each time and space data point. For net radi-
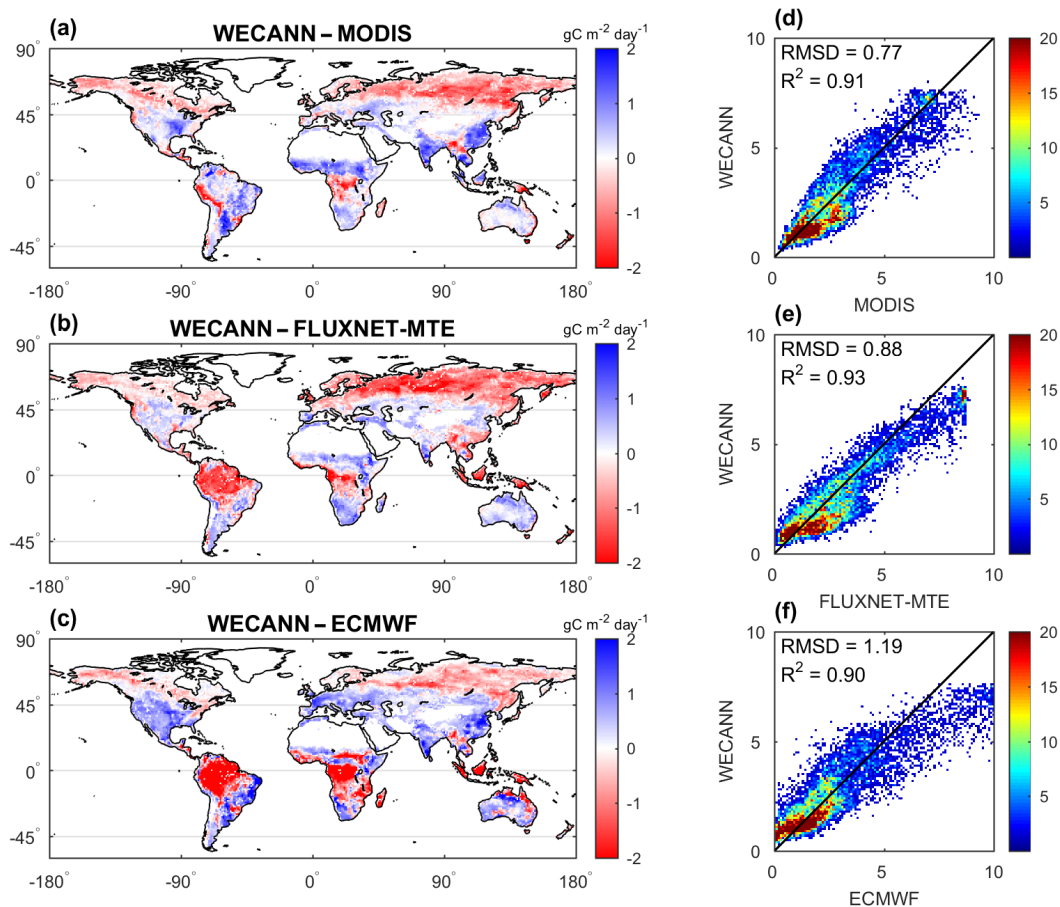
Figure 10. Similar to Fig. 8 but for GPP instead of LE.

ation, we use a constant standard deviation of $34.58\,\mathrm{W\,m^{-2}}$ based on the analysis by Pan et al. (2015). For precipitation and SWE estimates, we use a conservative 10 % of the estimates themselves as a standard deviation for error. For each bootstrap replicate, we sample from the error distribution of each input variable and add that to the input.

Figure 13 shows the results of the bootstrap for each of the LE, H, and GPP values globally and in different climatic zones (defined in Sect. 4.3). Each panel in Fig. 13 shows the uncertainty derived from the bootstrap experiment, relative to the interannual variability in the retrievals. GPP estimates are provided in units of petagrams of carbon per year as total productivity in each region. LE and H are provided in units of watts per square meter as an average rate of flux in each region.

On a global scale the GPP ranges between a minimum of $117.15 \pm 2.379\,\mathrm{PgC\,yr^{-1}}$ in 2015 and a maximum of $124.82 \pm 2.482\,\mathrm{PgC\,yr^{-1}}$ in 2007. Similarly, LE has a minimum of $37.40 \pm 0.54\,\mathrm{W\,m^{-2}}$ in 2015 and a maximum of $38.33 \pm 0.53\,\mathrm{W\,m^{-2}}$ in 2011. H has a maximum of $41.00 \pm 0.54\,\mathrm{W\,m^{-2}}$ in 2015 and a minimum of $39.43 \pm 0.52\,\mathrm{W\,m^{-2}}$ in 2011.

The interannual variations in surface fluxes and GPP show distinct patterns. For example, in the year 2015, which was an El Niño year, LE and GPP decreased notably, and H increased to an extreme value in the 9 years of the WECANN product. Moreover, from 2011 to 2015 both LE and GPP have a consistent decreasing trend on a global scale. The interannual variability in GPP and LE is similar on a global scale, while their regional patterns are different. For example, in the year 2015, GPP on a global scale and in all regions has decreased with respect to 2014, while LE in polar and NH mid-latitudes has increased, and LE on a global scale has decreased. As expected, the variability in LE and H is anticorrelated. We note that while WECANN is trained on three independent estimates of LE, H, and GPP, its interannual variability is driven by remote sensing observations that are input into the ANN.

4.7 Impact of SIF on the retrieval of surface fluxes and GPP

Satellite SIF observations are relatively new and have not been previously used to estimate LE and H on a global scale. Therefore, we want to assess the information content of SIF
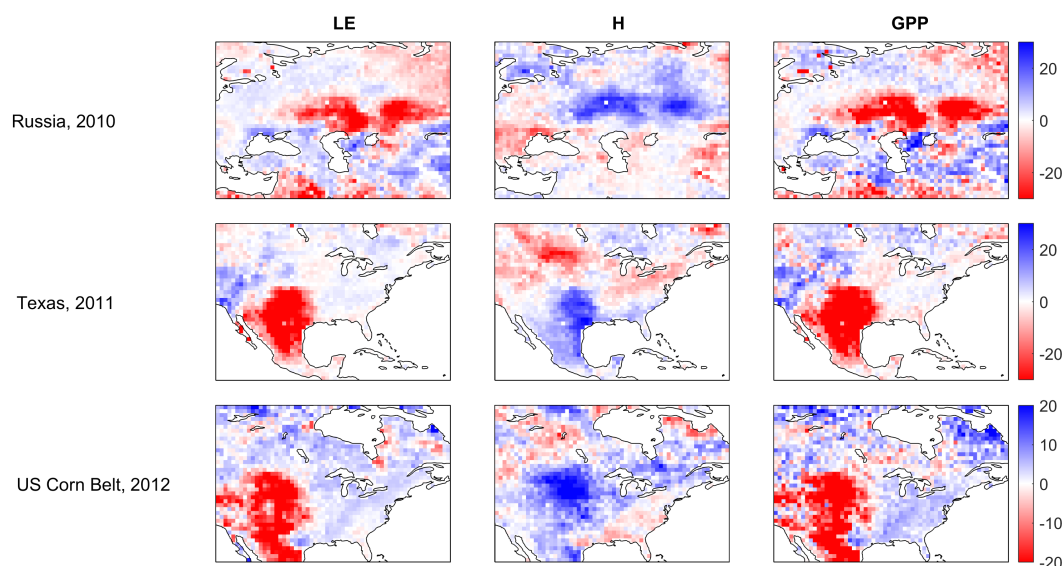
Figure 11. Mean monthly anomalies (in percentage with respect to mean value) for three extreme heat wave events.
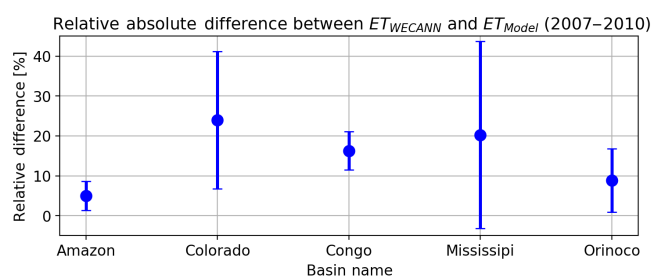

Figure 12. Relative absolute difference between ET estimates of WECANN compared to modeled ET from basin-scale water budget closure. Markers show the mean, and whiskers show 1-standard-deviation intervals.

observations in the WECANN retrievals by replacing them with more typical optical and/or near-infrared indices of vegetation (normalized difference vegetation index, NDVI, or enhanced vegetation index, EVI).

To do so, we trained two different ANNs with NDVI and EVI instead of SIF data on each of the three variables (LE, H, and GPP) and evaluated the retrievals against the same FLUXNET tower measurements used in Sect. 4.2 for evaluating WECANN retrievals. Tables S5–S7 show the results of evaluations of these three retrievals against the tower measurements for LE, H, and GPP. In terms of the correlation coefficient, on average all three retrievals have a relatively similar performance except in regions such as Spain where phenology (and incident radiation) is not the main contributor to the flux variability (ES-LgS). Indeed, in such regions, changes in canopy structure are more limited and changes in response to water stress (through changes in light- and water-use efficiency) are the primary reason for the seasonal variability. This emphasizes, similar to current thinking on the

SIF signal, that the monthly SIF signal is dominated by incident radiation and canopy structure but that in some conditions light-use efficiency changes are detected by SIF but not optical vegetation indices (Lee et al., 2013). We also point out that current SIF retrievals (such as those from GOME-2 used here) are still noisy as they were not obtained by satellites designed to measure SIF. Future SIF-designated missions such as Fluorescence Explorer (FLEX) will have higher accuracy and finer spatial and temporal resolution (Drusch et al., 2016). We expect they will further enhance the retrievals of surface fluxes and GPP such as those from WECANN.

5   Conclusions

This study introduces a new statistical approach to retrieving global surface latent and sensible heat fluxes as well as gross primary productivity using remotely sensed observations on a monthly timescale. The methodology is developed based on an artificial neural network that uses six input datasets including solar-induced fluorescence, precipitation, net radiation, soil moisture, snow water equivalent, and air temperature. Moreover, a Bayesian approach is implemented to optimally integrate information from three target datasets for training the ANN, using triple collocation to calculate a priori probabilities for each of the three target datasets based on their uncertainty estimates.

The new global product, referred to as WECANN, is evaluated using target datasets as well as FLUXNET tower observations. The evaluation results compared with training datasets show that our retrieval has similar correlation with the three products, while it has the smallest RMSD with FLUXNET-MTE for LE (RMSD = 6.42 W m^{-2}), H (RMSD = 7.84 W m^{-2}), and
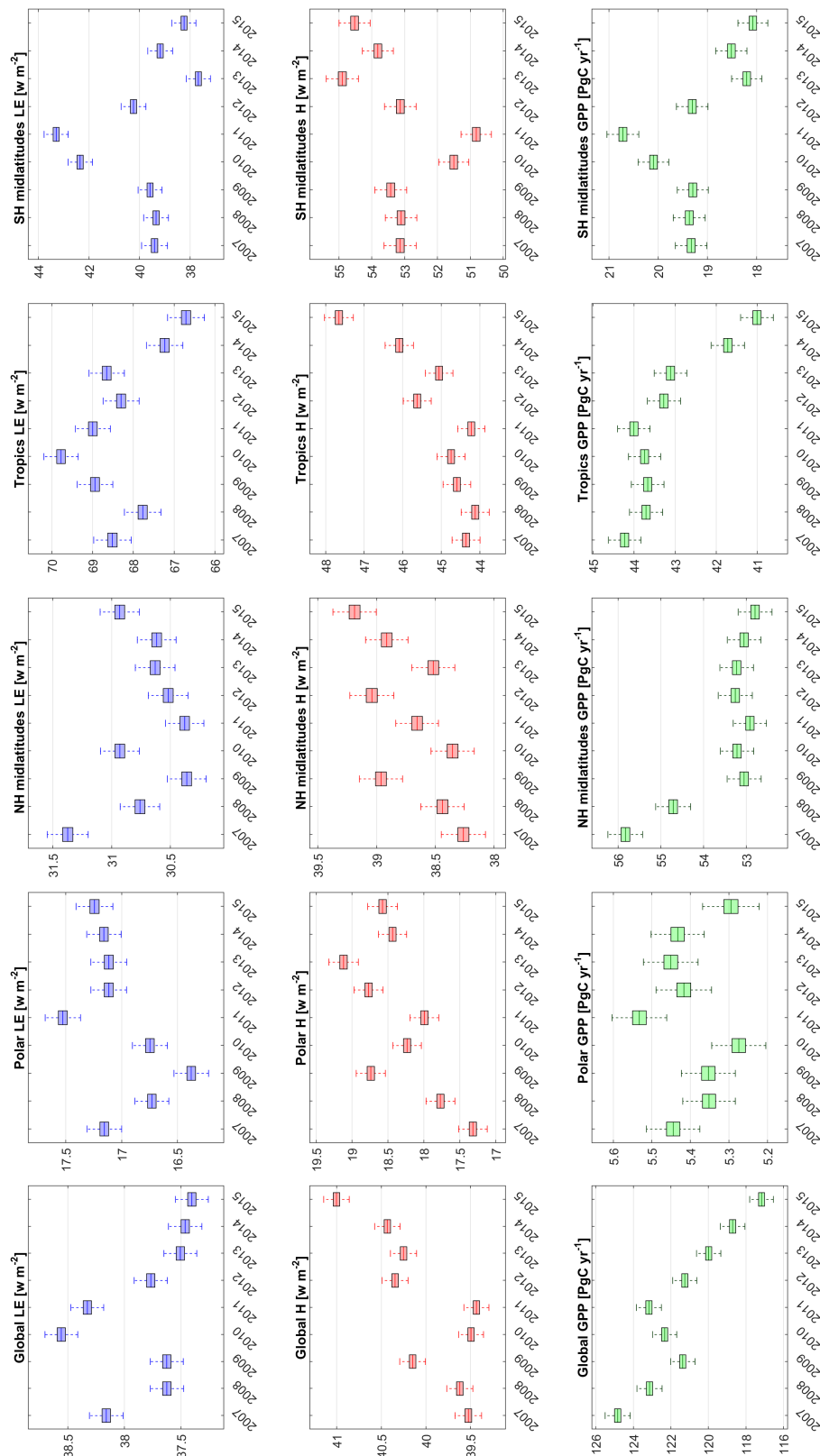
Figure 13. Annual mean estimates and uncertainty bounds of LE (top row), *H* (middle row), and GPP (bottom row) retrievals on global (left column) and regional (four right columns) scales between 2007 and 2015. The central line in each box indicates the mean, the edges of the box are the 25th and 75th percentiles, and the whiskers show the most extreme values.

GPP (RMSD $= 0.88 \, \text{gC m}^{-2} \, \text{day}^{-1}$), which is believed to be one of the most realistic global datasets. It also has the lowest RMSE based on our TC error estimates (Fig. S4–S6), despite its reported underestimated interannual variability due to the use of climatological values for several meteorological drivers (Miralles et al., 2014a, 2016). Such a tendency can also be summarized from the global difference maps, which show that FLUXNET-MTE has the best agreement with WECANN retrievals. The WECANN and FLUXNET-MTE approaches are both based on machine learning, although the FLUXNET-MTE retrievals use a regression tree rather than an ANN. Nevertheless, this commonality of methods may also contribute to the greater correspondence between these two datasets.

The retrieval maps indicate that LE, H, and GPP have similar seasonal variability and distribution, which are determined by the annual phenological cycle in energy-limited northern latitude regions, dryness in Mediterranean and monsoonal climates, and by light availability in rainforests. Seasonal radiation has a great impact for all three variables on some regions, such as the eastern US, Europe, and East Asia, which have wet conditions, are highly vegetated and located in midlatitudes. Conversely, the seasonal variability in LE, H, and GPP in some low-latitude and wet condition regions, such as Amazon rainforest, southern Africa, and Southeast Asia, as well as some low-latitude arid regions, such as the southwestern US, Western Australia, North Africa, and western Asia, are not significant, as there is less seasonal solar radiation variability in the aforementioned regions. Comparison between the LE, H, and GPP, shows that they all demonstrate generally similar patterns of seasonal variability through time.

We also assessed the impact of SIF on retrieval quality. In comparison to optical-based vegetation indices, SIF has better performance in regions where phenology and incident radiation are not the main contributors to flux variability, while it has similar performance in other regions.

From the evaluation results compared with FLUXNET tower observations, it is noted that WECANN has better performance compared to other global products. LE and H estimates from WECANN are more consistent with tower observations compared to GPP. WECANN retrievals have better correlation with tower observations at 76 % of sites for LE, 58 % of sites for H, and 55 % of sites for GPP compared to other products. Moreover, retrievals from WECANN outperform other global products in capturing the seasonality of surface fluxes and GPP across a wide range of sites with different climatic and biome conditions.

We also assessed the performance of WECANN in capturing extreme heat wave and drought events and showed that in the case of the 2010 Russia heat wave, 2011 Texas drought, and 2012 US Corn Belt drought, WECANN properly captures the extent of the anomalies in LE, H, and GPP. Moreover, an independent ET estimate from a water budget closure model was used to evaluate WECANN ET estimates

across five large basins, and it showed small to moderate errors for WECANN retrievals.

The Supplement related to this article is available online at https://doi.org/10.5194/bg-14-4101-2017-supplement.

References

Aires, F.: Combining Datasets of Satellite-Retrieved Products, Part I: Methodology and Water Budget Closure, J. Hydrometeorol., 15, 1677–1691, https://doi.org/10.1175/JHM-D-13-0148.1, 2014.

Aires, F., Prigent, C., and Rossow, W. B.: Sensitivity of satellite microwave and infrared observations to soil moisture at a global scale: 2. Global statistical relationships, J. Geophys. Res., 110, D11103, https://doi.org/10.1029/2004JD005094, 2005.

Aires, F., Aznay, O., Prigent, C., Paul, M., and Bernardo, F.: Synergistic multi-wavelength remote sensing versus a posteriori combination of retrieved products: Application for the retrieval of atmospheric profiles using MetOp-A, J. Geophys. Res.-Atmos., 117, D18304, https://doi.org/10.1029/2011JD017188, 2012.

Alemohammad, S. H., McColl, K. A., Konings, A. G., Entekhabi, D., and Stoffelen, A.: Characterization of precipitation product errors across the United States using multiplicative triple collocation, Hydrol. Earth Syst. Sci., 19, 3489–3503, https://doi.org/10.5194/hess-19-3489-2015, 2015.

Anber, U., Gentine, P., Wang, S., and Sobel, A. H.: Fog and rain in the Amazon, P. Natl. Acad. Sci. USA, 112, 11473–11477, https://doi.org/10.1073/pnas.1505077112, 2015.

Aumann, H. H., Chahine, M. T., Gautier, C., Goldberg, M. D., Kalnay, E., McMillin, L. M., Revercomb, H., Rosenkranz, P. W., Smith, W. L., Staelin, D. H., Strow, L. L., and Susskind, J.: AIRS/AMSU/HSB on the Aqua mission: design, science objectives, data products, and processing systems, IEEE Trans. Geosci. Remote Sens., 41, 253–264, https://doi.org/10.1109/TGRS.2002.808356, 2003.

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, Bull. Am. Meteorol. Soc., 82, 2415–2434, https://doi.org/10.1175/1520-0477(2001)082<2415:FANTTS>2.3.CO;2, 2001.

Balsamo, G., Beljaars, A., Scipal, K., Viterbo, P., van den Hurk, B., Hirschi, M., and Betts, A. K.: A Revised Hydrology for the ECMWF Model: Verification from Field Site to Terrestrial Water Storage and Impact in the Integrated Forecast System, J. Hydrometeorol., 10, 623–643, https://doi.org/10.1175/2008JHM1068.1, 2009.

Beer, C., Reichstein, M., Tomelleri, E., Ciais, P., Jung, M., Carvalhais, N., Rodenbeck, C., Arain, M. A., Baldocchi, D., Bonan, G.

B., Bondeau, A., Cescatti, A., Lasslop, G., Lindroth, A., Lomas, M., Luyssaert, S., Margolis, H., Oleson, K. W., Roupsard, O., Veenendaal, E., Viovy, N., Williams, C., Woodward, F. I., and Papale, D.: Terrestrial Gross Carbon Dioxide Uptake: Global Distribution and Covariation with Climate, Science, 329, 834–838, https://doi.org/10.1126/science.1184984, 2010.

Boyer, J. S., Byrne, P., Cassman, K. G., Cooper, M., Delmer, D., Greene, T., Gruis, F., Habben, J., Hausmann, N., Kenny, N., Lafitte, R., Paszkiewicz, S., Porter, D., Schlegel, A., Schussler, J., Setter, T., Shanahan, J., Sharp, R. E., Vyn, T. J., Warner, D., and Gaffney, J.: The U.S. drought of 2012 in perspective: A call to action, Global Food Security, 2, 139–143, https://doi.org/10.1016/j.gfs.2013.08.002, 2013.

Damour, G., Simonneau, T., Cochard, H., and Urban, L.: An overview of models of stomatal conductance at the leaf level, Plant. Cell Environ., 33, 1419–1438, https://doi.org/10.1111/j.1365-3040.2010.02181.x, 2010.

da Rocha, H. R., Manzi, A. O., Cabral, O. M., Miller, S. D., Goulden, M. L., Saleska, S. R., R.-Coupe, N., Wofsy, S. C., Borma, L. S., Artaxo, P., Vourlitis, G., Nogueira, J. S., Cardoso, F. L., Nobre, A. D., Kruijt, B., Freitas, H. C., von Randow, C., Aguiar, R. G., and Maia, J. F.: Patterns of water and heat flux across a biome gradient from tropical forest to savanna in Brazil, J. Geophys. Res., 114, G00B12, https://doi.org/10.1029/2007JG000640, 2009.

DeLucia, E. H. and Heckathorn, S. A.: The effect of soil drought on water-use efficiency in a contrasting Great Basin desert and Sierran montane species, Plant, Cell Environ., 12, 935–940, https://doi.org/10.1111/j.1365-3040.1989.tb01973.x, 1989.

Dewar, R. C.: The Ball-Berry-Leuning and Tardieu-Davies stomatal models: synthesis and extension within a spatially aggregated picture of guard cell function, Plant Cell Environ., 25, 1383–1398, https://doi.org/10.1046/j.1365-3040.2002.00909.x, 2002.

D'Odorico, P., Gonsamo, A., Pinty, B., Gobron, N., Coops, N., Mendez, E., and Schaepman, M. E.: Intercomparison of fraction of absorbed photosynthetically active radiation products derived from satellite data over Europe, Remote Sens. Environ., 142, 141–154, https://doi.org/10.1016/j.rse.2013.12.005, 2014.

Dole, R., Hoerling, M., Perlwitz, J., Eischeid, J., Pegion, P., Zhang, T., Quan, X.-W., Xu, T., and Murray, D.: Was there a basis for anticipating the 2010 Russian heat wave?, Geophys. Res. Lett., 38, L06702, https://doi.org/10.1029/2010GL046582, 2011.

Dorigo, W.: ESA CCI Soil Moisture for improved Earth system understanding: state-of-the art and future directions, Remote Sens. Environ., https://doi.org/10.1016/j.rse.2017.07.001, 2017.

Drusch, M., Moreno, J., Del Bello, U., Franco, R., Goulas, Y., Huth, A., Kraft, S., Middleton, E. M., Miglietta, F., Mohammed, G., Nedbal, L., Rascher, U., Schuttemeyer, D., and Verhoef, W.: The FLuorescence EXplorer Mission Concept-ESA's Earth Explorer 8, IEEE Trans. Geosci. Remote Sens., 1–12, https://doi.org/10.1109/TGRS.2016.2621820, 2016.

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 Sens. Environ., 112, 901–919, https://doi.org/10.1016/j.rse.2007.06.025, 2008.

Flexas, J., Escalona, J. M., Evain, S., Gulias, J., Moya, I., Osmond, C. B., and Medrano, H.: Steady-state chlorophyll fluorescence (Fs) measurements as a tool to follow varia-

tions of net CO_2 assimilation and stomatal conductance during water-stress in C3 plants, Physiol. Plant., 114, 231–240, https://doi.org/10.1034/j.1399-3054.2002.1140209.x, 2002.

Foken, T., Mauder, M., Liebethal, C., Wimmer, F., Beyrich, F., Leps, J.-P., Raasch, S., DeBruin, H. A. R., Meijninger, W. M. L., and Bange, J.: Energy balance closure for the LITFASS-2003 experiment, Theor. Appl. Climatol., 101, 149–160, https://doi.org/10.1007/s00704-009-0216-8, 2010.

Frankenberg, C., Fisher, J. B., Worden, J., Badgley, G., Saatchi, S. S., Lee, J.-E., Toon, G. C., Butz, A., Jung, M., Kuze, A., and Yokota, T.: New global observations of the terrestrial carbon cycle from GOSAT: Patterns of plant fluorescence with gross primary productivity, Geophys. Res. Lett., 38, L17706, https://doi.org/10.1029/2011GL048738, 2011.

Frankenberg, C., O'Dell, C., Guanter, L., and McDuffie, J.: Remote sensing of near-infrared chlorophyll fluorescence from space in scattering atmospheres: implications for its retrieval and interferences with atmospheric CO_2 retrievals, Atmos. Meas. Tech., 5, 2081–2094, https://doi.org/10.5194/amt-5-2081-2012, 2012.

Frankenberg, C., O'Dell, C., Berry, J., Guanter, L., Joiner, J., Köhler, P., Pollock, R., and Taylor, T. E.: Prospects for chlorophyll fluorescence remote sensing from the Orbiting Carbon Observatory-2, Remote Sens. Environ., 147, 1–12, https://doi.org/10.1016/j.rse.2014.02.007, 2014.

Gebremichael, M., Krajewski, W. F., Morrissey, M. L., Huffman, G. J., Adler, R. F., Gebremichael, M., Krajewski, W. F., Morrissey, M. L., Huffman, G. J., and Adler, R. F.: A Detailed Evaluation of GPCP 1° Daily Rainfall Estimates over the Mississippi River Basin, J. Appl. Meteorol., 44, 665–681, https://doi.org/10.1175/JAM2233.1, 2005.

Govindjee, Downton, W. J. S., Fork, D. C., and Armond, P. A.: Chlorophyll A fluorescence transient as an indicator of water potential of leaves, Plant Sci. Lett., 20, 191–194, https://doi.org/10.1016/0304-4211(81)90261-3, 1981.

Gruber, A., Su, C.-H., Zwieback, S., Crow, W., Dorigo, W., and Wagner, W.: Recent advances in (soil moisture) triple collocation analysis, Int. J. Appl. Earth Obs. Geoinf., 45, 200–211, https://doi.org/10.1016/j.jag.2015.09.002, 2016.

Guanter, L., Frankenberg, C., Dudhia, A., Lewis, P. E., Gómez-Dans, J., Kuze, A., Suto, H., and Grainger, R. G.: Retrieval and global assessment of terrestrial chlorophyll fluorescence from GOSAT space measurements, Remote Sens. Environ., 121, 236–251, https://doi.org/10.1016/j.rse.2012.02.006, 2012.

Guanter, L., Zhang, Y., Jung, M., Joiner, J., Voigt, M., Berry, J. A., Frankenberg, C., Huete, A. R., Zarco-Tejada, P., Lee, J.-E., Moran, M. S., Ponce-Campos, G., Beer, C., Camps-Valls, G., Buchmann, N., Gianelle, D., Klumpp, K., Cescatti, A., Baker, J. M., and Griffis, T. J.: Global and time-resolved monitoring of crop photosynthesis with chlorophyll fluorescence, P. Natl. Acad. Sci. USA, 111, E1327–E1333, https://doi.org/10.1073/pnas.1320008111, 2014.

Guillod, B. P., Orlowsky, B., Miralles, D., Teuling, A. J., Blanken, P. D., Buchmann, N., Ciais, P., Ek, M., Findell, K. L., Gentine, P., Lintner, B. R., Scott, R. L., Van den Hurk, B., and I. Seneviratne, S.: Land-surface controls on afternoon precipitation diagnosed from observational data: uncertainties and confounding factors, Atmos. Chem. Phys., 14, 8343–8367, https://doi.org/10.5194/acp-14-8343-2014, 2014.

Guillod, B. P., Orlowsky, B., Miralles, D. G., Teuling, A. J., and Seneviratne, S. I.: Reconciling spatial and temporal soil moisture effects on afternoon rainfall, Nat. Commun., 6, 6443, https://doi.org/10.1038/ncomms7443, 2015.

Hain, C. R., Crow, W. T., Mecikalski, J. R., Anderson, M. C., and Holmes, T.: An intercomparison of available soil moisture estimates from thermal infrared and passive microwave remote sensing and land surface modeling, J. Geophys. Res., 116, D15107, https://doi.org/10.1029/2011JD015633, 2011.

Havaux, M. and Lannoye, R.: Chlorophyll fluorescence induction: A sensitive indicator of water stress in maize plants, Irrig. Sci., 4, 147–151, https://doi.org/10.1007/BF00273382, 1983.

Huffman, G. J., Adler, R. F., Morrissey, M. M., Bolvin, D. T., Curtis, S., Joyce, R., McGavock, B., Susskind, J., Huffman, G. J., Adler, R. F., Morrissey, M. M., Bolvin, D. T., Curtis, S., Joyce, R., McGavock, B., and Susskind, J.: Global Precipitation at One-Degree Daily Resolution from Multisatellite Observations, J. Hydrometeorol., 2, 36–50, https://doi.org/10.1175/1525-7541(2001)002<0036:GPAODD>2.0.CO;2, 2001.

Jasechko, S., Sharp, Z. D., Gibson, J. J., Birks, S. J., Yi, Y., and Fawcett, P. J.: Terrestrial water fluxes dominated by transpiration, Nature, 496, 347–350, https://doi.org/10.1038/nature11983, 2013.

Jeong, S.-J., Schimel, D., Frankenberg, C., Drewry, D. T., Fisher, J. B., Verma, M., Berry, J. A., Lee, J.-E., and Joiner, J.: Application of satellite solar-induced chlorophyll fluorescence to understanding large-scale variations in vegetation phenology and function over northern high latitude forests, Remote Sens. Environ., 190, 178–187, https://doi.org/10.1016/j.rse.2016.11.021, 2017.

Jiang, C. and Ryu, Y.: Multi-scale evaluation of global gross primary productivity and evapotranspiration products derived from Breathing Earth System Simulator (BESS), Remote Sens. Environ., 186, 528–547, https://doi.org/10.1016/j.rse.2016.08.030, 2016.

Jiménez, C., Prigent, C., and Aires, F.: Toward an estimation of global land surface heat fluxes from multisatellite observations, J. Geophys. Res., 114, D06305, https://doi.org/10.1029/2008JD011392, 2009.

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

Jiménez, C., Clark, D. B., Kolassa, J., Aires, F., and Prigent, C.: A joint analysis of modeled soil moisture fields and satellite observations, J. Geophys. Res.-Atmos., 118, 6771–6782, https://doi.org/10.1002/jgrd.50430, 2013.

Joiner, J., Guanter, L., Lindstrot, R., Voigt, M., Vasilkov, A. P., Middleton, E. M., Huemmrich, K. F., Yoshida, Y., and Frankenberg, C.: Global monitoring of terrestrial chlorophyll fluorescence from moderate-spectral-resolution near-infrared satellite measurements: methodology, simulations, and application to GOME-2, Atmos. Meas. Tech., 6, 2803–2823, https://doi.org/10.5194/amt-6-2803-2013, 2013.

Joiner, J., Yoshida, Y., Vasilkov, A. P., Schaefer, K., Jung, M., Guanter, L., Zhang, Y., Garrity, S., Middleton, E. M., Huemmrich, K. F., Gu, L., and Belelli Marchesini, L.: The seasonal

cycle of satellite chlorophyll fluorescence observations and its relationship to vegetation phenology and ecosystem atmosphere carbon exchange, Remote Sens. Environ., 152, 375–391, https://doi.org/10.1016/j.rse.2014.06.022, 2014.

Joiner, J., Yoshida, Y., Guanter, L., and Middleton, E. M.: New methods for the retrieval of chlorophyll red fluorescence from hyperspectral satellite instruments: simulations and application to GOME-2 and SCIAMACHY, Atmos. Meas. Tech., 9, 3939–3967, https://doi.org/10.5194/amt-9-3939-2016, 2016.

Joshi, M. K., Rai, A., and Pandey, A. C.: Validation of TMPA and GPCP 1DD against the ground truth rain-gauge data for Indian region, Int. J. Climatol., 33, 2633–2648, https://doi.org/10.1002/joc.3612, 2012.

Jung, M., Reichstein, M., and Bondeau, A.: Towards global empirical upscaling of FLUXNET eddy covariance observations: validation of a model tree ensemble approach using a biosphere model, Biogeosciences, 6, 2001–2013, https://doi.org/10.5194/bg-6-2001-2009, 2009.

Jung, M., Reichstein, M., Ciais, P., Seneviratne, S. I., Sheffield, J., Goulden, M. L., Bonan, G., Cescatti, A., Chen, J., de Jeu, R., Dolman, A. J., Eugster, W., Gerten, D., Gianelle, D., Gobron, N., Heinke, J., Kimball, J., Law, B. E., Montagnani, L., Mu, Q., Mueller, B., Oleson, K., Papale, D., Richardson, A. D., Roupsard, O., Running, S., Tomelleri, E., Viovy, N., Weber, U., Williams, C., Wood, E., Zaehle, S., and Zhang, K.: Recent decline in the global land evapotranspiration trend due to limited moisture supply, Nature, 467, 951–954, https://doi.org/10.1038/nature09396, 2010.

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, J. Geophys. Res., 116, G00J07, https://doi.org/10.1029/2010JG001566, 2011.

Kato, S., Loeb, N. G., Rose, F. G., Doelling, D. R., Rutan, D. A., Caldwell, T. E., Yu, L., and Weller, R. A.: Surface Irradiances Consistent with CERES-Derived Top-of-Atmosphere Shortwave and Longwave Irradiances, J. Clim., 26, 2719–2740, https://doi.org/10.1175/JCLI-D-12-00436.1, 2013.

Kolassa, J., Aires, F., Polcher, J., Prigent, C., Jimenez, C., and Pereira, J. M.: Soil moisture retrieval from multi-instrument observations: Information content analysis and retrieval methodology, J. Geophys. Res.-Atmos., 118, 4847–4859, https://doi.org/10.1029/2012JD018150, 2013.

Kolassa, J., Gentine, P., Prigent, C., and Aires, F.: Soil moisture retrieval from AMSR-E and ASCAT microwave observation synergy, Part 1: Satellite data analysis, Remote Sens. Environ., 173, 1–14, https://doi.org/10.1016/j.rse.2015.11.011, 2016.

Krause, G. H. and Weis, E.: Chlorophyll Fluorescence and Photosynthesis: The Basics, Annu. Rev. Plant Physiol. Plant Mol. Biol., 42, 313–349, https://doi.org/10.1146/annurev.pp.42.060191.001525, 1991.

Landerer, F. W. and Swenson, S. C.: Accuracy of scaled GRACE terrestrial water storage estimates, Water Resour. Res., 48, W04531, https://doi.org/10.1029/2011WR011453, 2012.

Lau, W. K. M. and Kim, K.-M.: The 2010 Pakistan Flood and Russian Heat Wave: Teleconnection of Hydrometeorological Extremes, J. Hydrometeorol., 13, 392–403, https://doi.org/10.1175/JHM-D-11-016.1, 2012.

Lee, J.-E., Frankenberg, C., van der Tol, C., Berry, J. A., Guanter, L., Boyce, C. K., Fisher, J. B., Morrow, E., Worden, J. R., Asefi, S., Badgley, G., and Saatchi, S.: Forest productivity and water stress in Amazonia: observations from GOSAT chlorophyll fluorescence, P. R. Soc. B, 280, 20130171–20130171, https://doi.org/10.1098/rspb.2013.0171, 2013.

Lee, J.-E., Berry, J. A., van der Tol, C., Yang, X., Guanter, L., Damm, A., Baker, I., and Frankenberg, C.: Simulations of chlorophyll fluorescence incorporated into the Community Land Model version 4, Glob. Change Biol., 21, 3469–3477, https://doi.org/10.1111/gcb.12948, 2015.

Lei, F., Crow, W., Shen, H., Parinussa, R., and Holmes, T.: The Impact of Local Acquisition Time on the Accuracy of Microwave Surface Soil Moisture Retrievals over the Contiguous United States, Remote Sens., 7, 13448–13465, https://doi.org/10.3390/rs71013448, 2015.

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

Liu, Y. Y., Dorigo, W. A., Parinussa, R. M., de Jeu, R. A. M., Wagner, W., McCabe, M. F., Evans, J. P., and van Dijk, A. I. J. M.: Trend-preserving blending of passive and active microwave soil moisture retrievals, Remote Sens. Environ., 123, 280–297, https://doi.org/10.1016/j.rse.2012.03.014, 2012.

Loeb, N. G., Wielicki, B. A., Doelling, D. R., Smith, G. L., Keyes, D. F., Kato, S., Manalo-Smith, N., and Wong, T.: Toward Optimal Closure of the Earth's Top-of-Atmosphere Radiation Budget, J. Clim., 22, 748–766, https://doi.org/10.1175/2008JCLI2637.1, 2009.

Luo, L. and Zhang, Y.: Did we see the 2011 summer heat wave coming?, Geophys. Res. Lett., 39, L09708, https://doi.org/10.1029/2012GL051383, 2012.

Luojus, K., Pulliainen, J., Takala, M., Lemmetyinen, J., Kangwa, M., Smolander, T., and Derksen, C.: Global snow monitoring for climate research: Algorithm theoretical basis document (ATBD) – SWE Algorithm, Version/Revision 1.0/02, 2013.

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

McColl, K. A., Vogelzang, J., Konings, A. G., Entekhabi, D., Piles, M., and Stoffelen, A.: Extended triple collocation: Estimating errors and correlation coefficients with respect to an unknown target, Geophys. Res. Lett., 41, GL061322, https://doi.org/10.1002/2014GL061322, 2014.

McColl, K. A., Roy, A., Derksen, C., Konings, A. G., Alemohammed, S. H., and Entekhabi, D.: Triple collocation for binary and categorical variables: Application to validating landscape freeze/thaw retrievals, Remote Sens. Environ., 176, 31–42, https://doi.org/10.1016/j.rse.2016.01.010, 2016.

McFarlane, J. C., Watson, R. D., Theisen, A. F., Jackson, R. D., Ehrler, W. L., Pinter, P. J., Idso, S. B., and Reginato, R. J.: Plant stress detection by remote measurement of fluorescence, Appl. Opt., 19, 3287, https://doi.org/10.1364/AO.19.003287, 1980.

McPhee, J., Margulis, S. A., McPhee, J., and Margulis, S. A.: Validation and Error Characterization of the GPCP-1DD Precipitation Product over the Contiguous United States, J. Hydrometeorol., 6, 441–459, https://doi.org/10.1175/JHM429.1, 2005.

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

Miralles, D. G., Crow, W. T., and Cosh, M. H.: Estimating Spatial Sampling Errors in Coarse-Scale Soil Moisture Estimates Derived from Point-Scale Observations, J. Hydrometeorol., 11, 1423–1429, https://doi.org/10.1175/2010JHM1285.1, 2010.

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

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

Miralles, D. G., van den Berg, M. J., Gash, J. H., Parinussa, R. M., de Jeu, R. A. M., Beck, H. E., Holmes, T. R. H., Jiménez, C., Verhoest, N. E. C., Dorigo, W. A., Teuling, A. J., and Johannes Dolman, A.: El Niño–La Niña cycle and recent trends in continental evaporation, Nature Climate Change, 4, 122–126, https://doi.org/10.1038/nclimate2068, 2014a.

Miralles, D. G., Teuling, A. J., van Heerwaarden, C. C., and Vilà-Guerau de Arellano, J.: Mega-heatwave temperatures due to combined soil desiccation and atmospheric heat accumulation, Nat. Geosci., 7, 345–349, https://doi.org/10.1038/ngeo2141, 2014b.

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

Monteith, J. L. and Moss, C. J.: Climate and the Efficiency of Crop Production in Britain, Philos. Trans. R. Soc. B Biol. Sci., 281, 277–294, https://doi.org/10.1098/rstb.1977.0140, 1977.

Morton, D. C., Nagol, J., Carabajal, C. C., Rosette, J., Palace, M., Cook, B. D., Vermote, E. F., Harding, D. J., and North, P. R. J.: Amazon forests maintain consistent canopy structure and greenness during the dry season, Nature, 506, 221–224, https://doi.org/10.1038/nature13006, 2014.

Morton, D. C., Rubio, J., Cook, B. D., Gastellu-Etchegorry, J.-P., Longo, M., Choi, H., Hunter, M., and Keller, M.: Amazon forest structure generates diurnal and seasonal variability in light utilization, Biogeosciences, 13, 2195–2206, https://doi.org/10.5194/bg-13-2195-2016, 2016.

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

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

Munier, S. and Aires, F.: A new global method of satellite dataset merging and quality characterization constrained by the terrestrial water cycle budget, Remote Sens. Environ., in review, 2017.

Munier, S., Aires, F., Schlaffer, S., Prigent, C., Papa, F., Maisongrande, P., and Pan, M.: Combining data sets of satellite-retrieved products for basin-scale water balance study: 2. Evaluation on the Mississippi Basin and closure correction model, J. Geophys. Res.-Atmos., 119, 12100–12116, https://doi.org/10.1002/2014JD021953, 2014.

NASA LP DAAC: Land Cover Type Yearly L3, MCD12C1, V051, 2016.

Pan, X., Liu, Y., and Fan, X.: Comparative Assessment of Satellite-Retrieved Surface Net Radiation: An Examination on CERES and SRB Datasets in China, Remote Sens., 7, 4899–4918, https://doi.org/10.3390/rs70404899, 2015.

Parinussa, R. M., Meesters, A. G. C. A., Liu, Y. Y., Dorigo, W., Wagner, W., and de Jeu, R. A. M.: Error Estimates for Near-Real-Time Satellite Soil Moisture as Derived From the Land Parameter Retrieval Model, IEEE Geosci. Remote Sens. Lett., 8, 779–783, https://doi.org/10.1109/LGRS.2011.2114872, 2011.

Pastorello, G., Agarwal, D., Papale, D., Samak, T., Trotta, C., Ribeca, A., Poindexter, C., Faybishenko, B., Gunter, D., Hollowgrass, R., and Canfora, E.: Observational Data Patterns for Time Series Data Quality Assessment, in 2014 IEEE 10th International Conference on e-Science, 271–278, 2014.

Pulliainen, J.: Mapping of snow water equivalent and snow depth in boreal and sub-arctic zones by assimilating space-borne microwave radiometer data and ground-based observations, Remote Sens. Environ., 101, 257–269, https://doi.org/10.1016/j.rse.2006.01.002, 2006.

Restrepo-Coupe, N., da Rocha, H. R., Hutyra, L. R., da Araujo, A. C., Borma, L. S., Christoffersen, B., Cabral, O. M. R., de Camargo, P. B., Cardoso, F. L., da Costa, A. C. L., Fitzjarrald, D. R., Goulden, M. L., Kruijt, B., Maia, J. M. F., Malhi, Y. S., Manzi, A. O., Miller, S. D., Nobre, A. D., von Randow, C., Sá, L. D. A., Sakai, R. K., Tota, J., Wofsy, S. C., Zanchi, F. B., and Saleska, S. R.: What drives the seasonality of photosynthesis across the Amazon basin? A cross-site analysis of eddy flux tower measurements from the Brasil flux network, Agr. Forest Meteorol., 182/183, 128–144, https://doi.org/10.1016/j.agrformet.2013.04.031, 2013.

Rodrìguez-Fernández, N. J., Aires, F., Richaume, P., Kerr, Y. H., Prigent, C., Kolassa, J., Cabot, F., Jimenez, C., Mahmoodi, A., and Drusch, M.: Soil Moisture Retrieval Using Neural Networks: Application to SMOS, IEEE Trans. Geosci. Remote Sens., 53, 5991–6007, https://doi.org/10.1109/TGRS.2015.2430845, 2015.

Rubel, F., Skomorowski, P., and Rudolf, B.: Verification scores for the operational GPCP-1DD product over the European

Alps, Meteorol. Z., 11, 367–370, https://doi.org/10.1127/0941-2948/2002/0011-0367, 2002.

Running, S. W., Nemani, R. R., Heinsch, F. A., Zhao, M., Reeves, M., and Hashimoto, H.: A Continuous Satellite-Derived Measure of Global Terrestrial Primary Production, Bioscience, 54, 547, https://doi.org/10.1641/0006-3568(2004)054[0547:ACSMOG]2.0.CO;2, 2004.

Saleska, S. R., Wu, J., Guan, K., Araujo, A. C., Huete, A., Nobre, A. D., and Restrepo-Coupe, N.: Dry-season greening of Amazon forests, Nature, 531, E4–E5, https://doi.org/10.1038/nature16457, 2016.

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

Stoffelen, A.: Toward the true near-surface wind speed: Error modeling and calibration using triple collocation, J. Geophys. Res., 103, 7755, https://doi.org/10.1029/97JC03180, 1998.

Svoboda, M., LeComte, D., Hayes, M., Heim, R., Gleason, K., Angel, J., Rippey, B., Tinker, R., Palecki, M., Stooksbury, D., Miskus, D., Stephens, S., Svoboda, M., LeComte, D., Hayes, M., Heim, R., Gleason, K., Angel, J., Rippey, B., Tinker, R., Palecki, M., Stooksbury, D., Miskus, D., and Stephens, S.: The Drought Monitor, Bull. Am. Meteorol. Soc., 83, 1181–1190, https://doi.org/10.1175/1520-0477(2002)083<1181:TDM>2.3.CO;2, 2002.

Swenson, S. and Wahr, J.: Post-processing removal of correlated errors in GRACE data, Geophys. Res. Lett., 33, L08402, https://doi.org/10.1029/2005GL025285, 2006.

Toivonen, P. and Vidaver, W.: Variable Chlorophyll a Fluorescence and CO2 Uptake in Water-Stressed White Spruce Seedlings, Plant Physiol., 86, 744–748, https://doi.org/10.1104/pp.86.3.744, 1988.

Tramontana, G., Jung, M., Schwalm, C. R., Ichii, K., Camps-Valls, G., Ráduly, B., Reichstein, M., Arain, M. A., Cescatti, A., Kiely, G., Merbold, L., Serrano-Ortiz, P., Sickert, S., Wolf, S., and Papale, D.: Predicting carbon dioxide and energy fluxes across global FLUXNET sites with regression algorithms, Biogeosciences, 13, 4291–4313, https://doi.org/10.5194/bg-13-4291-2016, 2016.

USDA: Crop Production 2012 Summary, http://usda.mannlib.cornell.edu/usda/nass/CropProdSu/2010s/2013/CropProdSu-01-11-2013.pdf, 2013.

van der Tol, C., Verhoef, W., and Rosema, A.: A model for chlorophyll fluorescence and photosynthesis at leaf scale, Agr. Forest Meteorol., 149, 96–105, https://doi.org/10.1016/j.agrformet.2008.07.007, 2009.

van Kooten, O. and Snel, J. F. H.: The use of chlorophyll fluorescence nomenclature in plant stress physiology, Photosynth. Res., 25, 147–150, https://doi.org/10.1007/BF00033156, 1990.

Wagner, W., Dorigo, W., de Jeu, R., Fernandez, D., Benveniste, J., Haas, E., and Ertl, M.: Fusion of Active and Passive Microwave Observations To Create an Essential Climate Variable Data Record on Soil Moisture, ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci., I-7 (September), 315–321, https://doi.org/10.5194/isprsannals-I-7-315-2012, 2012.

Walther, S., Voigt, M., Thum, T., Gonsamo, A., Zhang, Y., Koehler, P., Jung, M., Varlagin, A., and Guanter, L.: Satellite chlorophyll fluorescence measurements reveal large-scale decoupling of photosynthesis and greenness dynamics in boreal evergreen forests, Glob. Change Biol., 22, 2979–2996, https://doi.org/10.1111/gcb.13200, 2016.

Wielicki, B. A., Barkstrom, B. R., Harrison, E. F., Lee, R. B., Louis Smith, G., and Cooper, J. E.: Clouds and the Earth's Radiant Energy System (CERES): An Earth Observing System Experiment, Bull. Am. Meteorol. Soc., 77, 853–868, https://doi.org/10.1175/1520-0477(1996)077<0853:CATERE>2.0.CO;2, 1996.

Xu, L., Saatchi, S. S., Yang, Y., Myneni, R. B., Frankenberg, C., Chowdhury, D., and Bi, J.: Satellite observation of tropical forest seasonality: spatial patterns of carbon exchange in Amazonia, Environ. Res. Lett., 10, 84005, https://doi.org/10.1088/1748-9326/10/8/084005, 2015.

Zhang, Y., Guanter, L., Berry, J. A., van der Tol, C., Yang, X., Tang, J., and Zhang, F.: Model-based analysis of the relationship between sun-induced chlorophyll fluorescence and gross primary production for remote sensing applications, Remote Sens. Environ., 187, 145–155, https://doi.org/10.1016/j.rse.2016.10.016, 2016a.

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

Zhao, M. and Running, S. W.: Drought-Induced Reduction in Global Terrestrial Net Primary Production from 2000 Through 2009, Science, 329, 940–943, https://doi.org/10.1126/science.1192666, 2010.

Zhao, M., Heinsch, F. A., Nemani, R. R., and Running, S. W.: Improvements of the MODIS terrestrial gross and net primary production global data set, Remote Sens. Environ., 95, 164–176, https://doi.org/10.1016/j.rse.2004.12.011, 2005.