Response to Referee 1

Water, Energy, and Carbon with Artificial Neural Networks (WECANN): A statistically-based estimate of global surface turbulent fluxes using solar-induced fluorescence (Manuscript # bg-2016-495)

Comments	Responses/Actions
Thank you for the efforts in trying to address the questions raised during the review process. I think the manuscript has been improved, specifically with the inclusion of the NDVI/EVI tests, the larger fluxnet dataset for validation and the analysis on the uncertainty estimation. However, I am still not satisfied on some issues. In particular on two points that were raised by both reviewers: (1) the circularity of using flux data in both the training and the validation and (2) the mismatch between the footprint of the towers and the 1 degree cells of the products. For both points, I would strongly encourage including the analyses suggested by Reviewer 1, that is: for point 1 to do an evaluation of ET at basin scale from water balance; and for point 2 to focus on an evaluation at larger scale such as catching big climate events, and perhaps the inter-annual variability of GPP and comparing them perhaps with atmospheric inversion estimates. I believe that these will be much more convincing in showing the added value of WECANN that the current individual pixel to tower comparisons/validations that are done. These could be placed in supplementary material in my opinion, if the length is really an issue.	We thank the referee for their comments and suggestions. In order to address these critiques, we have implemented the following three validation analyses: 1- Evaluating WECANN estimates in three extreme heatwave events: Russia 2010, Texas 2011, US corn belt 2013 (Section 4.4 in the new manuscript). 2- Estimating the error of WECANN ET estimates in five large basins (Amazon, Colorado, Congo, Mississippi, and Orinoco) using data from a water budget closure model (Section 4.5 in the new manuscript). 3- Calculating correlation between GRACE estimates of terrestrial water storage changes and P-ET from WECANN (Section 4.6 in the new manuscript). Results of these analyses show the capability of WECANN retrievals in capturing large-scale anomalies.
I do realize that both suggestions of Reviewer 1 where countered in part by using the argument that I (reviewer 2) asked to reduce the length of the manuscript. I am afraid that I was misunderstood on this point. I did not mean the whole manuscript should be reduced per se, but rather that some specific parts were unnecessarily lengthy. Large parts describing the plots and maps could (and still can in my opinion) be reduced, mostly the 'Results and Discussion' section. Also, phrases like "In this section, we [do	We changed the text throughout the manuscript to reduce the unnecessary sentences as needed.

this]… " after the subtitle saying precisely what 'this' is just adds unnecessary length and could be reformulated in my opinion e.g. top of page 9 and page 10. Bottom-line is that tackling the two points raised above should have a priority over reducing the length of the paper as a whole.	
Regarding the other arguments for keeping to a 'standard' comparison of 1 degree vs fluxtower footprints, such as the this is how it is often done or that fluxtowers is the only source of validation GPP, I believe now would be the time to raise the issue more prominently that doing these matches is sub optimal as we are unable to disentangle the signal from the confounding factors caused by intra-pixel spatial heterogeneity. The current manuscript would benefit from taking that point up front and show a more innovative 'validation' using the suggestions of reviewer 1. And I repeat, the current work on individual towers should be kept, as it is informative, but perhaps placed in supplementary if space is lacking.	We agree with the referee on the shortcomings of the FLUXNET tower comparisons given the coarse spatial resolution of WECANN retrievals. We have summarized the results of this comparison in a shorter section (4.2) now with only 5 flux towers being discussed in details instead of 17, and put more focus on the large scale validation efforts that are introduced in the new version. We have also summarized the results of flux tower comparisons in the new Figure 6 based on Plant Functional Types.
Other remarks/suggestions	
The comparison against a set of 97 fluxsite is welcome, but this should come to the forefront, instead of being relegated to supplementary material. Not with the full details of all individual fluxtower comparison (which could stay in the supplementary material), but with a synthetic table with the average values for each flux. In this table also include the average values for the RSME and Bias. You could also consider using a symmetric index of agreement (https://www.nature.com/articles/srep19401) as a generic value from 0 to 1 to specify the degree of agreement. This combines together the effects of RMSE, bias and correlation in a single metric, which also does not assume a reference. Indeed, here it could be argued that the fluxtower estimate is not an absolute reference because it is not measured on the same observational support as the 1 degree WECANN estimate.	As noted in the previous comment, we have summarized the flux tower comparison section and brought it up into the first evaluation section (4.2) following presenting the seasonal patterns from WECANN retrievals. We have added the average RMSE of each product in Tables S1-S6. But we tend to not report the average bias, because it can be misleading. For example, if a product has large negative and positive biases across different sites a mean value close to zero it does not mean that it has low bias compared to another product that has only small positive biases that on average are larger than the mean of the first product.
The point on C3/C4 plants raised by reviewer 1 is not explicitly addressed in the text. I would expect a phrase talking about it somewhere.	We added text to Page 9 Lines 20-23 in the current manuscript to explain this further.
Page 2, Line 31: you would need to specify that you " jointly retrieve surface turbulent fluxes" for	Thanks for the note on this. We have revised the text accordingly.

the first time using SIF, as GPP has been retrieved from SIF on many occasions	
Uncertainty analysis is very welcome. Clarify in the caption of figure 14 that these are the annual mean estimates of each flux with their uncertainty. As it is now, it seems the centre of the box plot represent the mean uncertainty instead of the mean flux.	We changed the figure caption to clarify the data presented in this figure.
You say the decreasing trends in GPP and LE are consistent in figure 14. They are consistent with each other (with is to be expected from your product) but is it consistent with the general knowledge of what is really happening across the globe (a marked reduction of GPP since 2011) ? Please put this in context with other studies. Also, how do you explain the strange pattern of sudden decrease in 2007-2009 for GPP in the Mid latitudes?	While we believe analysis of the trends in the retrievals is beyond the scope of this manuscript (and it is the focus of our next research study being carried out), here we present a figure to compare the yearly anomalies of LE, H, and GPP across the globe and different latitudinal zones and compare them with FLUXNET-MTE (Figure R1). Since FLUXNET-MTE is only available until 2011, we can't compare the anomalies from 2012-2015. However, comparison of the results from 2007-2011 shows consistent anomalies and trends in both datasets.
In figure S1, there seems to be a strange asymmetric effect of your land mask, e.g. there is a white line on the west coast of Africa and South America, but not on the East coast. Are the datasets properly aligned?	Thanks for noting this issue. This was caused by an error in using the plot function in MATLAB. It is corrected in the new version.

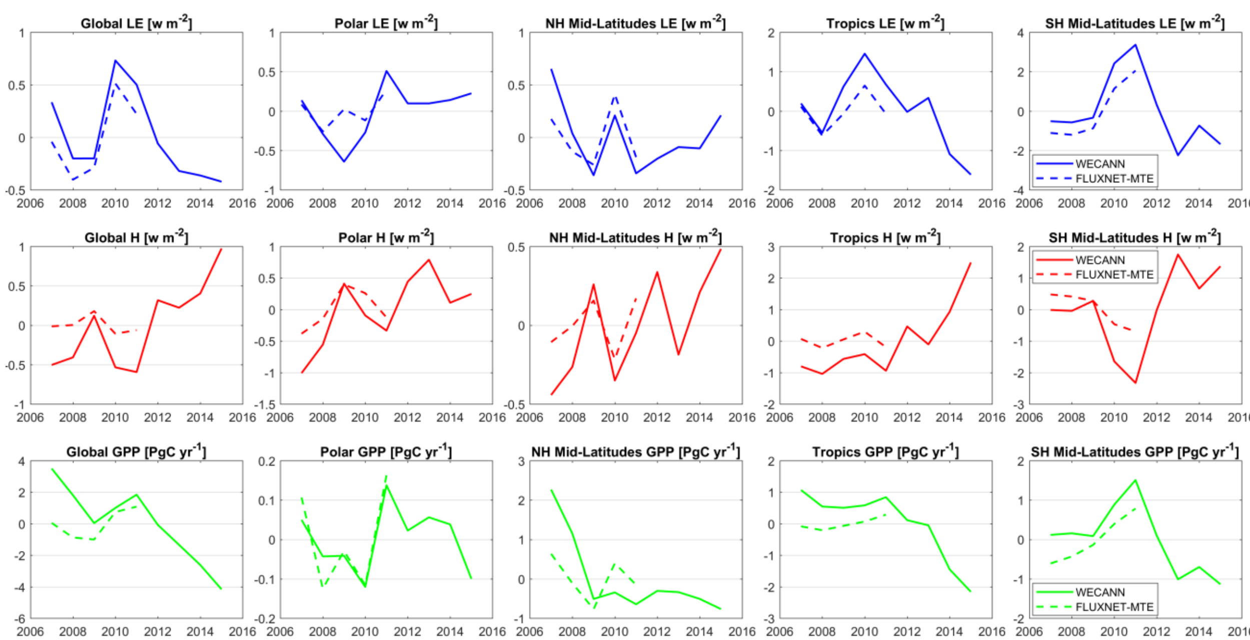


Figure R1- Annual mean anomalies of LE (top row), H (middle row) and GPP (bottom row) retrievals at global (left column) and regional (four right columns) scales for WECANN (solid lines) and FLUXNET-MTE (dashed lines).

Response to Referee 2

Water, Energy, and Carbon with Artificial Neural Networks (WECANN): A statistically-based estimate of global surface turbulent fluxes using solar-induced fluorescence (Manuscript # bg-2016-495)

Comments	Responses/Actions
Editor comments. Failing to find a suitable reviewer in time, I've made my own assessment. I largely agree with the other reviewer, but would like to point our that in addition I find this manuscript to be weak in discussing the implication of the findings in context to other studies. The reviewer and my comments may give a direction how to improve this, in general I think what is needed is a criticial assessment of what one can learn from this development, which we did not know in the past, and how for instance the trend estimates WECANN produces compare to other studies (or whether they maybe should not be regarded as trends given the brevity of the record.) One fairly simple addon would be to plot the other products as well in Figure 14, to have a more synthetic overview.	We thank the editor for his comments and suggestions. We have addressed each of the comments in the following and implemented necessary changes in the manuscript to address them. One of the key findings of this study is showing the capability of SIF measurements to be used for estimating latent and sensible heat fluxes along with GPP. GPP retrievals from SIF have been studied before, but our retrieval is the first of its kind to estimates latent and sensible heat. This finding is of interest to a large community of researchers in our field, in particular science teams of OCO-2, OCO-3, ECOSTRESS, geoCARB, and FLEX missions. During the last couple of months we have been invited to their science team meetings and presented our results on this topic which was highly received by their team members.
I would suggest to keep some (say 4 site from different biomes) of the FLUXNET comparison plots in the main manuscript (and in agreement to the other reviewer, but them more at the front of the results), and put the others into the Appendix	We have reduced the flux tower comparison section by including only 6 sites, and summarizing the statistics from Tables S1-S3 in Figure 6 based on Plant Functional Types.
Minor comments	
P1 L29: GPP and NPP are not generally considered turbulent exchanges, nor are they a flux from the land to the atmosphere. The turbulent flux observed is the net ecosystem exchange rate, of which GPP is a compartment flux. I also don't see the reason to include NPP in this list if it is not subject of this study.	We removed NPP from the discussion.
P1 L34 redundant sentence	We revised this in the new version.
P2 L1: this is not only the MPI-BGC, see Tramontana, G. et al. Predicting carbon dioxide	We revised the text here to appropriately refer to this product.

and energy fluxes across global FLUXNET sites with regression algorithms. Biogeosciences 13, 4291-4313 (2016).	
P2 L9ff: To be fair to the regression approach discussed above, this paragraph needs to mention that any model product that is used to train an ANN necessarily inherits all its imperfections in capturing means, extremes and so on to the ANN.	We already have mentioned this further down in the same paragraph, Page 2 Lines 14-16 in the new version.
P2 L20: It is necessary here to introduce the training data sources (at least in an overview).	Introduction on training data is added. (this is now in Page 3, Lines 7-8).
P2 L30: its is not clear what the first key innovation is at this point. Also SIF has been introduced already. Please edits the text such that the information flow is more logical, e.g. by mentioning SIF first here.	We revised the content in the first section to appropriately address this issue.
P3 L9: True, but how does this work in an ANN, which does not intrinsically know about this coupling?	The way ANN develops such a coupling implicitly is by using one set of weights and biases for its neurons to estimate the three fluxes together. These weights and biases are tuned during the training. Therefore, having a good target dataset for training is essential to develop an accurate network.
P3 L14-18 can be safely removed.	We have removed this section.
P3 L 22 why 1° when all input data are at a higher resolution?	As listed in Table 2, three of the input datasets (Air Temperature, Net Radiation, and Precipitation) are posted on 1° spatial grids. Therefore, the final product can't be on a higher spatial resolution.
P6 L8ff. Please clarify which data has been used, and why. It is probably best to include a supplementary table, listing all sites, data coverage and their sources.	As we have explained later in this section, we are using data accessed through the FLUXNET 2015 dataset. This is a product developed by the PIs of all flux towers contributing data, and all the specifications of data are specified in the product itself. So it would be redundant to list them again here. Based on the request of PIS, we have appropriately cited the data, acknowledged them in the acknowledgement section and used the correct abbreviation for site names.
P7 L20. To show that the method adds information, would it be helpful to compare the product against the point-wise average of the three input data sources?	As our results show WECANN has the best performance among the four products considered in this study. WECANN is developed by characterizing the errors of each of the training products and learning from them collectively while minimizing errors. Moreover, as our TC analysis shows error of each of the training products change from one pixel to another and there is no single product that has the lowest/highest error at all pixels. Therefore, a simple averaging of the pixel-wise estimates will not provide a reasonable estimate of the fluxes at all pixels and times.

P8 L13. Help the reader what the error means. Does the method simply fail, or is this evidence for the inconsistency between the different data sets. If so, what is the knowledge gain relative to simply looking at the difference between the three products?	Here, 'error' refers to the variance of the random error component of each measurement system estimated by TC technique. This information is an indicator of the uncertainty in each of the measurement systems with respect to the truth and can't be achieved by looking at the differences between the measurements themselves. We have clarified this in the new version.
P10 L4: Is this really surprising, given that MET and WECANN are both purely statistical models, whereas GLEAM and ECMWF contain process knowledge, which may obscure any direct relationship between fluxes and driving variables?	We believe this similarity between WECANN and FLUXNET-MTE is caused because of their accuracy. FLUXNET-MTE is driven by point based measurements from flux towers unlike GLEAM and ECMWF which are physical models driven by other environmental variables. Therefore, FLUXNET-MTE is expected to be more representative of the true fluxes. On the other hand, WECANN learns from all three products by minimizing the information from the most uncertain product in each case and it is driven by independent satellite based observations. This results in WECANN to have the best performance, and indirectly be more similar to FLUXNET-MTE.
P15 L18. I must have missed this, which figure/statistics demonstrates the added value of SIF at the global scale? Section 4.5 only discusses site-level analyses	This is based on the site level comparisons as we have discussed in Section 4.8.
P15 L 21: This is in parts not surprising, because WECANN isn't fully independent from these data. This needs to be mentioned here.	We tend to disagree with the editor in this case. Because FLUXNET-MTE (which is the other product used in training and our validation comparisons) is a direct upscaling of the flux tower data, while WECANN is indirectly trained on FLUXNET-MTE as well as three other products which are independent of flux tower data. Therefore, we would have expected that FLUXNET-MTE, which is not independent of flux tower data, be the best product. But this is not the case, and WECANN which learns from three independent products during training and uses SIF as an input has a better performance.
Figure 9ff: There seems to be an issue with some of the fluxnet data in the year 2014-2015. Please check and potentially remove flawed data.	There was no panel f in Figure 9 in the previous version of the manuscript. We are not sure what the editor is referring to. But we have investigated flux tower data as much as possible before including them in the analysis. In cases that the reported values seems unrealistic we discussed some of them in the manuscript.
Figure S1: Why is the land-mask shifted relative to the continents?	Thanks for noting this issue. This was caused by an error in using the plot function in MATLAB. It is corrected in the new version.
Tables S1-S4: Why is only R2 averaged over the sites, and not the others? I think it would be helpful to provide the average and standard deviation across sites of these statistics as a table	We had only averaged R2 because it is the index with similar range for all cases, unlike RMSE and bias that their values should be interpreted by considering the mean of the flux at each site.

for the manuscript (while keeping the SI tables), to make this material more accessible to the reader.	However, due to the request of the other reviewer as well, we have now added the average for RMSE. But we tend to not report the average bias, because it can be misleading. For example if a product has large negative and positive biases across different sites a mean value close to zero it does not mean that it has low bias compared to another product that has only small positive biases that on average are larger than the mean of the first product. Moreover, the new Figure 6 summarizes the mean and one standard deviation from these tables based on Plant Functional Types.

Water, Energy, and Carbon with Artificial Neural Networks (WECANN): A statistically-based estimate of global surface turbulent fluxes 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[7,8], Catherine Prigent[1,6], Pierre Gentine[1,2,9]

[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]Department of Earth Sciences, VU University Amsterdam, Amsterdam, 1081HV, The Netherlands
[8]Laboratory of Hydrology and Water Management, Ghent University, Ghent, B-9000, Belgium
[9]Earth Institute, Columbia University, New York, 10027, USA

Correspondence to: Seyed Hamed Alemohammad (sha2128@columbia.edu)

Abstract. A new global estimate of surface turbulent fluxes, including latent heat flux (LE), 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. The approach uses an artificial neural network (ANN) with a Bayesian perspective to learn from three training datasets~~.~~ and is the first global retrieval based on SIF. The combined target input dataset is generated using three independent data sources and a triple collocation (TC) algorithm to define a prior distribution. The new retrieval, named Water, Energy, and Carbon with Artificial Neural Networks (WECANN), provides surface turbulent fluxes from 2007 to 2015 at 1° × 1° spatial resolution and on monthly time resolution. The quality of ANN training is assessed using the target data, and the WECANN retrievals are validated using FLUXNET tower measurements across various climates and conditions. WECANN performs well in most cases and is constrained by the SIF information. When compared to *in situ* eddy covariance observations, WECANN typically outperforms other estimates, particularly for sensible and latent heat fluxes. Uncertainty estimates of the retrievals are analysed and the inter-annual variability in average global and regional fluxes show distinct climatic events such as the impact of El Niño on surface turbulent fluxes.

1 Introduction

Turbulent fluxes from the land surface to the atmosphere, particularly sensible heat flux (H), latent heat flux (LE), and gross primary production (GPP) ~~and net primary production (NPP)~~ 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 surface fluxes from remote sensing observations at 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 surface fluxes from remote sensing information. The first approach uses physically-based or semi-empirical models (e.g. the Priestley-Taylor or Penman-Monteith equations in the case of 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~~, employed by the Max Planck Institute for Biogeochemistry model (MPI-BGC)~~ uses machine learning (e.g. a model tree ensemble) to determine fluxes (LE, H, and GPP) from meteorological drivers and optical remote sensing data~~. Like all supervised machine learning models, the MPI-BGC method~~ (Tramontana et al., 2016). Like all supervised machine learning models, this approach relies on a training dataset to determine the non-linear statistical relationships. In this case, *in situ* turbulent flux measurements 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 fluxes 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 non-linear 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; Rodríguez-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 (Rodríguez-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).

In this study, we develop an ANN approach to retrieve monthly surface fluxes at the global scale. The network uses remotely sensed solar-induced fluorescence (SIF) estimates in addition to other data including precipitation, temperature, soil moisture, snow cover, and net radiation as inputs (predictor). ~~To reduce any biases, we introduce a Bayesian perspective to generate the target dataset for the ANN. Multiple estimates of each of the fluxes are selected according to a prior probability that reflects the quality and information content of the dataset at the particular pixel of interest (details are provided in Section 3.2). This approach enables us, for the first time, to generate a robust target dataset along with a statistical algorithm for the retrieval, while bypassing the need for a land surface model and radiative transfer scheme. This new global product of surface turbulent fluxes is named WECANN (Water, Energy, and Carbon Cycle fluxes with Artificial Neural Networks). WECANN monthly flux estimates for the period 2007 – 2015 are provided on a 1° × 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 Figure S1. It includes all the land areas, except for Greenland, Antarctica, and any 1° × 1° pixel that has more than 75% water, snow or ice permanently. To estimate the fraction of water, snow and ice in each pixel we used the 0.05° × 0.05° MODIS-based Land Cover Type product (MCD12C1 v051) (NASA LP DAAC, 2016).~~

~~A second key innovation of the WECANN methodology is that it uses the new remotely sensed SIF measurement as input.~~ To our knowledge, this is the first ~~time~~study that uses remotely-sensed SIF estimates ~~are used~~ at the global scale to retrieve LE and H surface turbulent fluxes ~~(LE, H, and~~along with GPP~~).~~ .

Previous studies show a strong relationship between the rate of photosynthesis and 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)~~(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 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 carbon (GPP) cycle. Using our machine learning approach we further demonstrate the usefulness of SIF for constraining surface evaporation.

~~The rest of the paper is organized as follows. The datasets used as input and target are introduced in Section 2. The ANN retrieval and Bayesian characterization methods are explained in Section 3. Section 4 provides the results of flux retrievals, validation of results, uncertainty analysis of the retrievals and discussions on the impact of SIF on the retrievals. Conclusions are presented in Section 5.~~

Moreover, to reduce any errors, we introduce a Bayesian perspective to generate the target dataset for the ANN. Multiple estimates of each of the fluxes 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 Section 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 for training of LE and H; and the triplet of MODIS-GPP, ECMWF and FLUXNET-MTE for GPP training.

This new global product of surface turbulent fluxes is named WECANN (Water, Energy, and Carbon Cycle fluxes with Artificial Neural Networks). WECANN monthly flux estimates for the period 2007 – 2015 are provided on a 1° × 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 Figure S1. It includes all the land areas, except for Greenland, Antarctica, and any 1° × 1° pixel that has more than 75% water, snow or ice permanently. To estimate the fraction of water, snow and ice in each pixel we used the 0.05° × 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 Section 2.2: SIF, net radiation, air temperature, soil moisture, precipitation, and snow water equivalent. These are used to retrieve the three surface fluxes (LE, H, and GPP). Different observation and/or model based datasets are used as the training dataset, and are explained in Section 2.1. All the data presented here are projected and gridded on a 1° × 1° geographic grid and averaged at monthly temporal resolution. Finally, the FLUXNET tower data used for validation of the ANN retrievals are presented in Section 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 (Section 3.2). For LE and H, training is performed based on GLEAM, FLUXNET-MTE, and ECWMF ERA HTESSEL. For GPP, training is performed on FLUXNET-MTE, ECWMF 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., 2016; 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 semi-empirical relationship between microwave VOD observations and root-zone soil moisture estimates (based on a running water balance for rainfall and assimilating satellite soil moisture). The data is provided on a 0.25° × 0.25° spatial resolution and daily temporal resolution and starts in 1980. GLEAM data have been used for studying land-atmosphere interactions, and the global water cycle (Guillod et al., 2014, 2015, Miralles et al., 2011a, 2014a, 2014b). In this study, we use LE and H estimates from the latest version v3.0a (Martens et al., 2016).

2.1.2 FLUXNET-MTE

The FLUXNET-MTE (Multi-Tree Ensemble) provides global surface fluxes at 0.5° × 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 learning diverse sequence of different model trees by perturbing the base learning algorithm (Jung et al., 2009, 2010, 2011). The data covers 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 European Centre for Medium-Range Weather Forecasts (ECMWF) Reanalysis (ERA) is a global 3D variational data assimilation (3DVAR) product that uses the Hydrology Tiled ECMWF Scheme for Surface Exchanges over Land (HTESSEL) in the forecast system. HTESSEL has a surface runoff component and accounts for a global non-uniform 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° × 0.25° geographic grid with daily temporal resolution.

2.1.4 MODIS-GPP

The Moderate Resolution Imaging Spectroradiometer (MODIS) sensor is onboard the sun-synchronous NASA satellites Terra (10:30 AM/PM overpasses) and Aqua (1:30 AM/PM 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 surface: atmosphere, land and ocean processes. The MODIS GPP/NPP project (MOD17) provides gross/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 provided on a 0.05° × 0.05° spatial resolution.

2.2 Input Datasets

Six sets of observations are used as input to the WECANN retrieval algorithm. These are selected in a way to provide 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 Meteorological Operational Satellite Program, MetOp-A and MetOp-B satellites, which were launched by the European Space Agency (ESA). GOME-2 was designed to monitor atmospheric ozone profile as wells as other trace gases and water vapor content. It senses Earth backscatter radiance and solar irradiance at a 40×40 km spatial resolution (prior to July 2013 the spatial resolution was 40×80 km). Recently, the retrieval of Solar-Induced chlorophyll Fluorescence (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° × 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 Radiation Energy System (CERES) is a suite of instruments which measure radiometric properties of solar reflected and Earth emitted radiation from the Top Of Atmosphere (TOA) to Earth surface, from three broadband channels at 0.3 – 100 μm. The CERES sensors are on board the Earth Observation Satellites (EOS) including Terra, Aqua and TRMM (Kato et al., 2013; Loeb et al., 2009). We use the net radiation estimates from Synoptic Radiative Fluxes and Clouds (SYN) product of CERES which are provided on a 1° × 1° geographic grid with monthly time resolution.

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°C in the lower troposphere under clear sky condition (Aumann et al., 2003). We use daily temperature estimates from the lowest layer of AIRS level-3 standard product that is provided on a 0.5° × 0.5° geographic grid.

2.2.4 Surface Soil Moisture

The European Space Agency (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 Sensor Microwave/Imager (SSM/I), AMSR-E) and active ones (e.g., the European Remote Sensing (ERS), Advanced Scatterometer (ASCAT)), based on a triple collocation error characterization (Dorigo, et al., in reivew; Liu, Parinussa, et al., 2011; Liu et al., 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° × 0.25° geographic grid.

2.2.5 Precipitation

The Global Precipitation Climatology Project (GPCP) provides global daily precipitation estimates at 1° × 1° spatial resolution from Oct. 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 at 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 × 25 km EASE grid.

2.3 Validation ~~Dataset:~~Datasets

2.3.1 Eddy-Covariance Flux Observations

Formatted: Heading 3

FLUXNET is a network of regional tower sites, which measure turbulent flux exchanges (water vapor, energy fluxes and carbon dioxide) between ecosystems and atmosphere (Baldocchi et al., 2001). FLUXNET comprises over 750 sites covering five continents. Measurements from the FLUXNET towers provide valuable information for validating 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 validation (details are provided in section 4.4).

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 Large-scale Biosphere-Atmosphere (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. We note that, of course, the inter-annual 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-fair use version of the dataset.

A total of 97 sites from the three datasets are selected for validation of WECANN retrievals spanning a large climatic and biome gradient (Fig. S2). For AmeriFlux towers, if measurements from both the FLUXNET 2015 dataset and the La Thuile dataset were available, we have used the FLUXNET 2015 data. We have only selected sites that had at least 24 month of continuous measurements during 2007-2015 years. Any site that would have fallen outside of the WECANN land mask (Fig. S1) is excluded (several sites in coastal regions).

2.3.2 GRACE Terrestrial Water Storage Changes

Gravity Recovery And Climate Experiment (GRACE) is a suit of two satellites orbiting the Earth and measuring changes in the gravitational field of the Earth. Observed changes are then related to changes in terrestrial water storage which includes groundwater, surface water, snow and soil moisture (Landerer and Swenson, 2012; Swenson and Wahr, 2006). Here, we use the retrievals from JPL's mascon solution release 5 which are posted on a 1° × 1° spatial grid at monthly time scale. These retrievals are used in section 4.6 to evaluate WECANN ET estimates from 2007 to 2015.

3. Methodology

3.1 Artificial Neural Network Setup

We developed an ANN retrieval algorithm to estimate the surface fluxes (LE, H, and GPP) based on our six sets of input observations: SIF, net radiation, air temperature, soil moisture, precipitation, and SWE (as described in Section 2.2). The ANN used here is a feedforward network consisting of 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 3 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), so 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 training, the target data is 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 are 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 1 to 5 neurons, the R^2 value between the target data and the ANN estimates increased with an increasing number of neurons. For more than 5 neurons, little change in the skill was observed when increasing the number of hidden layer neurons (Fig. S3). Thus an ANN with 5 hidden layer neurons represents the simplest ANN that can converge to a solution and model the non-linear 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 triple collocation based merging of triplets of the flux estimates introduced in Section 2.1 (details are discussed in Section 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 validated against other global products and eddy covariance tower data. Results of these comparisons are presented in section 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; Rodríguez-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 triple collocation (TC) technique. TC is a method to estimate the Root Mean Square Errors (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 ~~(Alemohammad 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)~~(Alemohammad 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:

Field Code Changed

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

where $X_i's$ are the measurements from the collocated system i (e.g. remote sensing observation, model output, etc~~)~~.), t is the true value, 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 variables in each measurement. By further assuming that the errors from the three measurements are uncorrelated $\left(Cov(\varepsilon_i, \varepsilon_j) = 0, \text{ for } i \neq j\right)$ and the errors are uncorrelated with the truth $(Cov(\varepsilon_i, t) = 0)$, the RMSE of each measurement error can be calculated as (McColl et al., 2014):

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

in which Q_{ij} is the (i^{th}, j^{th}) 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 semi-empirical 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 random error component of Equation (1) using TC for each of the surface fluxes, and use them as TC based errors of each product.

The TC estimated errors ~~from~~for the surface fluxes are shown in Figs. S4-S6. The white regions represent missing retrievals or discarded negative estimates due to insufficient data record. For LE, high TC errors are found in the Amazon rainforest and tropical

Africa for GLEAM, in Amazon rainforest and the Sahel for ECMWF, in Indian peninsula for FLUXNET-MTE and in U.S. Great Plains for ECMWF and FLUXNET-MTE. For H, beside 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 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 which 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 and Variability of Surface Fluxes

In this section, we present and compare the retrievals of LE, H and GPP fluxes 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 global average annual retrieved fluxes and scatterplots of flux retrievals vs target estimates. The spatial patterns of the WECANN retrievals are similar to expectations. The average global fluxes in 2011 are 38.33 W m^{-2} for LE, 39.44 W m^{-2} for H, and 2.34 gC m^{-2} day^{-1} (or 123.16 PgC yr^{-1}) for GPP. LE has the best R^2 (0.95) comparing to the other two flux variables 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 W m^{-2}; for H, RMSD = 13.13 W m^{-2}; and for GPP, RMSD=1.22 gC m^{-2} day^{-1}.

The seasonal variability and spatial pattern of the surface flux retrievals from 2011 (LE, H, GPP) 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 nutrient, soil type conditions as well as 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 are noticeable in wet mid-latitude regions for both Northern and Southern Hemisphere, i.e., East Asia, Eastern U.S. and Australian North and East Coast with over 60 W m^{-2} difference between winter and 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 re-evaporation 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, can be learnt by the network.

Seasonal variabilities in H (Fig. 4) are distributed in opposite pattern to LE, as expected. Deserts and dry regions i.e., the Sahara, Southwestern U.S. 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, Indonesia) exhibit limited seasonal variability. In mid-latitude energy-limited regions (Central/Eastern Europe, Easter US), H also reflects the course of available energy, and in more water-limited regimes (e.g. 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 of 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 continental U.S. In 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 Santarem (Restrepo-Coupe et al., 2013; da Rocha et al., 2009). Compared to LE, substantial geographical variability is observed 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 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, South-East Asia, Northern Australia and Central-Northern America are well captured with rapid rise in GPP following water availability. The highest GPP are observed in rainforests and the US agricultural Great Plains, in JJA for the latter. Northern latitude regions mainly exhibit substantial GPP in the summer and late spring, and small values throughout the rest of the year.

~~4.2~~4.2 Validation with FLUXNET Data

Direct validation of the WECANN fluxes is challenging by the fact that no global, error-free flux estimates 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, triple collocation provides a valuable technique to validate 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 fluxes and any of the datasets used for the training. Instead, we validate the fluxes by comparing them to data from a set of FLUXNET eddy-covariance towers. In situ estimates from eddy covariance towers with a footprint of a few 100 m to km may not be representative of the entire 1° × 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 at different spatial scales. For instance, the phenology has a strong impact on the seasonal cycle of the fluxes and in the following examples, it is clearly highlighted when comparing the different products to flux tower estimates.

Summary of statistics across 97 FLUXNET sites are provided in Tables S1 – S3. Overall, WECANN performs better than other alternative global products. In particular, WECANN has the highest correlation for 76% of sites for LE, 54% of sites for H, and 53% 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 S1 - S3 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 fluxes.

Figure 7 shows the comparison of monthly WECANN retrievals and three other global products' estimates with the tower estimates across 5 select sites that span a range of climatic and vegetation coverage conditions. At the Oklahoma agricultural site (US-ARM), H and LE are well reproduced, yet dry year H is underestimated (Fig. Figure 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 be again due to the difference in the footprint of the two estimates.

At the Brasschaat site, BE-Bra, Belgium (Fig. Figure 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 too early GPP rise and overestimate summer GPP. Again, the SIF data provides 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 very well captures the seasonal cycle of GPP and LE, as well as to a less extent H. WECANN better reproduces the seasonality, amplitude and interannual variability compared to other retrievals (Fig. Figure 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. Figure 7d). Yet, WECANN slightly underestimates LE and overestimates GPP. In fact, all flux retrievals overestimate GPP in the dry and cold seasons. The landscape in the region is highly heterogeneous with denser vegetation in 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. Figure 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 better captures the seasonal dynamics compared to other products.

Overall, across the different sites, the WECANN retrieval performs better than other products, especially in terms of the seasonality of the fluxes. 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 flux estimates. 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.

Formatted: Tab stops: Not at 1.63"

4.3 Comparison against other remote-sensing based products

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 scatterplot 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 W m^{-2} 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 is 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 FLUXNEWT-MTE.

The differences in H estimates are more complex (Fig. 9). First, the R^2 between WECANN and the other datasets are slightly lower than for LE. ECMWF and FLUXNET-MTE yield higher R^2 with WECANN (0.92) while GLEAM has an R^2 of 0.87. GLEAM exhibits lower H in most of the Northern hemisphere, 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 Western US, Brazilian Savannas, Southern Congo, the Sahel compared to WECANN and smaller values in the Amazon, Indonesia, and over desert areas of the Sahara and Arabic peninsula as well as South East Asia. The GLEAM and ECMWF H difference maps show many similar patterns: the Sahara, Eastern Europe, East Asia are underestimated, while Southern Africa and Eastern part of Amazon are overestimated. Similarly the errors patterns estimated from TC (Fig. S5) are consistent with the comparison of WECANN and target datasets. Figure S5 shows that ECMWF has higher errors in the Sahel, Southern Congo, and Brazilian Savana 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 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 higher GPP than WECANN. In Congo, MODIS and FLUXNET-MTE have higher GPP, while ECMWF has a lower one. In Central and 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.

~~4.3 Validation with FLUXNET Data~~

~~Direct validation of the WECANN fluxes is made more~~4.4 Extreme Events Assessment

In order to further assess WECANN at the global scale, we analyse its capacity in capturing extreme events. We thus selected three major heatwave and drought events that occurred during the temporal coverage of WECANN product. These events are the Russia 2010 heat wave, Texas 2011 drought and US Corn Belt 2012 drought. Figure 11 shows the percentage of mean monthly anomalies with respect to mean surface fluxes, 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 heatwave occurred over western Russia and resulted in 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 10-15% increase in H and 15-20% reduction in LE. In early 2011, drought conditions developed in southern US, particularly in the states of Texas and Louisiana (Luo and Zhang, 2012). By April, most of Texas, Oklahoma, Louisiana and Arkansas were classified in the D4 drought condition (exceptional drought), and the situation continued throughout the summer and fall of 2011 as reported by US Drought Monitor (Svoboda et al., 2002). As Figure 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 US Corn Belt was categorized as either D3 (extreme drought) or D4 (exceptional drought) condition. Figure 11 shows similar patterns in surface fluxes with significant positive anomaly in H (~20%) and reductions in LE and GPP (~20%), consistent with crop yield decrease.

4.5 Basin Scale Evaluation

We also assess the accuracy of WECANN ET retrievals using estimates of an independent water budget closure model across five major basins (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 at basin scale. These basins include 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 a nutshell 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. The analysis is carried out for years 2007 to 2010. Figure 12 shows the relative difference in ET estimates from WECANN compared to the ET estimates from the water budget study (mean value and one standard deviation across 48 months). Mean differences vary between a low of 5% in the Amazon and a larger 24% value in Colorado, while other three basins have a mean difference 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 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 heterogeneous 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 compare less favorably with these large-scale estimates. Downscaled version of those estimates would further help in the evaluation of ET products.

4.6 Comparison with GRACE

Terrestrial water storage changes estimated from GRACE satellite observations provide a valuable opportunity to evaluate global estimates of surface fluxes in combination with other observations. Terrestrial water storage includes groundwater, soil moisture, surface water and snow cover. Indeed, changes in storage equals precipitation (P) minus ET and runoff. We thus expect positive correlation between changes in storage and P-ET, except in snowy regions where delays in snow melt correlates storage compared to P-ET.

We estimate the correlation between seasonal changes in terrestrial water storage and the difference between precipitation and WECANN ET. This analysis provides another independent evaluation of ET estimates from WECANN over larger scales than over flux towers. We use the GPCP precipitation estimates, and calculate the correlation coefficient at seasonal time scale between 2007 and 2015 (Figure 13). Over almost all of the non-snowy regions the correlation is positive, as expected, meaning that the positive (negative) value of P-ET results in a positive (negative) change in terrestrial water storage and confirming that at a larger scale the ET seasonal cycle is consistent with changes in storage. Over snow dominated regions, such as the Rockies or Russia the correlation is negative which is due to the role of snow cover in the changes of the terrestrial water storage. This evaluation provides another perspective to the quality of the ET estimates from WECANN.

4.7 ~~challenging by the fact that no global, error-free flux estimates are available. Remote sensing or model products such as those used for training have their own errors. In situ estimates from eddy covariance towers with a footprint of a few 100 m may not be representative of the entire 1° × 1° pixel, and are known to have problems with energy closure. 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, triple collocation provides a valuable technique to validate large-scale datasets in the absence of a known truth. However, WECANN's use of different noisy training datasets may cause the presence of some correlated errors between WECANN fluxes and other possible large-scale triple collocation inputs. Instead, we validate the fluxes by comparing them to data from a set of FLUXNET eddy-covariance towers. Nevertheless, it is important to keep in mind that these flux estimates may themselves have errors relative to the true 1-degree scale fluxes and their footprint not be representative of the WECANN 1° × 1° pixels. However, in the comparison against tower data the impact of large-scale climate variability such as the seasonal cycle or interannual variability are comparable to pixel based retrievals. For instance, the phenology has a strong impact on the seasonal cycle of the fluxes and in the following examples, it is clearly highlighted when comparing the different products to flux tower estimates.~~

~~Summary of statistics across the 97 sites are provided in Tables S1 – S3. Overall, WECANN performs better than the alternative global products. In particular, WECANN has the highest correlation for 76% of sites for LE, 54% of sites for H, and 53% of sites for GPP. This high R^2 reflects the capacity of WECANN to correctly capture the seasonal cycle and interannual variability. 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, contrary to optical vegetation indices that are sensitive to vegetation greenness and canopy cover - factors which can lag fluxes or be out of phase (see e.g. the lower correlation with NDVI in 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. In the following, we analyze the retrievals across 17 select sites that span a range of climatic and vegetation coverage conditions. We provide interpretations of similarities and differences between the retrievals, flux tower measurements as well the three training datasets.~~

~~Figure 9 shows the comparison of monthly WECANN retrievals with the tower estimates across 5 European sites. At the AT-Neu site, Neusflit, Stubai Valley, Austria (Fig. 9a), the seasonal cycle is correctly captured for both LE and GPP. All flux retrievals perform relatively well at this site dominated by radiation and temperature. The GPP based on the eddy covariance has a sharper and earlier rise in the spring than LE, which seems unrealistic and may be an artifact of the GPP retrieval method. WECANN is~~

~~slightly delayed compared to the observed LE, possibly a reflection of the larger footprint encapsulating various conditions in this steep topography region. All flux retrievals overestimate the H observations, even though they capture some of the seasonality. The observed H lags the observed LE, which seems unrealistic given that the region is mostly radiation limited so that a spring increase in radiation and temperature should affect both fluxes. The large footprint of the retrieval could be another source of error, as it would sample multiple environmental conditions. Nonetheless, the ECMWF and GLEAM retrievals are the closest to the observed H and FLUXNET-MTE strongly overestimates the observed flux, similarly to WECANN, even though the bias is not as high.~~

~~At the Brasschaat site, BE-Bra, Belgium (Fig. 9b), 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. 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. WECANN captures the GPP seasonal cycle compared to other products, which display too early GPP rise and overestimate the summer GPP. Again, the SIF data provides independent useful data compared to other environmental information (radiation, temperature, vegetation indices) used by the other retrieval schemes.~~

~~At another seasonally cold site, in Switzerland, CH-Fru (Fig. 9c), WECANN again performs very well, correctly reproducing the seasonality of all fluxes, especially compared to the other products, which tend to rise too early in the spring. The magnitude of H and LE is very similar to the observations, yet GPP seems to be overestimated by WECANN, yet much less so than other products.~~

~~At the Mediterranean, Spanish site, ES-LgS (Fig. 9d) WECANN correctly reproduces H and LE yet overestimates the magnitude of GPP, even though it correctly captures its seasonal dynamics. We note; however, that the region is highly heterogeneous both in terms of topography and vegetation coverage and that the site is located at some of the driest location of the region.~~

~~At the cold Finland site (FI-Hyy), WECANN very well captures the seasonal cycle of GPP and LE, as well as to a less extent of H. WECANN better reproduces the seasonality, amplitude and interannual variability compared to other retrievals (Fig. 9e).~~

~~At the Brazilian sites, spanning the Savanna region to the Amazonian rainforest (Fig. 10), we only consider the climatology of the results, as most the data (ending in 2006) was not available during the GOME-2 satellite period. We acknowledge potential differences when considering the climatology of the fluxes, as interannual variability could modify the derived climatological seasonality. At the Rondônia sites Ji1, all flux retrievals tend to overestimate LE and GPP. This is most likely a reflection of the large landscape fragmentation with deforested and non-deforested patches. Similarly, the dryness perceived at the flux tower is not seen by most of the retrievals as forests can sustain photosynthesis during the dry season through deeper roots (da Rocha et al., 2009). At the nearby Ji2 site, on the other hand, most flux retrievals perform much better and correspondingly report a maintained GPP and LE in the dry season. GLEAM as well as ECMWF exaggerate the seasonal cycle of LE and H. WECANN is positively biased in H but correctly reproduces LE. FLUXNET-MTE better reproduces GPP than WECANN and both products outperform MODIS and the ECMWF retrievals. At the other site near Sao Paulo, with dry winter savanna, most flux retrievals correctly capture the seasonal cycle, yet most retrievals and especially WECANN are in seasonal advance over the observed eddy covariance with a too early increase in GPP and LE. The site is located in a highly heterogeneous agricultural landscape yet observes an evergreen broadleaf forest, which is not representative of the heterogeneous landscape seen by the remote sensing products.~~

~~In Canada, (Fig. 11), WECANN very well reproduces the seasonal cycle of LE, especially compared to the other products that produce a too early rise in LE during the spring season. WECANN also better reproduces the seasonal cycle of GPP compared to other products. Nonetheless, all GPP retrievals underestimate the reported eddy covariance GPP. This is true of both sites Qfo and~~

~~Qcu. The reported eddy-covariance GPP appears very small though, especially given the LE magnitude in the summer, pointing to potential problem in the magnitude of the surface fluxes, which is drastically impacted by the high-frequency corrections of the turbulent co-spectrum and its parameterization (Mamadou et al., 2016). H is well reproduced by WECANN at the Qcu site, but the Qfo site exhibits nearly twice the H magnitude of the Qcu site in the summer. This does not appear realistic given that the radiative and LE conditions are relatively similar at the two sites. WECANN again better reproduces the seasonal cycle compared to the other products.~~

~~Across the continental US Ameriflux sites (Fig. 12), WECANN performs well in terms of seasonal and interannual dynamics. At the Oklahoma agricultural site (US-ARM), H and LE are well reproduced, yet dry year H is underestimated (Fig. 12a). 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 be again due to the difference in the footprint of the two estimates, At the Illinois site, US-Ib2, the dynamics of LE is relatively well reproduced by most products except for ECMWF (Fig. 12b). All retrievals overestimate GPP, especially FLUXNET-MTE. WECANN exhibits a late delay in the GPP decay. The measured H is very noisy yet exhibits a summer decay which is only partially captured by the different products. At the evergreen needleleaf Maine site, US-Me2, WECANN reproduces the dynamics of H, LE and GPP well, even if it underestimates the peak fluxes (Fig. 12c). Over the irrigated maize site in Nebraska (US-Ne1), the retrievals underestimate the peak LE and GPP, as well as overestimate the H in the peak summer season (Fig. 12d). This is most likely a reflection of the larger area observed or modeled by the flux retrievals which do not include similar intensive irrigation practices, leading to lower peak LE (and correspondingly higher H) and GPP. Only FLUXNET-MTE reproduces the magnitude of this irrigated site (but US-Ne1 was included in the FLUXNET-MTE training database). Finally, at the monsoonal grassland site of Santa Rita, AZ, WECANN correctly captures the complex dynamics of H and LE at the site with sometimes rain periods preceding the Monsoon period (Fig. 12e). Yet, WECANN slightly underestimates LE and overestimates GPP. In fact, most flux retrieval overestimate GPP in the dry and cold seasons. The landscape in the region is highly heterogeneous with denser vegetation in 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.~~

Formatted: Tab stops: Not at 1.63"

~~Figure 13 shows the comparison of retrievals at two other sites. At the Daly River pasture, AU-DaP, Australia (Fig. 13a), WECANN reproduces very well the observed LE in terms of both seasonal and interannual variability. Compared to other products, WECANN better reproduces the seasonal cycle of this Monsoonal site, with a rapid rise in LE and lagged drying. Most retrievals fail to correctly reproduce the exact H seasonality, which is in opposite phase with LE, at this water limited site. All retrievals tend to overestimate the retrieved eddy-covariance GPP and fail to correctly capture the rapid rise in GPP, except for WECANN. The eddy-covariance GPP decay occurs significantly in advance over the LE decay. It seems unlikely that during the drying phase soil evaporation would explain nearly all of the LE and that transpiration would be so small (as indicated by the drop in GPP before LE). It is most likely due to an artifact in the model fitting of the respiration component, which implicitly assumes some stationarity. Nonetheless, all remote sensing retrievals seem to overestimate the dry season GPP.~~

~~At the South African Mediterranean site, ZA-Kru, WECANN reproduces some of the dynamics of the observed H, yet is typically smoother (Fig. 13b). Similarly, 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 better captures the seasonal dynamics.~~

~~Overall, across the different sites, the WECANN retrieval performs better than other products, especially in terms of the seasonality of the fluxes. Several factors contribute to the capability of WECANN in having a better retrieval compared to other products. The ANN approach in WECANN uses a novel training technique to remove highly uncertain and outlier estimates from its target~~

~~dataset. Therefore, WECANN retrievals are closer to the truth than each of the single target datasets. Moreover, the SIF measurements that are directly correlated with GPP provide a better constraint on flux estimates.~~

~~4.~~4 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 mean 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 are spatially and temporally varying and we used the associated value for each time and space data point. For net radiation, we use a constant standard deviation of 34.58 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 14 shows the results of the bootstrap for each of the three fluxes globally and in different climatic zones. The zones are defined as Polar (90° N - 60° N), Northern Hemisphere (NH) mid-latitude (60° N - 10° N), Tropics (10° N - 15° S), and Southern Hemisphere (SH) mid-latitude (15° S - 60° S). Each panel in Figure 14 shows the uncertainty derived from the bootstrap experiment, relative to the interannual variability of the fluxes. GPP estimates are provided in units of PgC yr^{-1} as total productivity in each region. LE and H are provided in units of W m^{-2} as an average rate of flux in each region.

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

The inter-annual variations of surface fluxes show distinct patterns. For example, in year 2015, which was an El Niño year, LE and GPP have reduced notably, and H increased to an extreme value in the 9 years of WECANN product. Moreover, from 2011 to 2015 both LE and GPP have a consistent decreasing trend at global scale. The inter-annual variability of GPP and LE are similar at global scale while their regional patterns are different. For example, in year 2015 GPP at global scale and in all regions has decreased with respect to 2014, while LE in Polar and NH mid-latitudes have increased and LE at global scale has decreased. As expected, the variability of LE and H are anti-correlated.

4.~~5~~8 Impact of SIF on the retrieval of surface fluxes

Satellite SIF observations are relatively new, and have not been used to estimate LE and H at the global scale previously. Therefore, we want to assess the information content of SIF observations in the WECANN retrievals by replacing them with more typical optical/near-infrared indices of vegetation (NDVI or EVI).

To do so, we trained two different ANNs with NDVI and EVI instead of SIF data on each of the three fluxes and evaluated the retrievals against the same FLUXNET tower measurements used in Section 4.~~3~~2 for validating WECANN retrievals. Tables S4 - S6 show the results of validations of these three retrievals against the tower measurements for LE, H and GPP, respectively. In terms of correlation coefficient, on average all three retrievals have relatively similar performance except in regions where phenology (and incident radiation) is not the main contributor to the flux variability such as in Spain (ES-LgS). Indeed, in such regions changes in canopy structure is 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, similarly 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)~~(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 such as those from WECANN.

5 Conclusion

This study introduces a new statistical approach to retrieve global surface latent and sensible heat fluxes as well as gross primary productivity using remotely sensed observations at a monthly time scale. The methodology is developed based on an Artificial Neural Network (ANN) that uses six input datasets including solar induced fluorescence (SIF), 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 validated using target datasets as well as FLUXNET tower observations. The validation results comparing 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 GPP (RMSD=0.88 gC m^{-2} day^{-1}), which is believed to be one of the most realistic global datasets and it has the lowest RMSE based on our TC error estimates (Fig. S4 – Fig. S6), despite its reported underestimated inter-annual variability due to the use of climatological values for several meteorological drivers (Miralles et al., 2014a, 2016). Such tendency also can 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 flux retrieval maps indicate that all three fluxes have similar seasonal variability and distribution which are determined by annual phenological cycle in energy limited Northern latitude regions, dryness in Mediterranean and Monsoonal climates and by light availability in rainforests. Seasonal radiation has great impact on some regions for all flux variables, such as Eastern U.S., Europe and East Asia, which have wet conditions, are highly vegetated and located in mid-latitudes. As opposed to this, the seasonal variability for all fluxes 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 Southwest U.S., Western Australia, North Africa and Western Asia are not significant, as there is less seasonal solar radiation variability in aforementioned regions. Comparison between the flux variables LE, H, and GPP, 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 contributor to flux variability, while it has similar performance in other regions.

~~Finally, from~~From the validation results comparing 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 in 76% of site for LE, 54% of sites for H, and 53% of sites for GPP compared to other products. Moreover, retrievals from WECANN outperform other global products in capturing the seasonality of surface fluxes across a wide range of sites with different climatic and biome conditions.

We also assessed the performance of WECANN in capturing extreme heatwave and drought events, and showed that in the case of Russia 2010 heatwave, Texas 2011 drought and US Corn Belt 2012 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. Finally, a correlation analysis of P-ET (from WECANN) with water storage changes from GRACE mission showed positive correlation coefficients across non-snowy regions, confirming the accuracy of ET estimates from WECANN.

Data Availability

WECANN product is publicly available for download on Aura Validation Data Center (AVDC) at Goddard Space Flight Center via https://avdc.gsfc.nasa.gov/pub/data/project/WECANN/

Competing Interests

The authors declare that they have no conflict of interest.

Acknowledgments

The funding for this study is provided by the NASA grant # NNX15AB30G. PG acknowledges funding from NSF CAREER Award # EAR - 1552304, and NASA grant # 14-AIST14-0096. DM and PG acknowledge funding from the Belgian Science Policy Office (BELSPO) in the frame of the STEREO III programme project STR3S (SR/02/329). WECANN product is hosted on AVDC server, and we would like to thank Michael M. Yan and Ghassan Taha for their help in this regard. The authors would like to thank all the producers and distributors of the data used in this study. The ECMWF team (Dr. Gianpaolo Balsamo and Dr. Souhail Bousetta, in particular) for providing the ECMWF data. We also thank NASA and Prof. Running for providing the MODIS GPP estimates and Dr. Johanna Joiner for the GOME-2 data. The GPCP 1DD data were provided by the NASA/Goddard Space Flight Center's Mesoscale Atmospheric Processes Laboratory, which develops and computes the 1DD as a contribution to the GEWEX Global Precipitation Climatology Project. The MCD12C1 data product was retrieved from the online Data Pool, courtesy of the NASA Land Processes Distributed Active Archive Center (LP DAAC), USGS/Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota, https://lpdaac.usgs.gov/data_access/data_pool. This work used eddy covariance data acquired and shared by the FLUXNET community, including these networks: AmeriFlux (U.S. Department of Energy, Biological and Environmental Research, Terrestrial Carbon Program (DE-FG02-04ER63917 and DE-FG02-04ER63911)), AfriFlux, AsiaFlux, CarboAfrica, CarboEuropeIP, CarboItaly, CarboMont, ChinaFlux, Fluxnet-Canada (supported by CFCAS, NSERC, BIOCAP, Environment Canada, and NRCan), GreenGrass, ICOS, KoFlux, LBA, NECC, OzFlux-TERN, TCOS-Siberia, and USCCC. The FLUXNET eddy covariance data processing and harmonization was carried out by the ICOS Ecosystem Thematic Center, AmeriFlux Management Project and Fluxdata project of FLUXNET, with the support of CDIAC, and the OzFlux, ChinaFlux and AsiaFlux offices. We acknowledge the financial support to the eddy covariance data harmonization provided by CarboEuropeIP, FAO-GTOS-TCO, iLEAPS, Max Planck Institute for Biogeochemistry, National Science Foundation, University of Tuscia, Université Laval and Environment Canada and US Department of Energy and the database development and technical support from Berkeley Water Center, Lawrence Berkeley National Laboratory, Microsoft Research eScience, Oak Ridge National Laboratory, University of California - Berkeley, University of Virginia. GRACE data are retrieved from GRCTellus Land data on JPL's website: https://grace.jpl.nasa.gov/.

References

Aires, F.: Combining Datasets of Satellite-Retrieved Products. Part I: Methodology and Water Budget Closure, J. Hydrometeorol., 15(4), 1677–1691, doi: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(D11), D11103, doi: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(D18), doi: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(8), 3489–3503, doi:10.5194/hess-19-3489-2015, 2015.

Anber, U., Gentine, P., Wang, S. and Sobel, A. H.: Fog and rain in the Amazon, Proc. Natl. Acad. Sci., 112(37), 11473–11477, doi: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(2), 253–264, doi: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(11), 2415–2434, doi: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(3), 623–643, doi: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 (80-.)., 329(5993), 834–838, doi: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, Glob. Food Sec., 2(3), 139–143, doi:10.1016/j.gfs.2013.08.002, 2013.

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~~,~~(0), 141–154, doi:10.1016/j.rse.2013.12.005, 2014.

Damour, G., Simonneau, T., Cochard, H. and Urban, L.: An overview of models of stomatal conductance at the leaf level, Plant. Cell Environ., 33(9), no-no, doi:10.1111/j.1365-3040.2010.02181.x, 2010.

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(9), 935–940, doi: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(11), 1383–1398, doi:10.1046/j.1365-3040.2002.00909.x, 2002.

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(6), n/a-n/a, doi: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., n.d.

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, doi: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(3), 901–919, doi: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 variations of net CO2 assimilation and stomatal conductance during water-stress in C3 plants, Physiol. Plant., 114(2), 231–240, doi: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(1–2), 149–160, doi: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(17), doi: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(8), 2081–2094, doi: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, doi: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(5), 665–681, doi: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(3), 191–194, doi: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, doi: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, doi: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, Proc. Natl. Acad. Sci., 111(14), E1327–E1333, doi: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(16), 8343–8367, doi: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, doi: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(D15), D15107, doi:10.1029/2011JD015633, 2011.

Havaux, M. and Lannoye, R.: Chlorophyll fluorescence induction: A sensitive indicator of water stress in maize plants, Irrig. Sci., 4(2), 147–151, doi: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(1), 36–50, doi: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(7445), 347–350, doi: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, doi: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, doi: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(D6), D06305, doi: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(D2), D02102, doi: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(12), 6771–6782, doi: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(10), 2803–2823, doi: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, doi: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(8), 3939–3967, doi: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(12), doi: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(10), 2001–2013, doi: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(7318), 951–954, doi: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(G3), G00J07, doi: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(9), 2719–2740, doi: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(10), 4847–4859, doi: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, doi:10.1016/j.rse.2015.11.011, 2016.

van Kooten, O. and Snel, J. F. H.: The use of chlorophyll fluorescence nomenclature in plant stress physiology, Photosynth. Res., 25(3), 147–150, doi:10.1007/BF00033156, 1990.

Krause, G. H. and Weis, E.: Chlorophyll Fluorescence and Photosynthesis: The Basics, Annu. Rev. Plant Physiol. Plant Mol. Biol., 42(1), 313–349, doi: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(4), doi: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(1), 392–403, doi: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, Proc. R. Soc. B Biol. Sci., 280(1761), 20130171–20130171, doi: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. Chang. Biol., 21(9), 3469–3477, doi: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(10), 13448–13465, doi: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(2), 425–436, doi: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, doi: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(3), 748–766, doi:10.1175/2008JCLI2637.1, 2009.

Luo, L. and Zhang, Y.: Did we see the 2011 summer heat wave coming?, Geophys. Res. Lett., 39(9), n/a-n/a, doi: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.

~~Mamadou, O., Gourlez de la Motte, L., De Ligne, A., Heinesch, B. and Aubinet, M.: Sensitivity of the annual net ecosystem exchange to the cospectral model used for high frequency loss corrections at a grazed grassland site, Agric. For. Meteorol., 228–229, 360–369, doi:10.1016/j.agrformet.2016.06.008, 2016.~~

Martens, B., Miralles, D. G., Lievens, H., van der Schalie, R., de Jeu, R. A. M., Férnandez-Prieto, D., Beck, H. E., Dorigo, W. A. and Verhoest, N. E. C.: GLEAM v3: satellite-based land evaporation and root-zone soil moisture, Geosci. Model Dev. Discuss., 0, 1–36, doi:10.5194/gmd-2016-162, 2016.

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(17), 2014GL061322, doi: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, doi: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(19), 3287, doi: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(4), 441–459, doi: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(2), 803–822, doi: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(6), 1423–1429, doi: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(2), 453–469, doi:10.5194/hess-15-453-2011, 2011a.

Miralles, D. G., De Jeu, R. A. M., Gash, J. H., Holmes, T. R. H. and Dolman, A. J.: Magnitude and variability of land evaporation and its components at the global scale, Hydrol. Earth Syst. Sci., 15(3), 967–981, doi: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, Nat. Clim. Chang., 4(2), 122–126, doi: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(5), 345–349, doi: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(2), 823–842, doi: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(980), 277–294, doi: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(7487), 221–224, doi: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(7), 2195–2206, doi: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(4), 519–536, doi: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(6), n/a-n/a, doi: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 revisio, 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(21), 12,100-12,116, doi: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(4), 4899–4918, doi: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(4), 779–783, doi: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, pp. 271–278, IEEE., 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(2), 257–269, doi: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, Agric. For. Meteorol., 182–183, 128–144, doi:10.1016/j.agrformet.2013.04.031, 2013.

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(G1), G00B12, doi:10.1029/2007JG000640, 2009.

Rodríguez-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(11), 5991–6007, doi: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. Zeitschrift, 11(5), 367–370, doi: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(6), 547, doi: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(7594), E4–E5, doi: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. Chang. Biol., 21(5), 1762–1776, doi: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, doi: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(8), 1181–1190, doi: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(8), L08402, doi: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(3), 744–748, doi:10.1104/pp.86.3.744, 1988.

van der Tol, C., Verhoef, W. and Rosema, A.: A model for chlorophyll fluorescence and photosynthesis at leaf scale, Agric. For. Meteorol., 149(1), 96–105, doi:10.1016/j.agrformet.2008.07.007, 2009.

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(14), 4291–4313, doi:10.5194/bg-13-4291-2016, 2016.

USDA: Crop Production 2012 Summary., 2013.

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, doi: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. Chang. Biol., 22, 2979–2996, doi: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(5), 853–868, doi: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(8), 84005, doi: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, doi: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(August 2015), 19124, doi:10.1038/srep19124, 2016b.

Zhao, M. and Running, S. W.: Drought-Induced Reduction in Global Terrestrial Net Primary Production from 2000 Through 2009, Science (80-.)., 329(5994), 940–943, doi: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(2), 164–176, doi:10.1016/j.rse.2004.12.011, 2005.

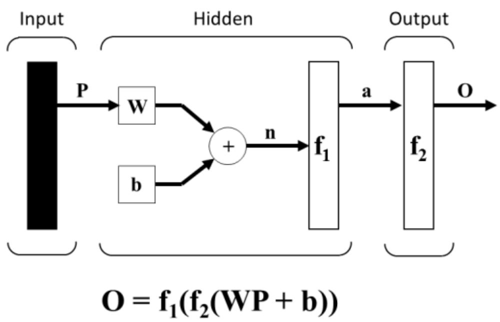


Figure 1: Architecture of the ANN layers. Input layer provides the matrix P of the inputs to the Hidden layer. Hidden layer has a matrix W of weights and b of biases for the neurons, and the f_1 transfer function. The output of the Hidden layer ($a = f_1(WP + b)$) is an input to the Output layer that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

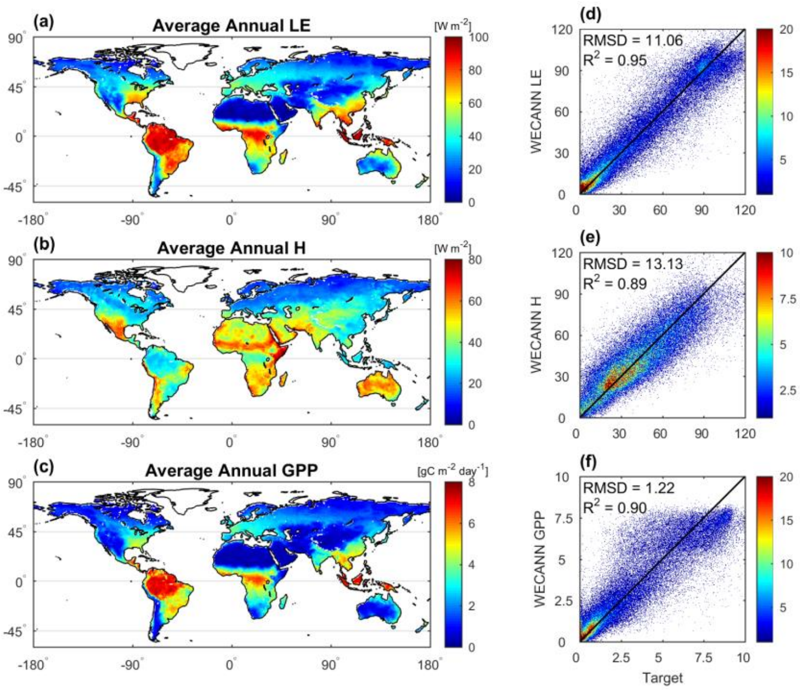


Figure 2: Left column: Annual average surface fluxes in 2011 for (a) LE, (b) H, and (c) GPP. Right column: Density scatterplot between estimates of ANN and target data for (d) LE, (e) H, and (f) GPP during the validation period (2011). The density of scatter points is represented by the shading color. The diagonal black line depicts the 1:1 relationship.

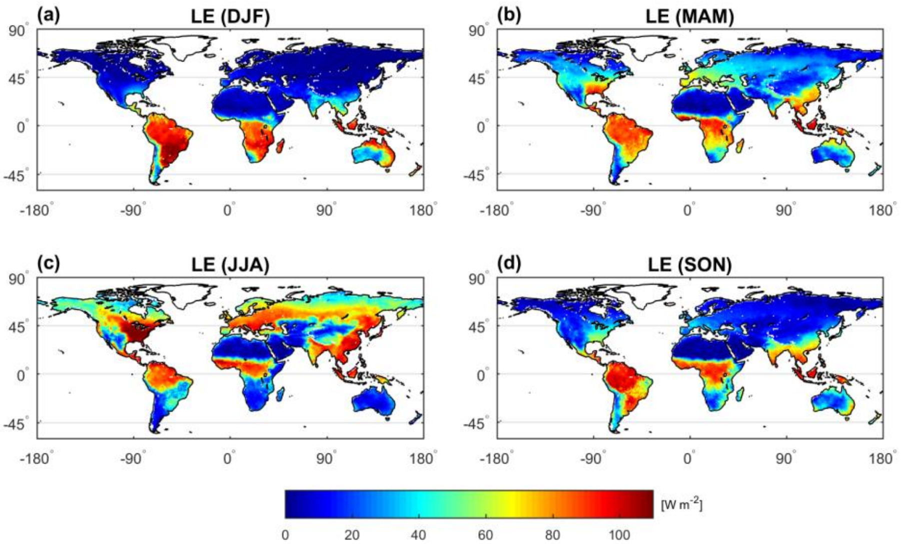


Figure 3: Global patterns of seasonal average LE from WECANN in 2011, (a) December - February, (b) March - May, (c) June - August, and (d) September - November.

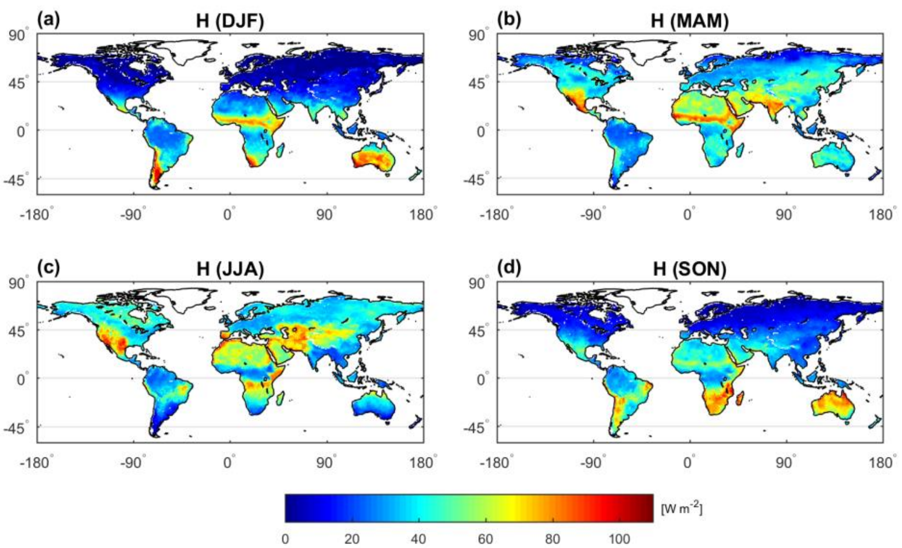


Figure 4: Similar to Figure 3 but for H instead of LE

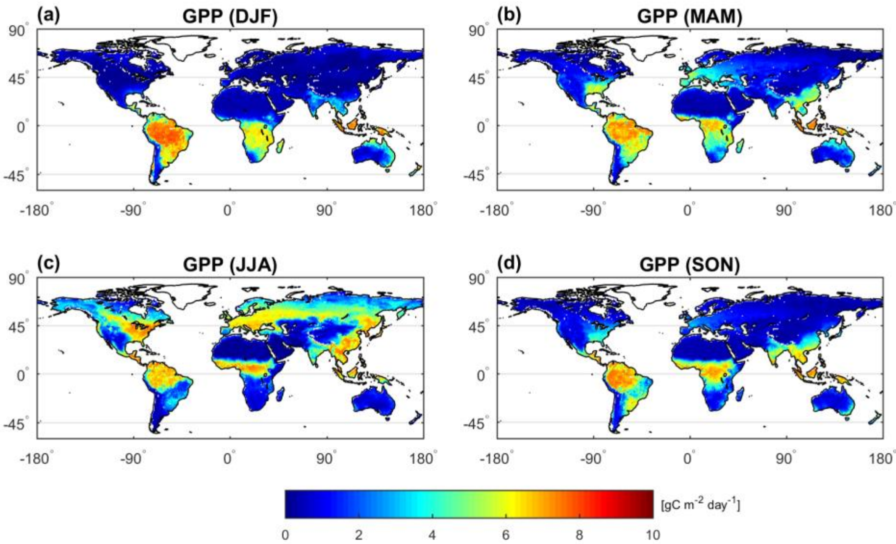
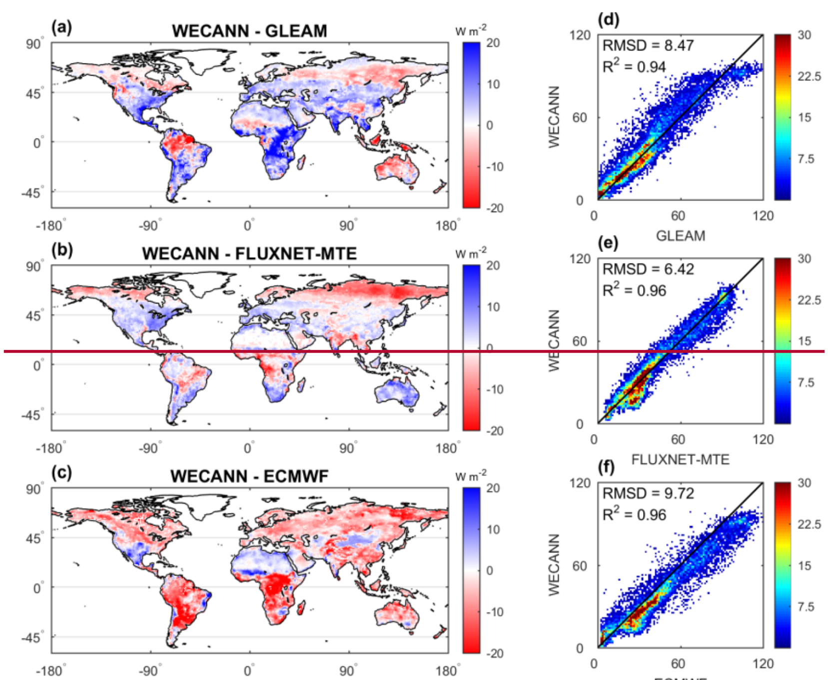


Figure 5: Similar to Figure 3 but for GPP instead of LE



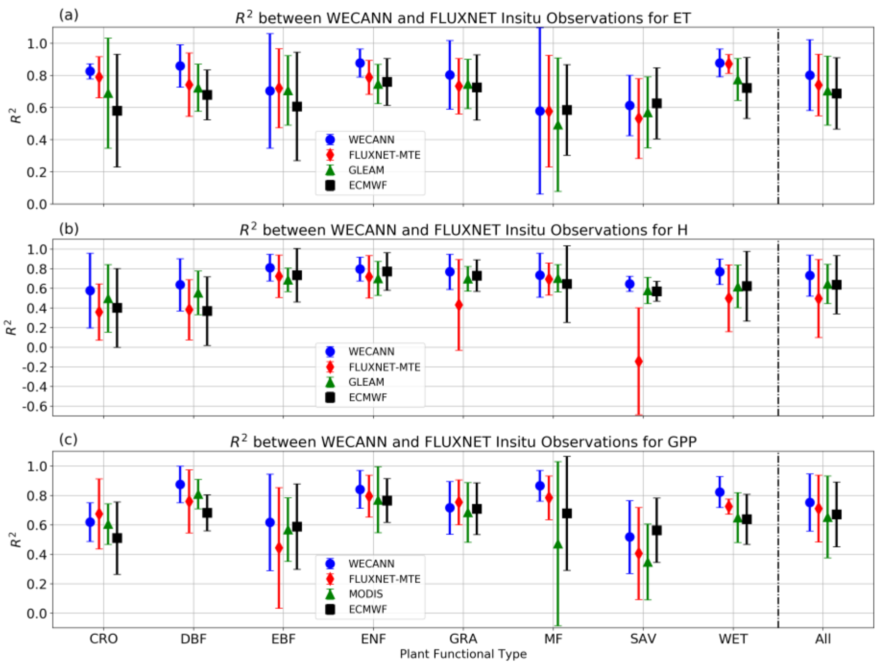


Figure 6: Correlation coefficient (R^2) between WECANN retrievals and FLUXNET tower estimates categorized across different plant functional types for (a) LE, (b) H, and (c) GPP. Markers show mean, and whiskers show one standard deviation intervals. (CRO=Croplands, DBF=Deciduous Broadleaf Forests, EBF=Evergreen Broadleaf Forests, ENF=Evergreen Needleleaf Forests, GRA=Grasslands, MF=Mixed Forests, SAV=Savannas, and WET=Permanent Wetlands)

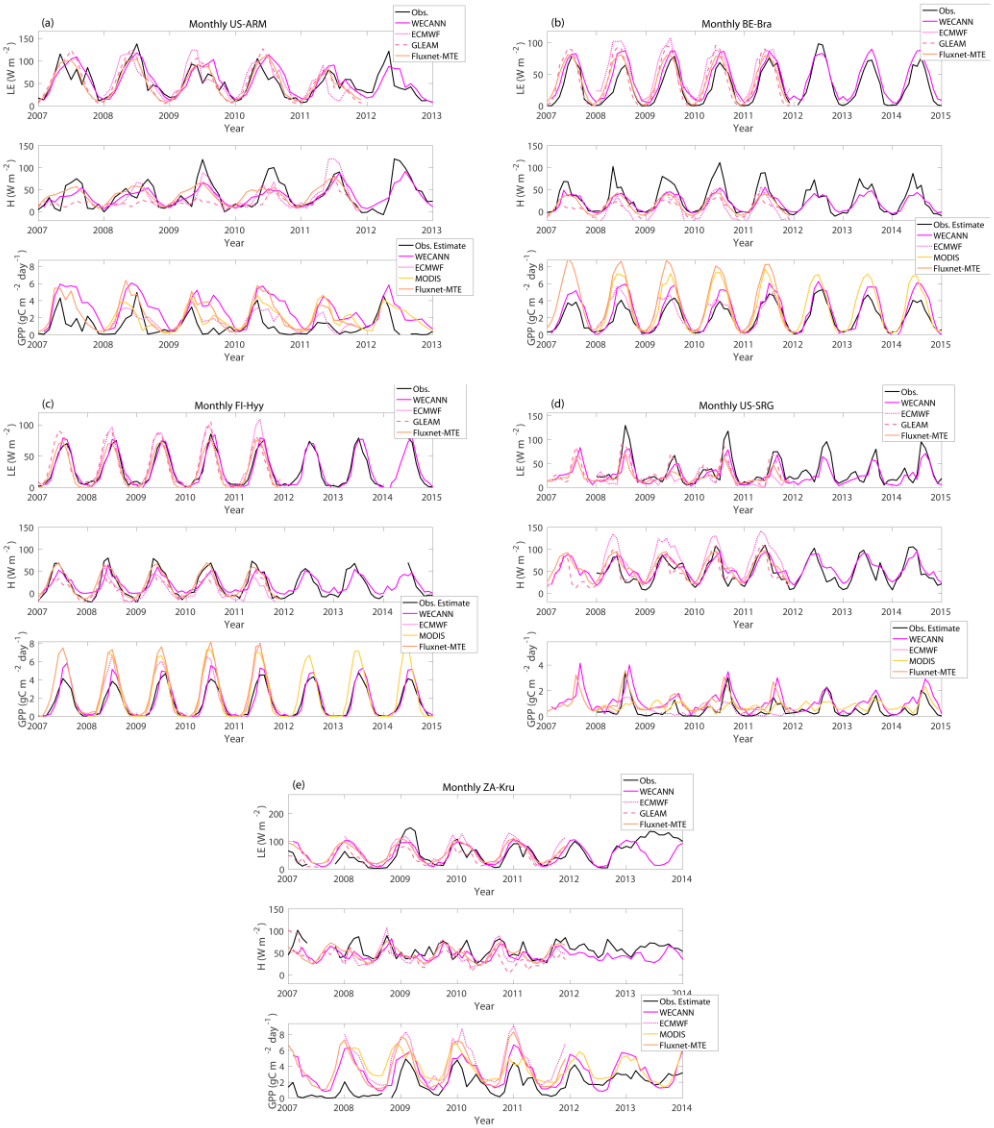


Figure 7: Comparison of the flux retrievals with eddy covariance observations of LE, H and GPP across 5 sites (a) US-ARM site, USA, (b) AT-Neu site, Austria, (c) BE-Bra site, Belgium, (d) FI-Hyy site, Finland, (e) US-SRG, USA, and (f) ZA-Kru, South Africa

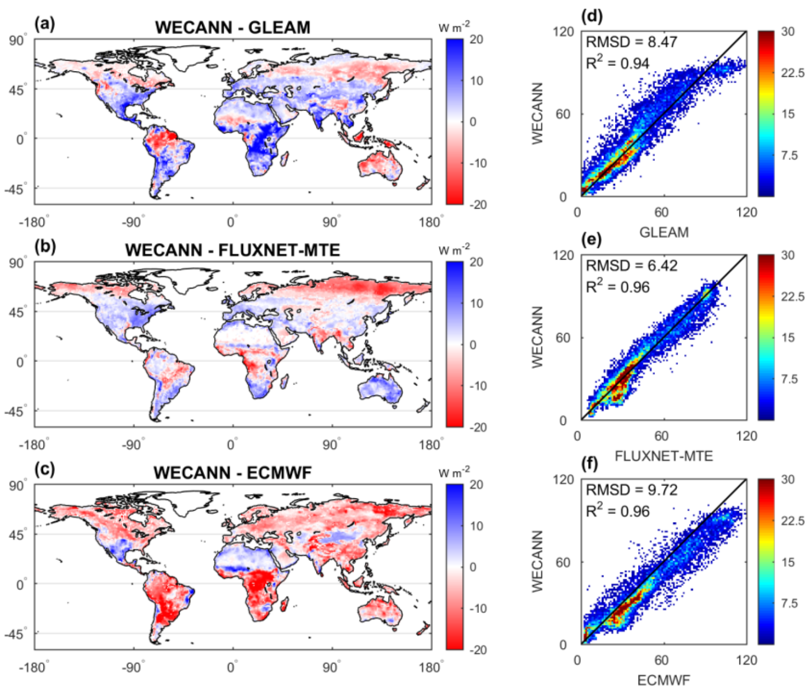


Figure 8: Difference between annual mean LE retrieved by WECANN and the three target datasets (a-c). Scatter plots of LE retrieved from WECANN vs. from each of the target datasets (d-f). Data used are from 2011.

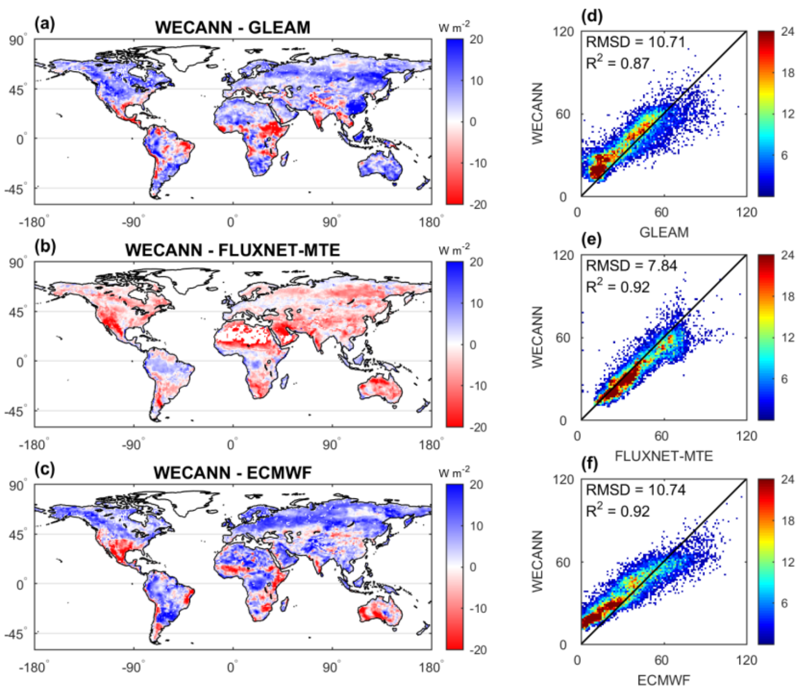


Figure 9: Similar to Figure 8 but for H instead of LE

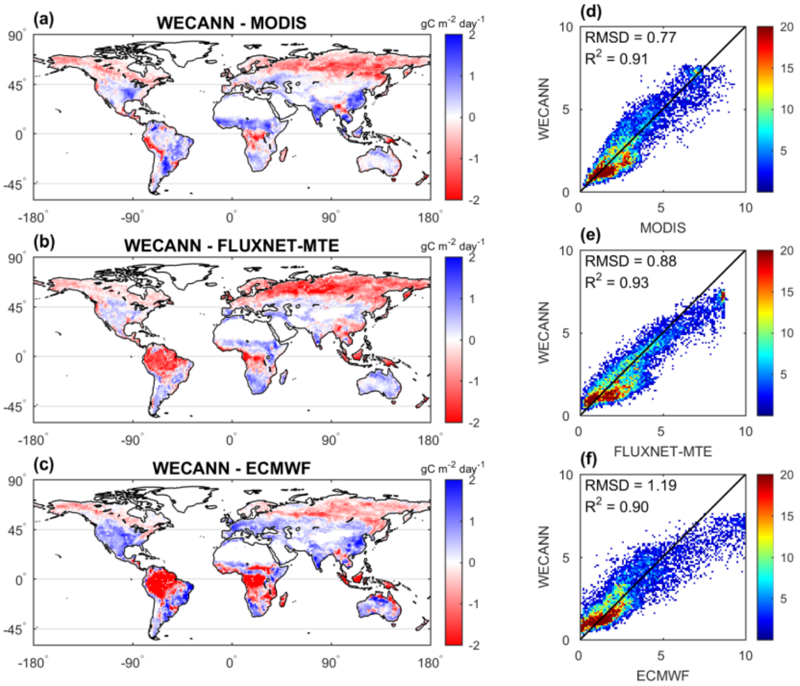
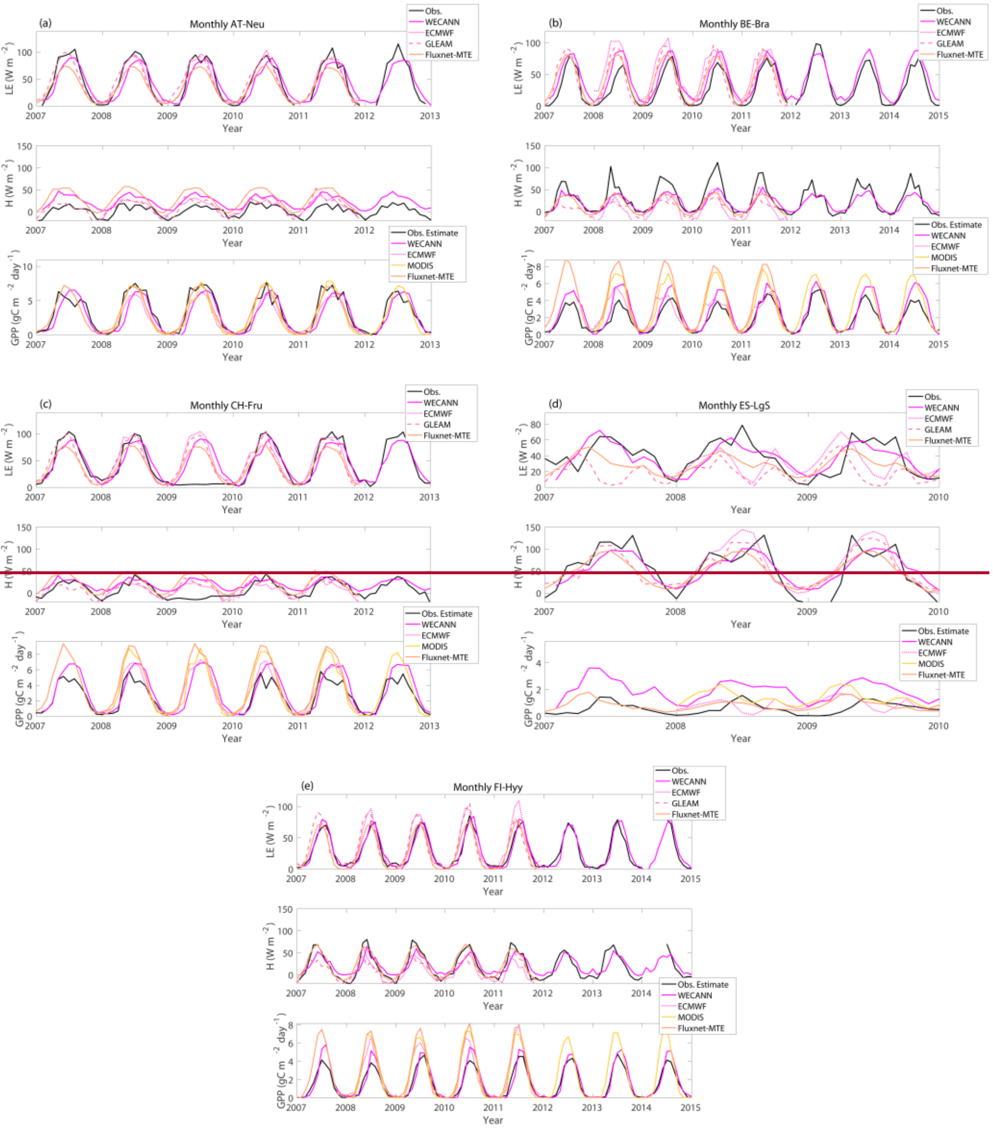
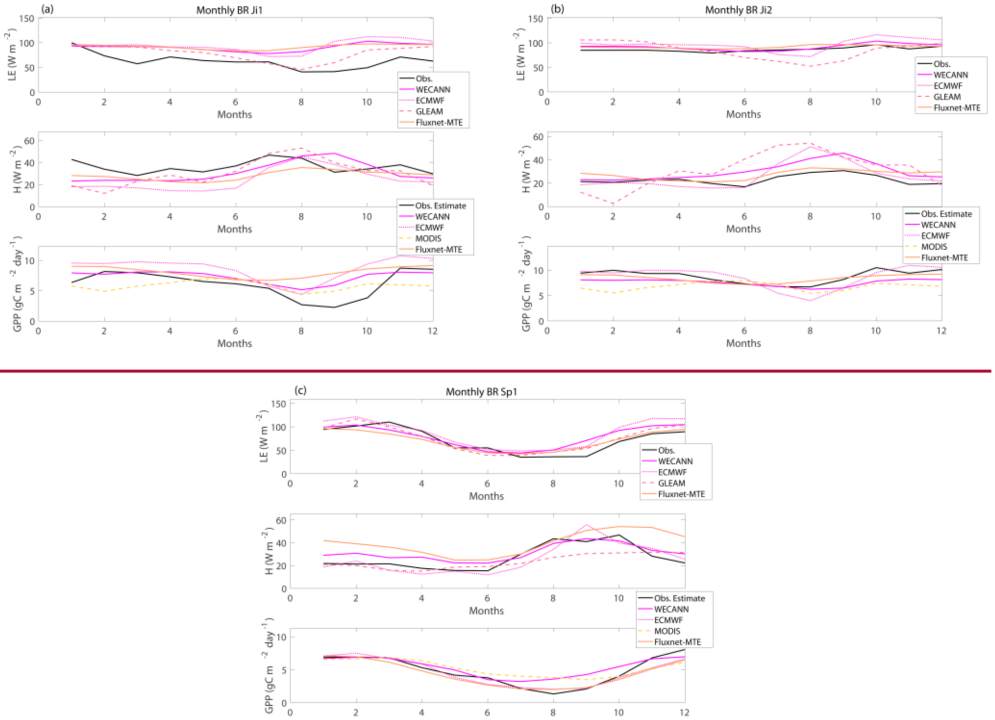


Figure 10: Similar to Figure 8 but for GPP instead of LE

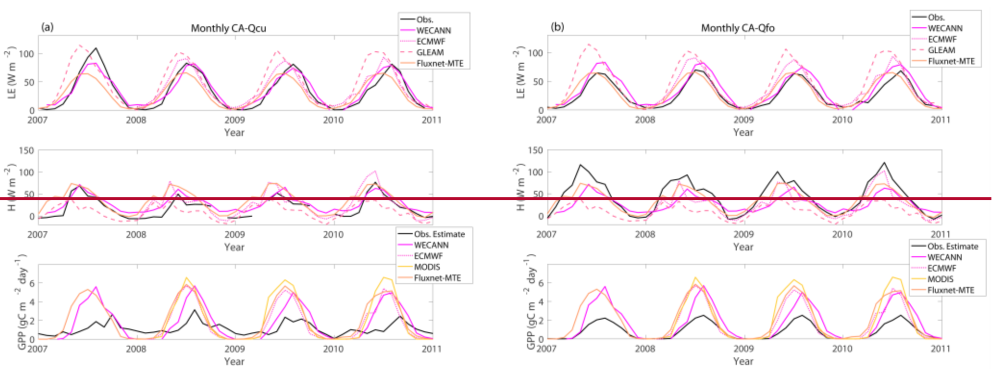
Formatted: Font: 9 pt



~~Figure 9: Comparison of the flux retrievals with eddy covariance observations of LE, H and GPP across European sites (a) AT-Neu site, Austria, (b) BE-Bra site, Belgium, (c) CH-Fru site, Switzerland, (d) ES-LgS site, Spain, and (e) FI-Hyy site, Finland~~



~~**Figure 10: Same as Figure 9 but for Brazilian sites (a) BR-Ji1, (b) BR-Ji2, (c) BR-Sa3, and (d) BR-Sp1.**~~



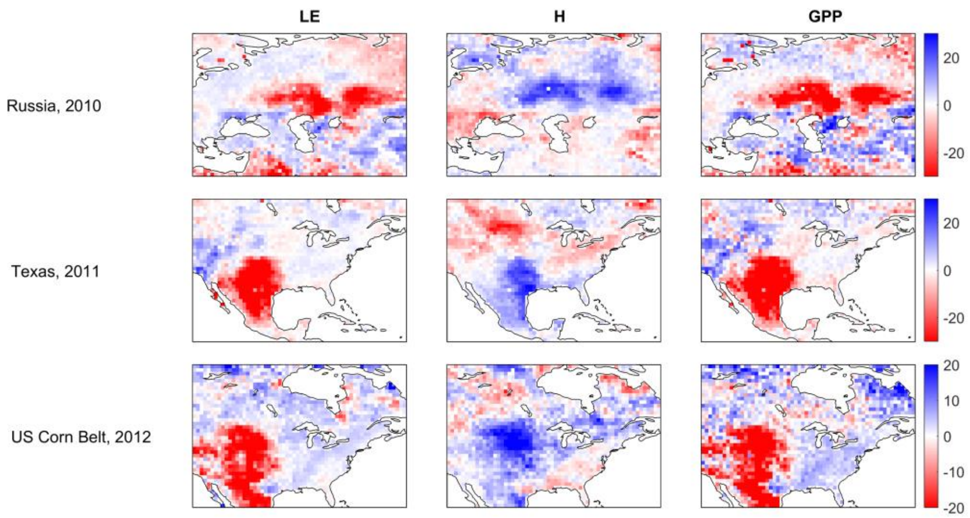
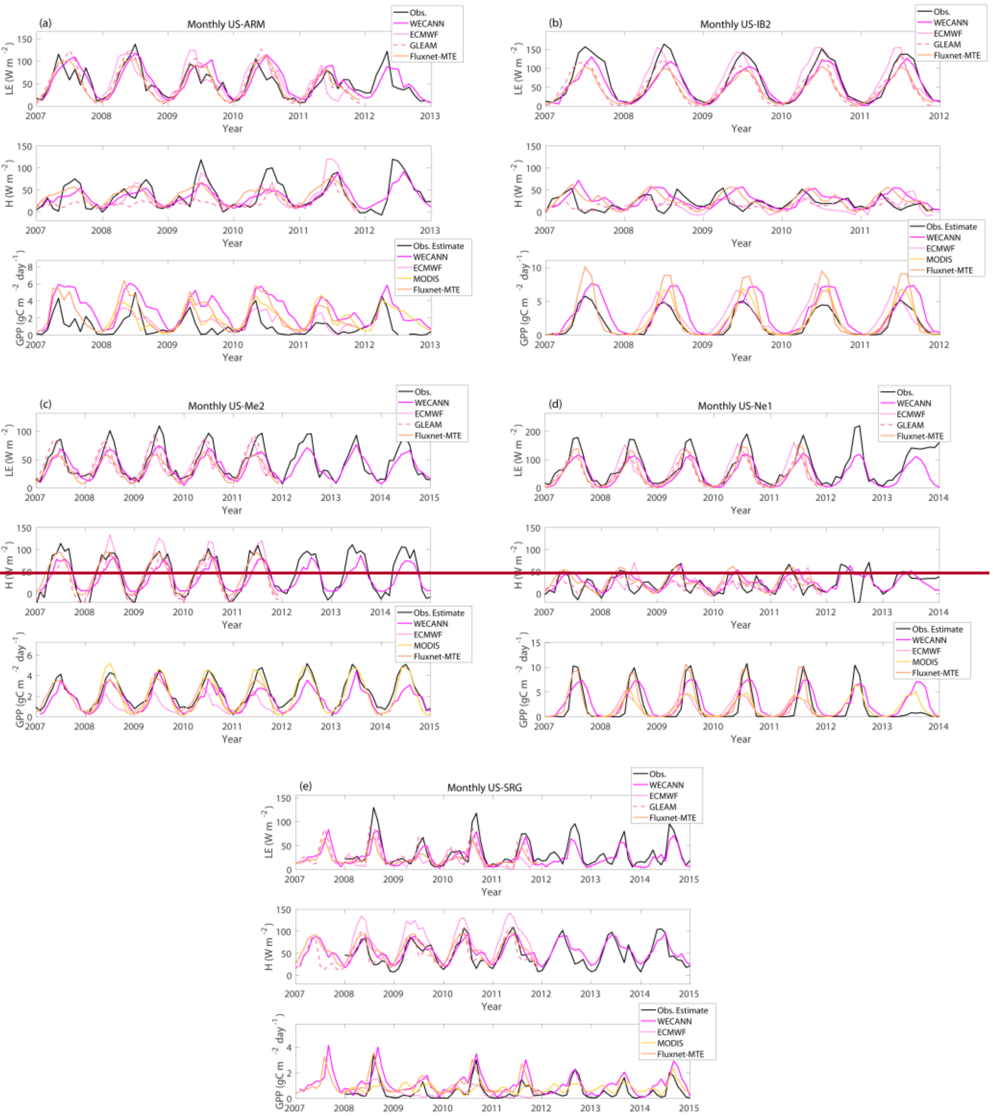


Figure 11: ~~**Same as Figure 9 but for Canadian sites (a) CA-Qfo, and (b) CA-Qcu**~~**Mean monthly anomalies (in percentage with respect to mean value of each flux) for three extreme heatwave events.**

Formatted: Font: 10 pt, Not Bold, Font color: Auto

5


(a)
Monthly US-ARM
Obs.
WECANN
ECMWF
GLEAM
Fluxnet-MTE
LE (W m $^{-2}$)
H (W m $^{-2}$)
Obs. Estimate
WECANN
ECMWF
MODIS
Fluxnet-MTE
GPP (gC m $^{-2}$ day $^{-1}$)
Year
(b)
Monthly US-IB2
(c)
Monthly US-Me2
(d)
Monthly US-Ne1
(e)
Monthly US-SRG

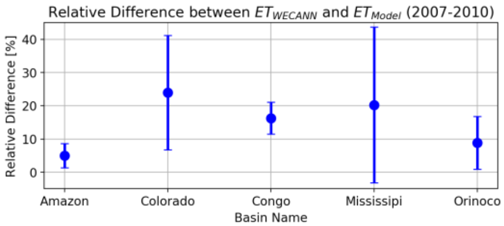
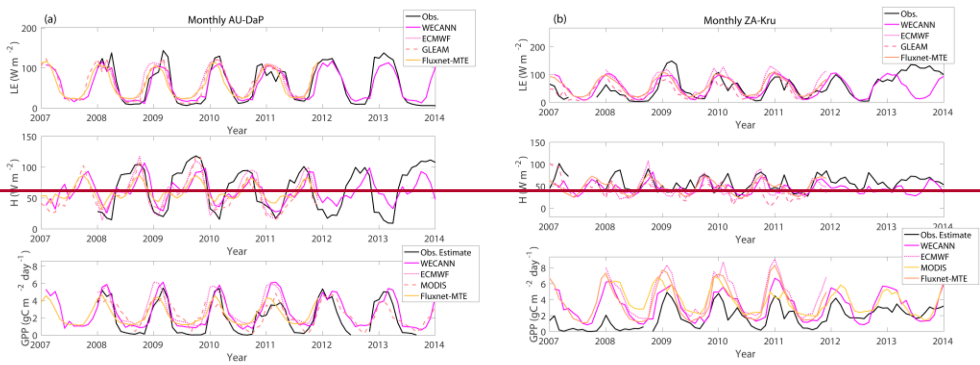


Figure 12: ~~Same as Figure 9 but for US sites (a) US-ARM, (b) US-IB2, (c) US-ME2, (d) US-Ne1, and (e) US-SRG~~**Relative difference between ET estimates of WECANN compared to modeled ET from basin scale water budget closure. Markers show mean, and whiskers show one standard deviation intervals.**

Formatted: Font: 10 pt, Not Bold, Font color: Auto



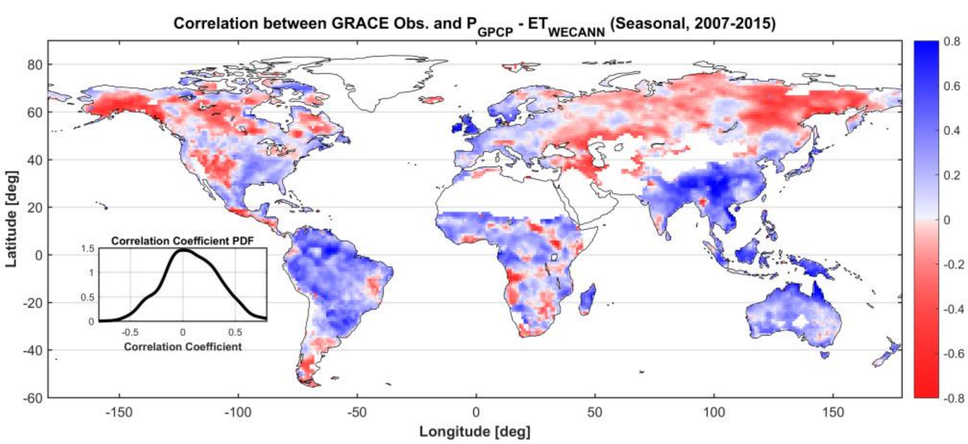


Figure 13~~: Same as Figure 9 but for (a) AU-DaP, Australia, (b) ZA-Kru, South Africa~~

: Correlation between seasonal values of GRACE terrestrial water storage anomaly and difference between precipitation and evapotranspiration (estimated by WECANN). Regions with an annual precipitation of less than 200 mm are excluded from the analysis.

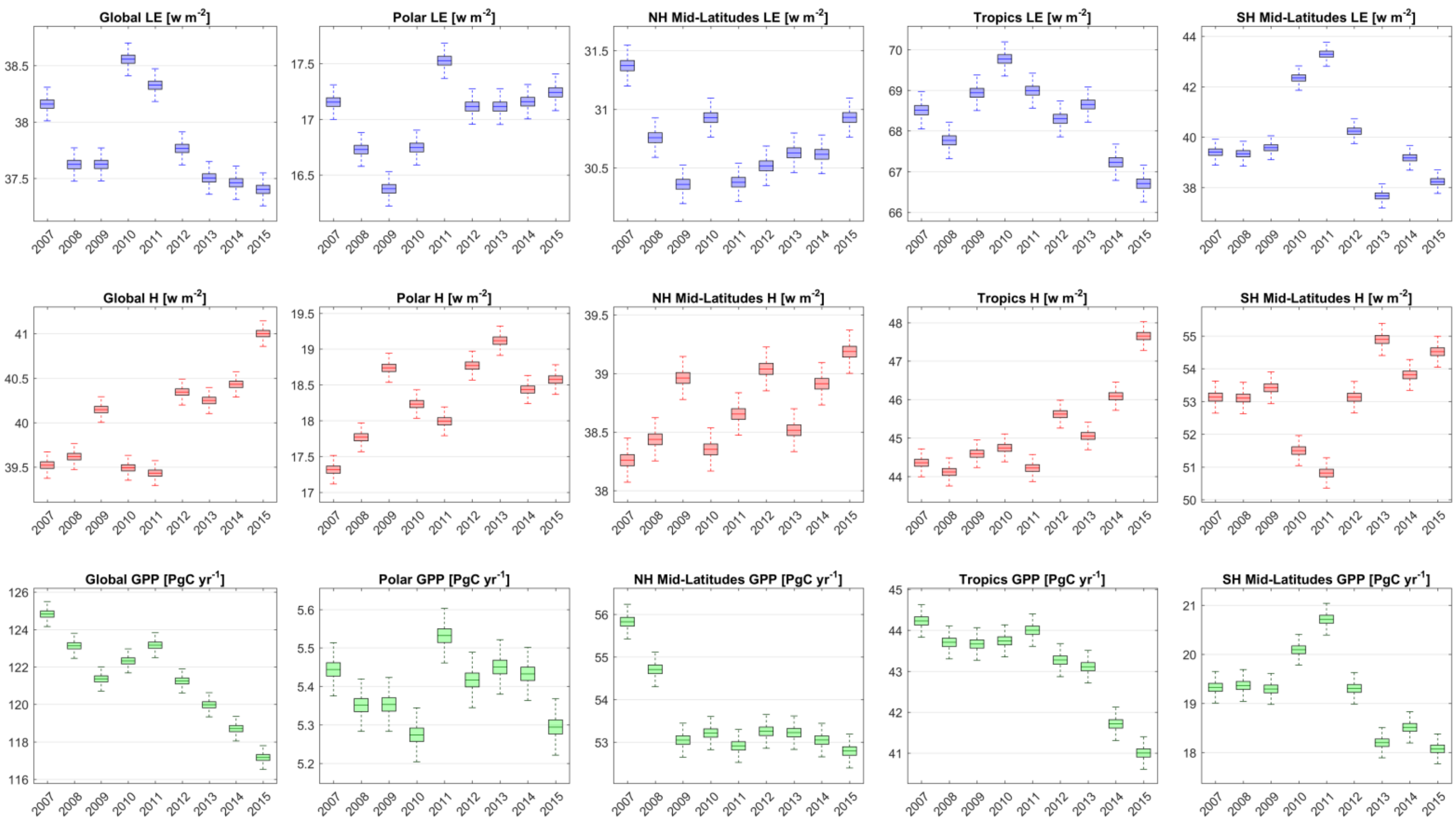


Figure 14: ~~Uncertainty~~Annual mean estimates and uncertainty bounds of LE (top row), H (middle row) and GPP (bottom row) retrievals at global (left column) and regional (four right columns) scales between 2007 and 2015. The central line in each box indicates the mean, the edges of the box are 25th and 75th percentiles, and the whiskers show the most extreme values.

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° × 0.25°	Martens et al., 2016
ECMWF ERA HTESSEL	LE, H, GPP	2008 - 2015	Global	Daily	0.25° × 0.25°	Balsamo et al., 2009
FLUXNET-MTE	LE, H, GPP	1982 - 2012	Global	Monthly	0.5° × 0.5°	Jung et al., 2009
MODIS-GPP	GPP	2000 - 2015	Global	Monthly	0.5° × 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° × 0.5°	Joiner et al., 2013
Net Radiation	CERES L3 SYN 1deg	2002-present	Global	Monthly	1° × 1°	Wielicki et al., 1996
Air Temperature	AIRS3STD v6.0	2002-present	Global	Daily	1° × 1°	Aumann et al., 2003
Soil Moisture	ESA-CCI v2.3	1978-2015	Global	Daily	0.25° × 0.25°	Liu et al., 2012
Precipitation	GPCP 1DD v1.2	1996-2015	Global	Daily	1° × 1°	Huffman et al., 2001
Snow Water Equivalent	GLOBSNOW L3A v2	1979-present	Global	Daily	25 km × 25 km	Luojus et al., 2013