

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
<p>The authors proposed a new global product of GPP, ET and H by using ANN. The manuscript is well written and the topic falls on to the scope of the journal. I do have several concerns.</p>	<p>We thank the referee for his/her positive comments.</p>
<p>First, the authors highlighted the use of SiF as input data. I see SiF was the only input data related to vegetation. Therefore, with/without SiF in WECANN must give different flux estimates. What happens if the authors use EVI or NDVI instead of SiF? Any significant difference in WECANN performance?</p>	<p>We agree with the referee's point on the evaluation of no-SiF retrieval. Therefore, in the revised manuscript we included comparisons with an artificial neural networks retrieval that has either NDVI or EVI as input instead of SiF. We comment on the differences and similarities, and why SiF is a better input for this retrieval, in particular highlighting the differences in terms of vegetation structure impact on SiF and the impact of saturation of vegetation indices (especially in forested areas and agricultural regions). Results are summarized in section 4.5 and Tables S1-S3 in the supplementary materials.</p>
<p>Second, what is the significant contribution from this work? Spatial (1 degree) and temporal (monthly) resolutions are too coarse. The approach is on the similar family of other machine learning methods (e.g. see Tramontana et al 2016 Biogeosciences). Stress the novelty of this manuscript. If there is any new discovery, then highlight it.</p>	<p>There are two major new contributions in this study:</p> <ol style="list-style-type: none"> 1- Using remotely-sensed SiF to estimate surface fluxes. 2- Using a machine learning algorithm (in this case artificial neural networks) to estimate fluxes from remote sensing observations at global scale. <p>The Tramontana et al 2016 paper uses a regression model to upscale fluxes from FLUXNET observations. However, we use remote sensing observations to estimate fluxes rather than relying on the representativeness of spatially limited FLUXNET eddy-covariance data like the Tramontana et al approach and its predecessors. Most importantly we use SiF as an indicator of vegetation activity. Therefore, the strategy is pretty different compared to the Tramontana et al retrieval. In addition, our main objective is to show that SiF provides useful information on the rates of photosynthesis and evapotranspiration. To our knowledge, this is the first direct estimate of fluxes based on SiF data. We revised portions of the text in the introduction section (Page 2, Lines 30-33) to make sure the novelty of our approach is clearly stated.</p>

<p>Third, the authors used MPI-BGC product as a training dataset while testing the product against FLUXNET data. As MPI-BGC product was trained against FLUXNET dataset, the approach is self-correlated. Why not evaluating the product against independent datasets from MPI-BGC? E.g. water balance derived ET in basin scale.</p>	<p>We do not share the referee’s perspective. While we acknowledge that there is some information carried from FLUXNET tower data into the FLUXNET-MTE dataset and therefore in some part of the training data, we believe this cross-correlation is likely to be small.</p> <p>We train our algorithm against a target dataset which is derived from three products (including MPI-BGC) by using the Triple Collocation method and assigning <i>a priori</i> weights to every product in each pixel. This means that our target dataset has collective information from all three products and not just MPI-BGC. Furthermore, the correlation between FLUXNET and FLUXNET-MTE data is also imperfect (cfg. Figure 2 in Tramontana et al 2016).</p> <p>Interestingly, WECANN typically outperforms other products (including FLUXNET-ME, which is expected to have a stronger correlation to FLUXNET data) especially in terms of seasonal cycle. This further emphasizes the information content provided by remote sensing data which are used as additional inputs</p> <p>This is consistent with previous work (see Jimenez et al. 2009 for instance) that the spatial and temporal correlations of a global artificial neural network are not due to the initial training dataset but to the remote sensing observations used as input.</p> <p>Jimenez, C., Prigent, C., Aires, F. (2009). Toward an estimation of global land surface heat fluxes from multi-satellite observations. <i>Journal of Geophysical Research-Atmospheres</i>, 114(D6), D06305.</p> <p>Conducting a water balance analysis is an interesting idea that might be informative, but it has its own challenges because multiple sources of information need to be used to close the water budget, each of which has its own uncertainties. Furthermore, such an approach would only be useful for validating the ET data but would not provide information about the GPP and H performance. In view of these reasons, and because the other referee asked us to reduce the length of the manuscript, we have chosen not to include this analysis.</p> <p>Nevertheless, to provide an additional line of evidence investigating the WECANN quality, we have now added a new section 4.4 with an uncertainty analysis (which is much briefer than would be required for a full discussion of a water budget comparison).</p>
<p>Fourth, the spatial domain should be clearly defined. The authors said it is global product, but it did not include Antarctica and Greenland.</p>	<p>Thank you for the comments. We have now revised the description in the introduction section to clearly note what</p>

<p>Given the coarse resolution (100 km), most islands are likely uncovered but the global map (Fig 2) showed fluxes in some islands. How did it happen? Also, how to treat with water fraction for each 1-degree pixel?</p>	<p>the coverage of the new product is (Page 2, Lines 26-29), and provided a land mask in Figure S1.</p>
<p>Fifth, I recommend showing global uncertainty maps for GPP, LE, H. I think one of strengths in WECANN is its ability to quantify uncertainty. Show the uncertainty map and discuss where and why uncertainties are high. Also quantify uncertainties in global values (e.g. XXX PgC yr-1 +- Y PgC yr-1).</p>	<p>In the revised manuscript, we now include uncertainty estimates based on errors in the input data propagated into the network. We report a global average value as error is spatially and temporally variable. The new section 4.4 in the revised manuscript provides details on our uncertainty analysis and the results that are provided in Figure 14.</p>
<p>Sixth, test global more carefully. When I look at Fig 2, I found higher ET in mid to south east South America (e.g. cerrado) compared to other global ET products. Also, your ET in this region is relatively very high compared to your GPP map. So, water use efficiency will be very low in this region, which is unlikely. See global distribution of C4 maps. Higher proportion in C4 in this area is likely to lead higher water use efficiency. It is notable that your ANN did not consider C4 information.</p>	<p>The referee's point is an important one. The SIF relationship with GPP will likely change in C4 plants. However, we explicitly did not want to impose the C4/C3 (or even CAM) delimitation in the artificial neural network as it would be highly dependent on the quality of the classification map used. Given that we do not have partitioning of transpiration to total ET, it would be impossible to say whether the water use efficiency is indeed low or if rain re-evaporation and soil evaporation is the main process explaining the difference. We note that all training products include C3/C4 delimitation and therefore the C3/C4 delimitation is implicit in the training dataset and therefore can be learnt by the network. We have added a comment in the text emphasizing the referee's points (Page 9, Lines 17-19).</p>
<p>Specific comments</p>	
<p>P6: why only 21 FLUXNET sites were used? More than 150 sites data are open to public</p>	<p>We had selected these 21 sites to represent a range of climatic conditions along a geographical gradient for validation of our retrieval. Presenting evaluation metrics and temporal time series for 150 sites would lengthen the manuscript and make it now hard to read. However, in the revised manuscript we present summary statistics from a comparison of WECANN retrievals against a much larger number of tower data (97) from the FLUXNET 2015 and the La Thuile synthesis dataset in the supplementary tables S1-S3. We also comment on the results in Section 4.4.</p>
<p>P6 L23-24: The authors explained that target data is used for training, validation, and testing. I am confused with the terminology of validation and testing. How do they differ? Also, in L36, "after training, . . . was evaluated". Here, does "evaluation" indicate validation or testing? I recommend clearly defining each term, and use them consistently across the whole manuscript.</p>	<p>We apologize for the confusion. The training, validation and testing proportions are related to the training phase of the retrieval. The back propagation algorithm uses a portion of the training data for training (basically estimating the weights of each neuron), and other portions of the training data for validation and testing that aims at checking the convergence of the training step. While after the training is done, we use a subset of data that were not used in the</p>

	training process for evaluation. We revised the text in the new version of the manuscript to clarify these terminologies. (Page 7, Lines 5-10)
P6 L30: NN -> ANN (?)	Our apologies, this has been corrected in the revised manuscript.
P7 L9: Please define “multiple datasets.” Is this training dataset?}	This refers to the three products that we use (together with error weights from Triple Collocation) to define a target dataset for training. We revised the text in the new version of the manuscript to clarify this.
P7 L12: What is “this” in “this prior distribution”?	It refers to the pseudo Bayesian training mentioned in the lines before. We revised the text in the new version and clarified the point.
P8 L20: Is this “target estimate” from 3.2?	Yes, this is the same. We made changes to section 3 and 4 of the manuscript in the new version to clarify all these terminologies.
P8 L22: Add another unit for GPP as PgC yr-1, which could be easily compared to the other studies. Same for LE (km3).}	Thanks for noting this. We included the new units along with previous ones in the new version of the manuscript.
P9 L29: I was surprised to see the reduction of GPP in the Saharan Desert after removing SiF. How to interpret this as we know there must be zero GPP? Also, exclusion of SiF in LE made mixed tendencies in this region. As we are confident LE and GPP are close to nil in this area, it will be interesting to test the impacts of inclusion/exclusion in SiF on LE and GPP here.}	This observation is true, and is caused by noise in the SiF data in deserts. As noted correctly by the referee, LE and GPP are close to zero in this region; therefore, the difference between the two retrievals (with and without SiF) divided by the small amount of flux in this region is on the order of the noise level in the retrievals. While the percentages of change are notable, the absolute values of difference between the two retrievals are less than 2 w m^{-2} for LE and less than $0.7 \text{ gC m}^{-2} \text{ day}^{-1}$ for GPP. In addition, the noisy pattern does not show up in the H retrievals in this region. This is also another sign that the change patterns in LE and GPP are due to noise. However, due to the request of both referees we have revised our section on the impact of SiF, and our analysis now focuses on the differences between a retrieval with SiF or with NDIV/EVI (the new Section 4.5). Therefore, this figure was removed from the manuscript.
P10 L3: All three R2 looks too similar, so it is hard to tell 0.96 is higher than 0.94.}	Yes, we agree and have typically highlighted with bold fonts all comparable products for a fair comparison. This is further emphasized in the text of the revised version of the manuscript.
P10: The authors compared WECANN to FLUXNET-MTE, ECMWF, GLEAM and MODIS-GPP which were the training data for WECANN. I feel there should be self-correlation, so I am curious whether this is a reasonable approach.	The focus of this comparison is not validation. Since we used the three training products to generate the target dataset, we compare WECANN to these three to examine how similar is it to each of those training datasets. And we show that spatially WECANN is more similar to the product that has the lower RMSE in our TC estimates.
P10 L8: I know there are few eddy flux tower data in India, so FLUXNET-MTE might involve higher	It is true that there are few towers in India, but our retrieval does not rely solely on regional towers to estimate surface

<p>uncertainty. However, this is the same situation for WECANN as it used FLUXNET-MTE and others, which are all uncertain as training dataset.</p>	<p>fluxes. Indeed this is a major advantage over FLUXNET-MTE and others. We train an artificial neural network algorithm using the three training products (two of which are not based on flux towers, and so do not necessary have higher uncertainty in certain areas because there are few towers there) mentioned in the manuscript for all the pixels. However, the actual time-scale retrieval is mostly informed by the remote sensing observations (see discussion in Jimenez et al. 2009). That means we use the information from all the pixels over the globe to train one retrieval algorithm. This algorithm uses remote sensing observation at each point in time and space to retrieve surface fluxes. Therefore, lack of FLUXNET towers in any part of the globe would not impact the accuracy of WECANN retrievals, while this would expected to be more be an issue for products that upscale tower-based observations to estimate fluxes across the globe.</p>
<p>P10 L4: Be quantitative. Report bias.</p>	<p>Thanks for noting this. The point we have raised in this line (on the spread of scatter plots) can be quantitatively compared using the RMSD value that are provided in the figures. In the revised manuscript, we report this in the text.</p>
<p>P10 L20: Define "G"</p>	<p>Corrected in the revised manuscript.</p>
<p>P11: Many contents in this page should move to Methods.</p>	<p>In the revised manuscript, we re-organized the text and moved these contents to section 2.3.</p>
<p>P12 L5-6: Then why not removing this site given obvious deficiencies?</p>	<p>We believe that it is informative to include this comparison, as it illustrates that the representativeness area can be a challenge in comparing large-scale remote sensing based retrievals to point based tower data. In this case, we have good knowledge of the site and its surrounding region so it is possible to investigate if the tower data is representative of the larger scale fluxes.</p>
<p>P12- : As the authors well recognized, I feel it is odd to compare 1 degree WECANN to several hundred meters in flux towers. All discussion from this comparison seems too subjective. I think "validation" of 0.5-degree product is unlikely possible. As your products are too coarse, I would recommend evaluating at larger scales. For example, look interannual variability of global GPP (PgC yr-1), ET (W m-2), and H (W m-2) and compare to atmospheric inversion estimates. Test whether your product could capture big climate extreme events such as Russian heatwave, Texas drought etc. Compare to other existing global land surface products</p>	<p>While there is some caveat in validation against point based tower data, these are the only ground based observations available for such a validation. Moreover, in the comparison against tower data many large scale variabilities, including but not limited to the seasonal cycle are comparable to pixel based retrievals. This is also the case for interannual variability, and we have discussed them in detail, in section 4.4 of the original manuscript (section 4.3 of the revised manuscript). For instance the phenology has a strong impact on the seasonal cycle of the fluxes and is here clearly highlighted when comparing the different products to flux tower estimates. In the revised manuscript, we highlighted this limitation clearly in section 4.3, while noting that comparison against ground-based tower observations is common practice and</p>

<p>which were not used as input/training dataset in WECANN.</p>	<p>is what the community indeed looks for when a new retrieval algorithm is developed. We believe that specific drought or flood events would lack the generality provided here when comparing all years/months. Moreover, such a comparison needs detailed analysis that would further lengthen the manuscript (Indeed, the other referee asked us to reduce the length of the manuscript). In addition, in the new section 4.4 we provide uncertainty estimates of the retrievals along with interannual variability of surface fluxes at global scale to provide an additional line of evidence on the quality of the WECANN dataset.</p>
---	--

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
<p>This manuscript is well written and deserves consideration for publication in this journal. However, I have the following issues that need to be addressed.</p>	<p>We appreciate the referee’s positive feedback and provide responses to his/her comments below.</p>
<p>The paper proposes an empirical machine learning ‘meta-model’ to try to learn from different existing datasets to combine their strengths and factor out their limitations. On one hand, I appreciate this effort to bring together different datastreams and somehow harmonize them through this new consolidated product, but on the other, I am wary of this approach of blindly adding further algorithmic layers without really trying to understand mechanistically why the initial datasets have shortcomings. If all products are equally off in some parts, combining them just gives the false impression we are going in the right direction while reality is still off. Also, the FLUXNET-MTE used as training is already a machine learning product driven by various input variables, very much like WECANN is. Furthermore, there is quite some circularity in the work since the FLUXNET-MTE and MODIS GPP are both strongly based on the same fluxtowers used here for validation. I deem that all these points need to be acknowledged clearly and discussed thoroughly.</p>	<p>We acknowledge this concern, and would like to bring the following points to the referee’s attention:</p> <ol style="list-style-type: none"> 1- The WECANN machine learning retrieval is quite different from FLUXNET-MTE in the sense that we use remote sensing observations as inputs while FLUXNET-MTE upscales tower-based observations to estimate surface fluxes at global scale. Although both approaches use machine learning techniques (artificial neural networks in the case of WECANN and regression in the case of FLUXNET-MTE) their retrieval algorithms are quite different and directly informed by only remote sensing observations in WECANN, which we believe is an important means of better constraining the retrievals. 2- Moreover, our training approach uses all the spatial and temporal observations during the training period (2008-2010) to develop one single neural network for the global retrievals. This network is then used with remote sensing observations as input to retrieve surface fluxes. Therefore, if a few percentage of times and pixels, all the three training products are equally off this will be mitigated by the larger number of pixel/time data points that have more accurate estimates in other places and other times. In addition, the network can even correct the seasonal cycle when learning from an incorrect seasonal cycle training data, as the remote sensing inputs provide the information on the seasonal cycle directly. This has already been demonstrated previously, c.fg. Jimenez et al. (2009). Jimenez, C., Prigent, C., Aires, F. (2009). Toward an estimation of global land surface heat fluxes from

multisatellite observations. *Journal of Geophysical Research-Atmospheres*, 114(D6), D06305.

- 3- On the issue of validation against FLUXNET tower data, we acknowledge that two of the training products use FLUXNET data for their calibration or as input. However, virtually all products have been calibrated in some ways or tested against eddy-flux tower, so implicit circularity is hard to avoid; there simply isn't another high quality data-set available. This does not however mean that the products are not independent: indeed the training products we use are typically calibrated to reproduce either the annual mean or are adjusted per season at very few sites (but not the exact temporal structures of the eddy-covariance observations except for FLUXNET-MTE). In addition the specific years of observations used here were not used in the calibration of MODIS and FLUXNET-MTE.

Here, we use the three training products together with a priori weights calculated from Triple Collocation to define a target dataset that has collective information from all three of them. And then we train our network on the target dataset. Finally, we validate the retrievals of WECANN against FLUXNET tower data and compare its performance with the performance of the three training products. While some information from FLUXNET observations propagates through the training products to WECANN training, the comparison results against FLUXNET observations show that WECANN learns from the three products collectively and performs better than any of them individually, emphasizing that our strategy works well. In addition, it is clear that WECANN does not have the seasonal biases seen in most retrievals (see e.g. FI-Hyy site where WECANN correctly captures this cold region's photosynthesis and evapotranspiration compared to the other products). Nevertheless, we have tried to also provide alternative lines of evidence to support the WECANN data quality, including an entirely new uncertainty analysis in section 4.5

We also made changes to sections 3 and 4 of the manuscript in the revised version to better reflect on these points. Moreover, in the new section 4.4 we now provide uncertainty estimates on WECANN retrievals to provide an additional line of evidence on the quality of the WECANN dataset.

<p>Could you specify why you use the SYN products (Level 3) from CERES instead of the EBAF ones (Level 3B)? The later have been energy balanced according to the product specifications. Wouldn't this be an advantage in your case?</p>	<p>Our goal here, as also mentioned in section 1 of the manuscript, is to only use remotely-sensed observations as input. The EBAF product is based on a model with some remote sensing observations; therefore, we decided to use the SYN product to avoid any model addition.</p>
<p>In the construction of the ANN, I would welcome to have some justification of why tangent sigmoid transfer functions are used instead of linear ones. I know this is often done, but it seems very arbitrary.</p>	<p>In this case, we tried the tangent sigmoid (the common choice) as well as linear, and did not see any notable changes in the performance of the network. Therefore, we chose to use the typical tangent sigmoid function. This has been added to the text (Page 6, Line 36 – Page7, Line 2).</p>
<p>Also, I did not quite understand how the 20% of 'testing' data is used. I clearly see that 60% are used for training and 20% for validation, but how exactly do you use the other 20%? Perhaps this just needs some rephrasing in the text for clarification.</p>	<p>We apologize for the confusion on this matter. In the revised manuscript, we explained this in more detail (Page 7, Lines 5-10). In summary, these percentages are for the data that are used in the training process. This is standard practice in artificial neural networks training. The back-propagation algorithm uses the training portion of the data for estimating the weights of the neuron in the network, and the validation and test data are used to evaluate convergence of the training. These are separate than the data that we used for validation later on. Our validation, uses a subset of data that are not used in the training, to make sure the network is not over-fitted to the training data. We revised the text in the new version of the manuscript and clarified the definitions.</p>
<p>Comparison with fluxtower measurements is not appropriate as the difference in spatial support is just too different (1 squared degree vs <1km²). Saying that WECANN performs better than other products based on individual towers while all these products cover such a larger area (by several orders of magnitude) just does not make much sense (even if it has been done in other studies). The authors would need to do some filtering of the towers to select only those that can be considered representative (e.g. http://doi.org/10.1016/j.rse.2016.04.027), although I doubt this would leave many valid towers for pixels of 1 squared degree. Another option that may be more feasible would be to make an evaluation at a larger aggregation scale, such as for clusters of similar climates and plant functional types. Making such averages from the fluxtowers on one side and from all pixels that are comparable in this respect on the other would reduce the number of measurements for validation, but would render them more credible. I would also suggest to exploit more of the</p>	<p>We acknowledge that comparison against point based tower data has its own limitation (as we also have noted in the manuscript), but these are the only ground based validation data that is available for evaluating a new global product. For this reason, we used a selection of sites spanning a geographical gradient and provided detail explanation on the comparison results in each site based on the knowledge of the land cover / land use around the site to make sure the differences, if any, can be explained either by uncertainties in WECANN retrieval or representativeness of the of towers. As the referee notes, filtering the towers based on representativeness might not leave us with any tower to use.</p> <p>In the original manuscript, we only used 21 towers that were selected to represent a wide range of climatic conditions and we would be able to explain the results of each one of them, including the time series plots in detail. However, due to the request of both referees in the revised manuscript we include summary statistics from comparison of WECANN against 97 FLUXNET sites from three datasets: FLUXNET2015, La Thuile Synthesis Dataset and the Large-scale Biosphere-Atmosphere (LBA) experiment in Brazil. Results are provided in Tables S1 – S3 and discussed in Section 4.3 of the revised manuscript. We also want to</p>

<p>available towers in the Fluxnet2015 dataset instead of only 21.</p>	<p>emphasize that some features of the flux towers such as phenology, seasonality are correctly picked up by our retrieval compared to other products and are only moderately affected by the heterogeneity within the pixels (except if there would be a very different even composition of deciduous and conifers for instance).</p>
<p>The part pretending to demonstrate the value of SIF is also inadequate as the authors only test the effect of removing this one input. By doing so, any information of the actual vegetation phenology is lost, which would necessarily reduce the performance. What would be interesting would be to show that SIF provides better information than the classical vegetation indices like NDVI or EVI. To do so, the SIF input of the ANN should be replaced by one of these and then a judgement on the pertinence of SIF can be made.</p>	<p>We appreciate the referee's comment on this point. We have now included comparisons with retrievals that have only NDVI or EVI instead of SIF in the revised manuscript. This better shows the value of having SIF as an input in retrieving surface fluxes. Thank you for this important comment. The results further emphasize the difference between SIF and purely vegetation structure and phenology (as well as saturation effects of vegetation indices). Section 4.5 in the revised manuscript provides the detailed comparison of these retrievals.</p>
<p>Finally, the manuscript is often too long and too descriptive in several parts describing the graphs and maps. This needs to be reduced drastically. Most of what is being said can be easily inferred from the reader by looking at the graphs, while deeper discussion on why discrepancies occur between products and fluxtowers would be more welcome. Also, please remove the extensive references to different parts of the text and the description of the structure of the paper (e.g. page 3 lines 10-20), I think they are lengthening the text needlessly.</p>	<p>Given the novelty of the approach we feel that it is important to correctly describe the different steps of the analysis as many are relatively new such as the machine learning and the triple collocation. We had received the opposite comments before that we were not sufficiently describing the details; hence, the reason why the article goes into the details of the retrievals. We have edited the manuscript throughout and shortened it where possible.</p>

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

5 Seyed Hamed Alemohammad^{1,2}, Bin Fang^{1,2}, Alexandra G. Konings³, Filipe Aires^{1,4}, Julia K. Green^{1,2}, Jana ~~Kolassa^{4,5}~~, ~~Catherine Prigent^{4,6}~~, ~~Filipe Aires⁴~~ Kolassa^{5,6}, Diego Miralles^{7,8}, Catherine Prigent^{1,6}, Pierre Gentine^{1,2,9}

¹Department of Earth and Environmental Engineering, Columbia University, New York, 10027, USA

²Columbia Water Center, Columbia University, New York, 10027, USA

³Department of Earth System Science, Stanford University, Stanford, 94305, USA

10 ~~⁴Universities⁴Observatoire de Paris, Paris, 75014, France~~

⁵Universities Space Research Association/NPP, Columbia, MD, 21046, USA

⁶~~Global⁶Global~~ Modeling and Assimilation Office, NASA Goddard Spaceflight Center, Greenbelt, MD, 20771, USA

~~⁶Observatoire de Paris, Paris, 75014, France~~

⁷Department of Earth Sciences, VU University Amsterdam, Amsterdam, 1081HV, The Netherlands

15 ⁸Laboratory of Hydrology and Water Management, Ghent University, Ghent, B-9000, Belgium

⁹Earth 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 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: a. 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 $1^\circ \times 1^\circ$ 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 ~~strongly constrained by SIF information. The impact of SIF on WECANN retrievals is evaluated by removing it from the input dataset of the ANN, and it shows that SIF has significant influence, especially in regions of high vegetation cover and in humid conditions 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.

20
25
30

1 Introduction

Turbulent fluxes from the land surface to the atmosphere, particularly sensible heat flux (H), latent heat flux (LE), 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 different 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

35

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
5 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 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
10 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) ~~approach~~, trained on globally-representative but ~~noisy/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
15 moisture retrieval (~~Aires et al., 2012; Kolassa et al., 2013, 2016; Rodríguez-Fernández et al., 2015~~)(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 ~~the subsequent data~~ assimilation ~~of retrievals into the model~~ efforts (~~Aires et al., 2012; Kolassa et al., 2013, 2016~~)(Aires
20 et al., 2012; Kolassa et al., 2013, 2016). However, the ~~downside of this approach is that the~~ resulting fluxes estimated by the ANN often ~~show 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,
25 and net radiation as inputs (predictor). To reduce ~~any~~ biases, we introduce a Bayesian perspective to generate the ~~training 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 ~~training 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
30 WECANN (Water, Energy, and Carbon Cycle fluxes with Artificial Neural Networks). WECANN monthly flux estimates for the period 2007 – 2015 are provided on a $1^\circ \times 1^\circ$ resolution grid and with units of W m^{-2} for LE and H, and $\text{gC m}^{-2} \text{ day}^{-1}$ for GPP. ~~The spatial coverage of WECANN is presented in Figure S1. It includes all the land areas, except for Greenland, Antarctica, and any $1^\circ \times 1^\circ$ 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 $0.05^\circ \times 0.05^\circ$ 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 that remotely-sensed SIF estimates are used at the global scale to retrieve surface turbulent fluxes (LE, H, and 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.,
35 1981; Havaux and Lannoye, 1983; van Kooten and Snel, 1990; Krause and Weis, 1991; McFarlane et al., 1980; Toivonen and Vidaver, 1988; van der Tol et al., 2009). Recently, satellite observations of SIF have become available, opening new possibilities

for the global monitoring of photosynthesis (Frankenberg et al., 2011, 2012, 2014; Guanter et al., 2012; Joiner et al., 2013; Schimel et al., 2015; Xu et al., 2015).

~~SIF observations from the Global Ozone Monitoring Experiment-2 (GOME-2) instrument are shown to be more sensitive to plant photosynthesis (both seasonal variability and intensity) compared to typical optical-based vegetation index estimates (such as the Enhanced Vegetation Index—EVI) (Joiner et al., 2011). Another SIF product retrieved from the Greenhouse gases Observing SATellite (GOSAT) has been used to study the impact of seasonal variability on vegetation productivity in Amazon rainforest and shows that SIF is~~
5 ~~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 ~~strong-near~~ linear relationship between ~~GOSAT-based~~ monthly SIF retrievals and GPP is found for different vegetation types which suggests that SIF estimates can ~~be combined with plant physiological fluorescence models for future global carbon cycle research~~ 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
15 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 ~~directly~~ 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.,
20 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.

2 Data

~~This section provides details of each of the remote sensing and/or model based estimates of the variables used as input or target data in the ANN framework, as well as the tower data used to validate the retrievals.~~
30 ~~The inputs to~~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 $1^\circ \times 1^\circ$ geographic grid and averaged at monthly temporal resolution. ~~The ocean and ice covered pixels were masked using the~~
35 ~~land mask data from National Snow and Ice Data Center (NSIDC) (Brodzik and Armstrong, 2013).~~ 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 HTESSSEL. For GPP, training is performed on FLUXNET-MTE, ECWMF ERA HTESSSEL, 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 $0.25^\circ \times 0.25^\circ$ 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 (Guilod 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 $0.5^\circ \times 0.5^\circ$ 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 HTESSSEL

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. HTESSSEL 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 HTESSSEL 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 HTESSSEL provided on a $0.25^\circ \times 0.25^\circ$ 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)~~(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 $0.05^\circ \times 0.05^\circ$ 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.

10 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 well as other trace gases and water vapor content. It senses Earth backscatter radiance and solar irradiance at a 40×80 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 ~~(Frankenberg et al., 2011; Joiner et al., 2011)~~(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 $0.5^\circ \times 0.5^\circ$ grid spacing.

2.2.2 Net Radiation

20 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 $1^\circ \times 1^\circ$ 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 $1^\circ C$ 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 $0.5^\circ \times 0.5^\circ$ 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 review; 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 $0.25^\circ \times 0.25^\circ$ geographic grid.

5 2.2.5 Precipitation

The Global Precipitation Climatology Project (GPCP) provides global daily precipitation estimates at $1^\circ \times 1^\circ$ 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 $25 \text{ km} \times 25 \text{ km}$ EASE grid.

2.3 Validation Dataset: Eddy-Covariance Flux Observations

20 FLUXNET is a network of regional ~~micrometeorological~~ 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 ~~FLUXNET2015~~ FLUXNET 2015, the La Thuile Synthesis dataset for 24 stations and the Large-scale Biosphere-Atmosphere (LBA) experiment in Brazil are used for validation (details are provided in section 4.4).

25 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.

30 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.

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

35 A total of 97 sites from the three datasets are selected for validation of WECANN retrievals. ~~These sites are selected to span spanning~~ a large climatic and biome gradient (~~details are provided in section 4.4~~). 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).

3. Methodology

3.1 Artificial Neural Network Setup

5 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
10 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
15 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 ~~and validation estimates~~ 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 ~~test data~~testing subset are used to assess the ANN performance after the training phase.

25 As an additional measure to avoid overfitting, we repeated the training for several ~~ANNs~~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~~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. ~~S1~~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.

30 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.

35 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 training~~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 most some of the biases of the original training target dataset even if the ANN is able to make corrections to its original training-target data (e.g. correction of an imperfect seasonal cycle, as demonstrated by Jiménez et al., 2009). To address this issue, we use multiple 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), as well as in validating categorical variables such as the soil freeze/thaw state (McColl et al., 2016). The relationship between each measurement and the true value is assumed to follow a linear model:

$$X_i = \alpha_i + \beta_i t + \varepsilon_i \quad i = 1,2,3 \quad (1)$$

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 ($Cov(\varepsilon_i, \varepsilon_j) = 0$, for $i \neq j$) 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):

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

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.

The TC errors from the surface fluxes are shown in Figs. S2-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).

- 5 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:

$$P_i = \frac{\frac{1}{\sigma_{\varepsilon_i}^2}}{\sum_{i=1}^3 \frac{1}{\sigma_{\varepsilon_i}^2}} \quad (3)$$

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
10 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.~~ Thus, it is used here to evaluate the ANN fit to the target values.

- 15 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 ~~36.26~~38.33 W m⁻² for LE, ~~34.82~~39.44 W m⁻² for H, and ~~2.20~~34 gC m⁻² day⁻¹ (~~or 123.16 PgC yr⁻¹~~) for GPP. LE has the best R² (0.95) comparing to the other ~~three~~two flux variables H (R²=0.89), and GPP (R²=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 = ~~11.43~~06 W m⁻²; for H, RMSD = ~~13.35~~13 W m⁻²; and for GPP,
20 RMSD=~~1.23~~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). ~~Tropical regions~~Wet tropical forests exhibit subtle seasonal variability in LE; ~~such as in the Amazon rainforest, Congo basin and Southeast Asia~~. 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-
30 evaporation and soil evaporation.

- 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.
35

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 ~~follow~~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 ~~are~~is observed in the Amazon, because of the strong variabilities in soil type, green up, biodiversity and ~~soil water availability~~soil water availability-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 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 global scales before. Therefore, we assess the information content of SIF observations in the WECANN retrievals by excluding them from the ANN inputs. We trained an ANN without SIF data on each of the three fluxes and evaluated the difference between the retrieved fluxes. Figure 6 shows the percentage difference maps between flux retrievals trained with SIF and without SIF for the year 2011 as well as the scatter plots with respect to target dataset. Including SIF decreases LE estimates in parts of Australia, Central Asia, Tibetan plateau and Southern Africa. In Indian peninsula, the Sahel, Eastern US and Southeastern Asia LE tends to reduce by adding the SIF information. Comparison of the ANN retrieval without SIF compared to the target data shows that the overall statistics of the ANN retrieval are comparable to WECANN retrievals, and inclusion of SIF slightly improves R^2 and lowers RMSD (Fig. 6d, and Table 3). Including SIF decreases H in the Sahel, Arabian Peninsula, Europe, Eastern US and in most of South America. In most of the other regions H is increased when SIF is added. The global ANN fits against the target H are relatively similar with and without SIF (Fig. 6e, Table 3).~~

~~Including SIF increases GPP in Central US, as well as in Europe, Northern India, and Southern Brazil capturing intense cropping regions. In Northern Canada, Central Asia, Australia, Southern Africa and the Tibetan Plateau GPP is strongly reduced by adding SIF into the ANN retrieval. In the Congo and Amazon, photosynthesis is slightly increased locally by the inclusion of SIF. Similarly to the other retrievals the global statistics of the retrieval with and without SIF compared to the target are relatively similar (Fig. 6f, and Table 3), hiding some of the changes in the spatial structure.~~

~~This comparison shows the significant role that SIF estimates play in the flux retrievals from WECANN. Given that GOME 2 instrument was originally designed to measure ozone in the atmosphere and not SIF, the future estimates of SIF from designated missions such as Fluorescence Explorer (FLEX) will have higher accuracy and finer spatial and temporal resolution (Kraft et al., 2012). Those SIF estimates will further enhance the retrievals of surface fluxes.~~

4.3.4.2 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 6 shows the comparisons for LE, and indicates that our product has higher a relatively similar R^2 with the three products ($R^2 = 0.96$ with FLUXNET-MTE (~~$R^2 = 0.96$~~) and ECMWF (~~$R^2 = 0.96$~~) than 94 with GLEAM (~~$R^2 = 0.94$~~). However, the scatterplot with FLUXNET-MTE is more concentrated and aligned along the 1:1 line, further emphasizing the consistency between the two datasets- Difference (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. 6a-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 ~~our product~~ 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. [S2S4](#) shows that FLUXNET-MTE has the smallest error in LE estimates globally compared to GLEAM and ECMWF, other than across India. WECANN retrievals also have better agreement with FLUXNET-MTE.

The differences in H estimates are more complex (Fig. 7). First, the R^2 between WECANN and the other datasets is always are slightly lower than for LE. ECMWF and FLUXNET-MTE again yield higher R^2 with WECANN (0.85 and 0.84, respectively) while GLEAM has an R^2 of 0.8087. GLEAM exhibits lower H in most of the Northern hemisphere, especially in seasonally dry regions, potentially due to its simple formulation of G-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. [S3S5](#)) are consistent with the comparison of WECANN and target datasets. Figure [S3S5](#) 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. 8). WECANN compares the best against FLUXNET-MTE ($R^2 = 0.9293$), with MODIS ($R^2 = 0.9091$) and ECMWF ($R^2 = 0.8790$) following. While ~~FLUXNET-MTE and MODIS~~ 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. [S4S6](#)) shows that FLUXNET-MTE has the lowest errors globally, while ECMWF has the largest errors in the Amazon.

4.43 Validation with FLUXNET Data

Direct validation of the WECANN fluxes is made more 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 $1^\circ \times 1^\circ$ 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 ~~several~~ a set of FLUXNET eddy-covariance towers. ~~However~~ 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 $1^\circ \times 1^\circ$ 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.

~~We compare the model outputs to eddy covariance towers from the FLUXNET 2015 database (tier 1 and tier 2, <http://fluxnet.fluxdata.org/data/fluxnet2015-dataset/>) spanning a large climatic and biome gradient (Fig. 10). Summary 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. The NEE is partitioned as the sum of Gross Primary Production (GPP) and Ecosystem Respiration (RECO) using one of two methods. The first method is based on the extrapolation of nighttime data (Reichstein et al., 2005), which is used to parameterize a respiration model that is then applied over the daytime NEE to estimate RECO. GPP is then calculated as the difference between RECO and NEE. The second method uses daytime data to parameterize a model of both GPP and RECO (Lasslop et al., 2010). The partitioning method used varies from site to site.~~

~~In addition to the FLUXNET 2015 dataset, we use data from the Large-scale Biosphere Atmosphere (LBA) experiment in Brazil. Specifically, we use data from sites near Santarem, Pará (Site code BR-Sa3), 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.~~

~~A summary of statistics across the different sites combining the FLUXNET 2015 tier 1 database is~~ 97 sites are provided in ~~Table 4-Table 6-Tables S1 – S3~~. Overall, WECANN performs better than the alternative global products. In particular, WECANN has the highest correlation for 6476% of sites for LE, 6054% of sites for H, and 5653% 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 5671% of sites for LE, 5046% of sites for H, and 4451% 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. ~~Relatively similar results are obtained at the wet Santarem site, Sa3, where both WECANN and FLUXNET-MTE perform well in reproducing all fluxes. ECMWF and MODIS show the incorrect seasonality of the fluxes at the site, as GPP at the site reflects subtle leaf aging and flushing (Lopes et al., 2016; Saleska et al., 2016; Wehr et al., 2016), and~~

~~radiation structure not captured by those models (Anber et al., 2015; Morton et al., 2014, 2016).~~ 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.

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.

- 5 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.

10 **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.

- 15 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.

- 20 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 ($90^\circ \text{ N} - 60^\circ \text{ N}$), Northern Hemisphere (NH) mid-latitude ($60^\circ \text{ N} - 10^\circ \text{ N}$), Tropics ($10^\circ \text{ N} - 15^\circ \text{ S}$), and Southern Hemisphere (SH) mid-latitude ($15^\circ \text{ S} - 60^\circ \text{ S}$). 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.

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

- 30 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 Impact of SIF on the retrieval of surface fluxes

- 35 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 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). 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 ~~target output~~ training datasets show that our retrieval ~~is best correlated~~ has similar correlation with the three products while it has the smallest RMSD with FLUXNET-MTE for LE ($R^2=0.96$ $RMSD=6.42 \text{ W m}^{-2}$), H ($R^2=0$ $RMSD=7.84 \text{ W m}^{-2}$) and GPP (R^2 $RMSD=0.9288 \text{ gC m}^{-2} \text{ day}^{-1}$), 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. S2S4 – Fig. S4S6), 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. ~~The difference maps between neural network outputs trained with SIF and without SIF demonstrate that SIF has high influence on all three flux retrievals (Fig. 6).~~ 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.

5 Finally, 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 ~~6476%~~ 6054% of sites for LE, ~~6054%~~ 5653% of sites for H, and ~~5653%~~ 5653% 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.

10 Data Availability

~~WECANN product is available for free upon request. Please contact the corresponding author to request access.~~

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

15 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

20 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 Boussetta, 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

25 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,

30 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,

35 FAO-GTOS-TCO, iLEAPS, Max Planck Institute for Biogeochemistry, National Science Foundation, University of Tuscia,

References

- 5 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), ~~n/a-n/a~~, doi:10.1029/2011JD017188, 2012.
- 10 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.
- 15 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.
- 20 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., Luysaert, S., Margolis, H., Oleson, K. W., Rouspard, 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.
- 30 ~~Brodzik, M. and Armstrong, R.: Northern Hemisphere EASE Grid 2.0 Weekly Snow Cover and Sea Ice Extent, Version 4., doi:http://dx.doi.org/10.5067/P7O0HGJLYUQU, 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, 141–154, doi:10.1016/j.rse.2013.12.005, 2014.
- 35 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.
- Dorigo, W.: ESA CCI Soil Moisture for improved Earth system understanding: state-of-the art and future directions, *Remote Sens. Environ.*, n.d.
- 5 [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.
- 10 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.
- 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
- 15 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 CO₂ 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
- 20 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.
- 25 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
- 30 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.
- 35 Guillod, B. P., Orlowsky, B., Miralles, D., Teuling, A. J., Blanken, P. D., Buchmann, N., Ciais, P., Ek, M., Findell, K. L., Gentile, 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.
- 40 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.
- 5 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:10.1175/1525-10 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.
- 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.
- 15 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, 20 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., ~~Yoshida, Y.~~ Guanter, L., Lindstrot, R., Voigt, M., Vasilkov, A. P., ~~Yoshida, Y., Corp, L. A. and~~ Middleton, E. M.: ~~First observations of global and seasonal~~ M., Huemmrich, K. F., Yoshida, Y. and Frankenberg, C.: Global monitoring of terrestrial chlorophyll fluorescence from ~~space, *Biogeosciences*, 8(3), 637–654~~ moderate-spectral-resolution near-infrared satellite measurements: methodology, simulations, and application to GOME-2, *Atmos. Meas. Tech.*, 6(10), 2803–2823, doi:10.5194/bg-8-637-2011, 2014 amt-6-2803-2013, 2013.
- 25 Joiner, J., ~~Guanter, L., Lindstrot, R., Voigt, M., Yoshida, Y.,~~ Vasilkov, A. P., ~~Schaefer, K., Jung, M., Guanter, L., Zhang, Y., Garrity, S.,~~ Middleton, E. M., Huemmrich, K. F., ~~Yoshida, Y.~~ Gu, L. and Frankenberg, C.: Global monitoring Belelli Marchesini, L.: The seasonal cycle of ~~terrestrial 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.
- 30 Joiner, J., Yoshida, Y., Guanter, L. and Middleton, E. M.: New methods for the retrieval of chlorophyll red fluorescence from moderate-spectral-resolution near-infrared hyperspectral satellite measurements: methodology, instruments: simulations, and application to GOME-2 and SCIAMACHY, *Atmos. Meas. Tech.*, 6(10), 2803–2823 (8), 3939–3967, doi:10.5194/amt-6-2803-2013, 2013 39-3939-2016, 2016.
- 35 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- 40 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., Rouspard, 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.
- ~~Kraft, S., Del Bello, U., Bouvet, M., Drusch, M. and Moreno, J.: FLEX: ESA's Earth Explorer 8 candidate mission, in 2012 IEEE International Geoscience and Remote Sensing Symposium, pp. 7125–7128, IEEE., 2012.~~
- 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.
- ~~Lasslop, G., Reichstein, M., Papale, D., Richardson, A. D., Arneth, A., Barr, A., Stoy, P. C. and Wohlfahrt, G.: Separation of net ecosystem exchange into assimilation and respiration using a light response curve approach: critical issues and global evaluation, *Glob. Chang. Biol.*, 16(1), 187–208, doi:10.1111/j.1365-2486.2009.02041.x, 2010.~~
- 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, 5 2009.
- ~~Lopes, A. P., Nelson, B. W., Wu, J., Graça, P. M. L. de A., Tavares, J. V., Prohaska, N., Martins, G. A. and Saleska, S. R.: Leaf flush drives dry season green-up of the Central Amazon, *Remote Sens. Environ.*, 182, 90–98, doi:10.1016/j.rse.2016.05.009, 2016.~~
- 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.
- 10 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., Fernández-Prieto, D., Beck, H. E., Dorigo, W. A. and Verhoest, N. E. C.: GLEAM v3: satellite-based land evaporation and root-zone soil moisture, *Geosci. Model Dev. Discuss.*, 0, 15 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 20 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 25 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 30 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 35 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 40 combined soil desiccation and atmospheric heat accumulation, *Nat. Geosci.*, 7(5), 345–349, doi:10.1038/ngeo2141, 2014b.

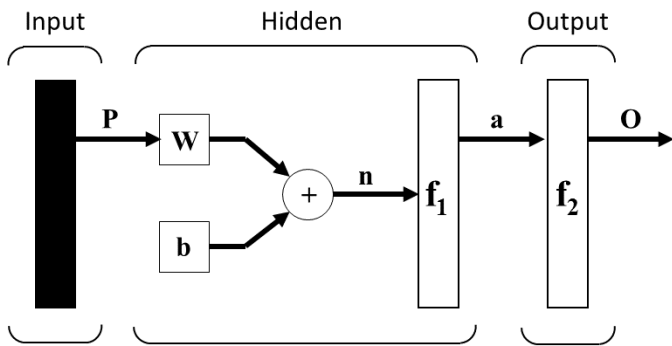
- Miralles, D. G., Jiménez, C., Jung, M., Michel, D., Ershadi, A., McCabe, M. F., Hirschi, M., Martens, B., Dolman, A. J., Fisher, J. B., Mu, Q., Seneviratne, S. I., Wood, E. F. and Fernández-Prieto, D.: The WACMOS-ET project - Part 2: Evaluation of global terrestrial evaporation data sets, *Hydrol. 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 ~~[and Discussion]~~, *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.
- [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.
- ~~Reichstein, M., Falge, E., Baldocchi, D., Papale, D., Aubinet, M., Berbigier, P., Bernhofer, C., Buchmann, N., Gilmanov, T., Granier, A., Grunwald, T., Havrankova, K., Iivesniemi, H., Janous, D., Knohl, A., Laurila, T., Lohila, A., Loustau, D., Matteucci, G., Meyers, T., Miglietta, F., Ourcival, J. M., Pumpanen, J., Rambal, S., Rotenberg, E., Sanz, M., Tenhunen, J., Seufert, G., Vaccari, F., Vesala, T., Yakir, D. and Valentini, R.: On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm, *Glob. Chang. Biol.*, 11(9), 1424–1439, doi:10.1111/j.1365-2486.2005.001002.x, 2005.~~
- 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.
- 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.
- 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/isprsnals-I-7-315-2012, 2012.
- ~~Wehr, R., Munger, J. W., McManus, J. B., Nelson, D. D., Zahniser, M., Walther, S., Davidson, E., Voigt, M., Thum, T., Gonsamo, A., Wofsy, S., C. Zhang, Y., Koehler, P., Jung, M., Varlagin, A. and Saleska, S. R.: Seasonality~~ Guanter, L.: Satellite chlorophyll fluorescence measurements reveal large-scale decoupling of temperate forest photosynthesis and daytime respiration, *Nature*, 534(7609), 680–683 greenness dynamics in boreal evergreen forests, *Glob. Chang. Biol.*, 22, 2979–2996, doi:10.1038/nature17966 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.

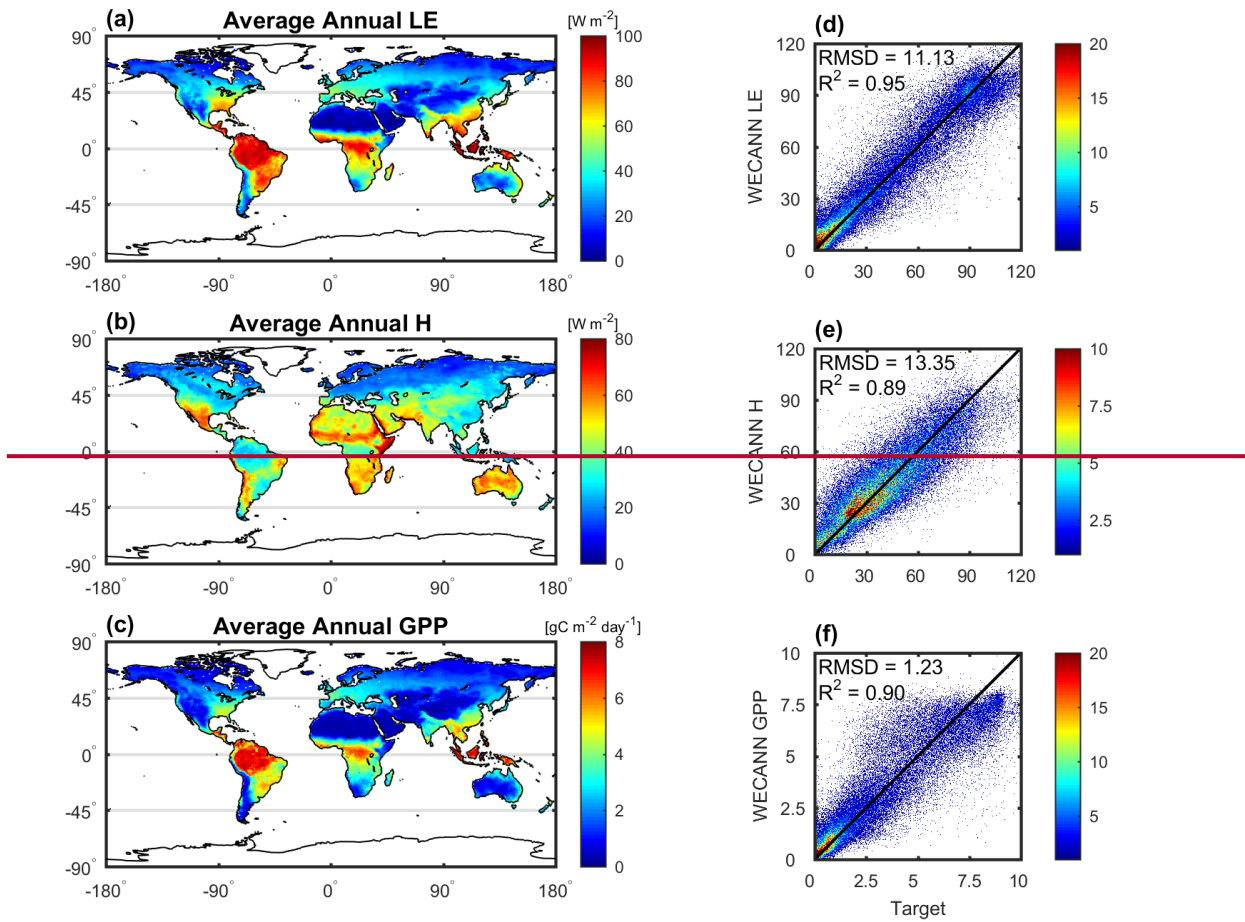
Zhao, M. and Running, S. W.: Drought-Induced Reduction in Global Terrestrial Net Primary Production from 2000 Through 2009, *Science* (80-.), 329(5994), 940–943, doi:10.1126/science.1192666, 2010.

- 5 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.



$$O = f_1(f_2(WP + b))$$

5 **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.**



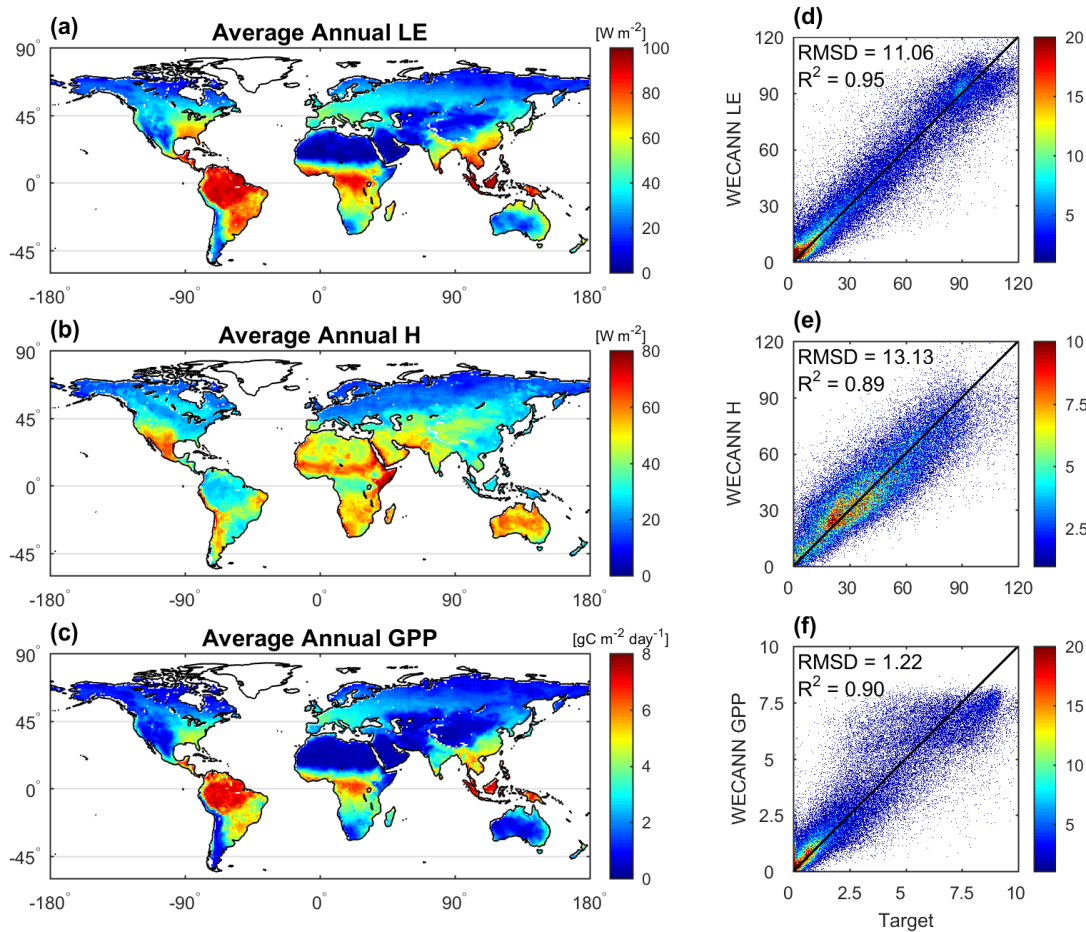
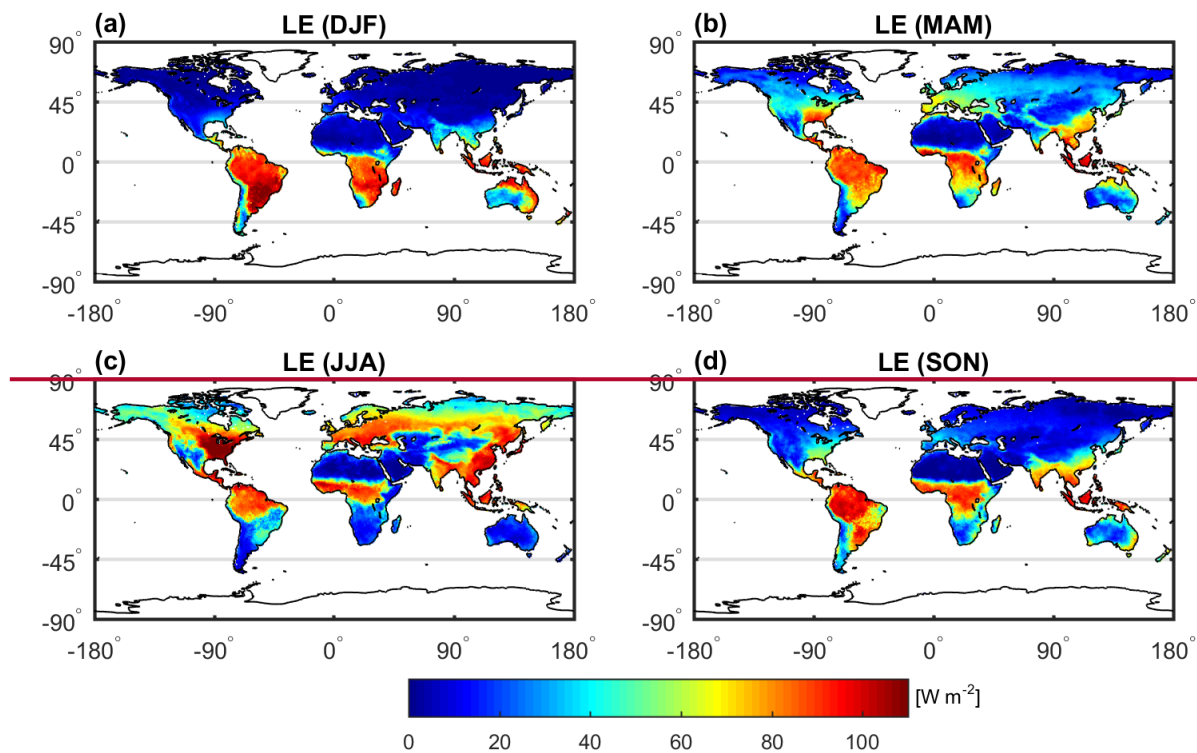


Figure 2: Left column: Annual average surface fluxes in 2011 for (a) LE, (b) H, and (c) GPP. Right column: Density scatterplot between estimates of ANN and target data for (d) LE, (e) H, and (f) GPP during the validation period (2011). The density of scatter points is represented by the shading color.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



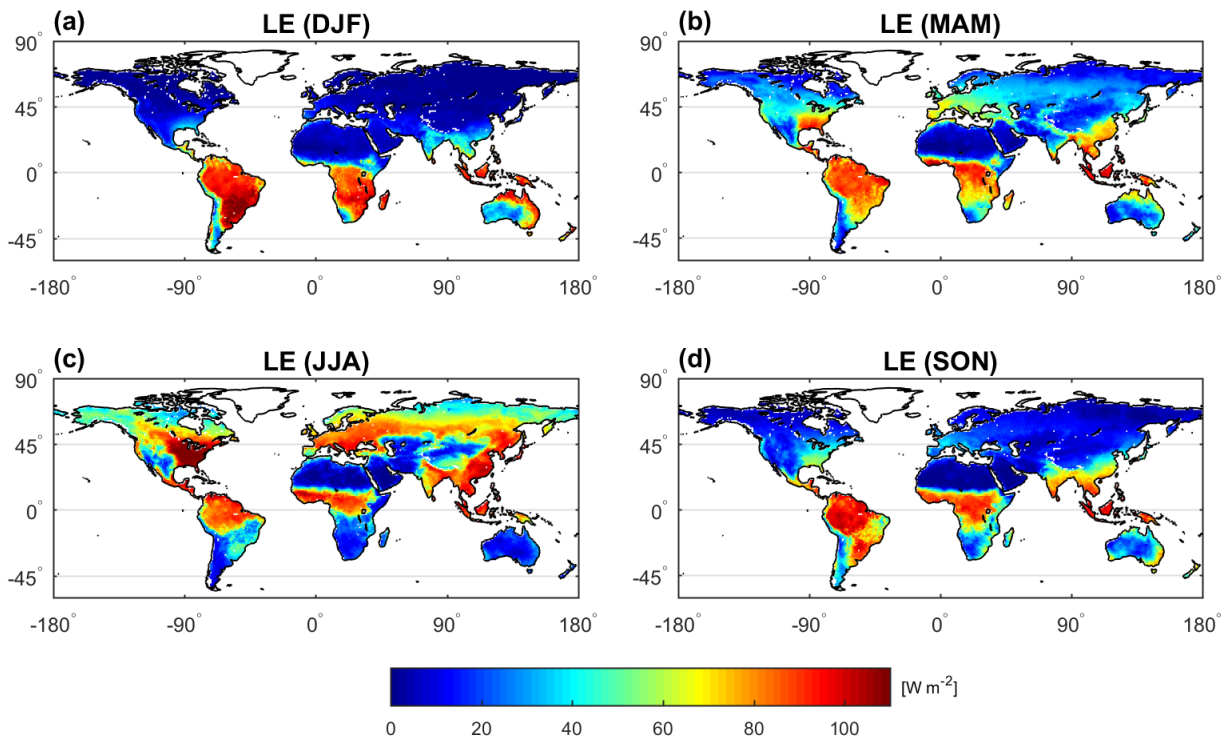
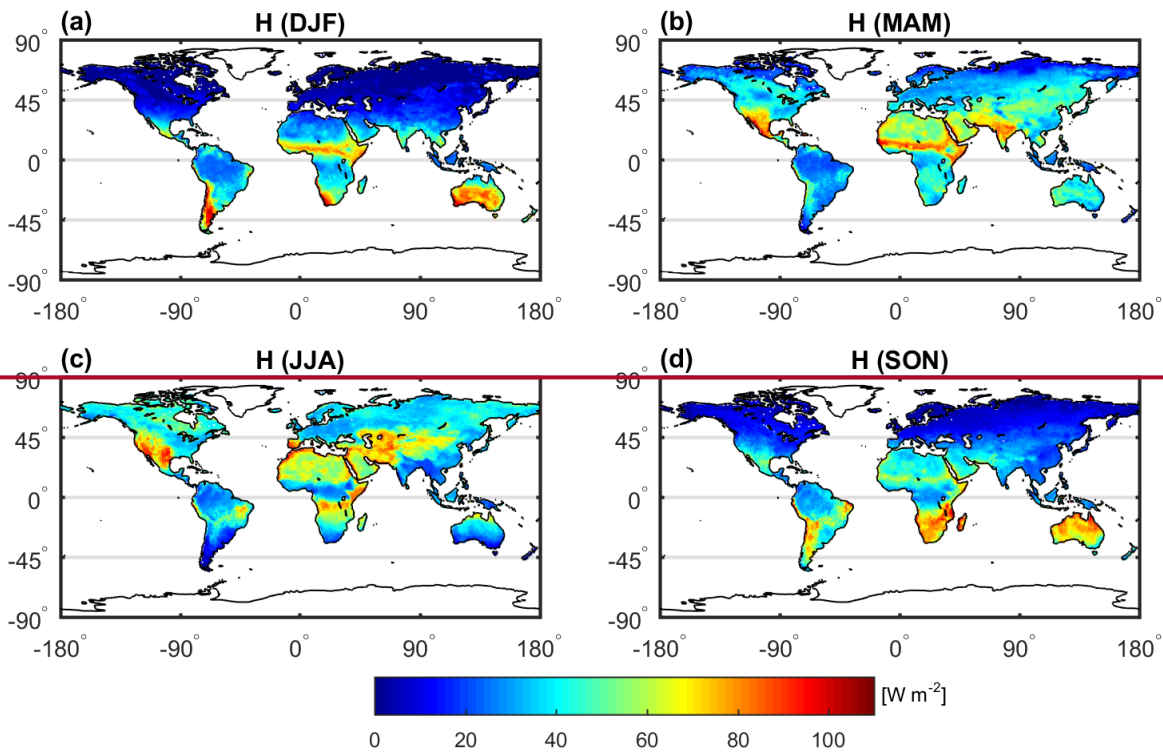


Figure 3: Global patterns of seasonal average LE from WECANN in 2011, (a) December - February, (b) March - May, (c) June - August, and (d) September - November.



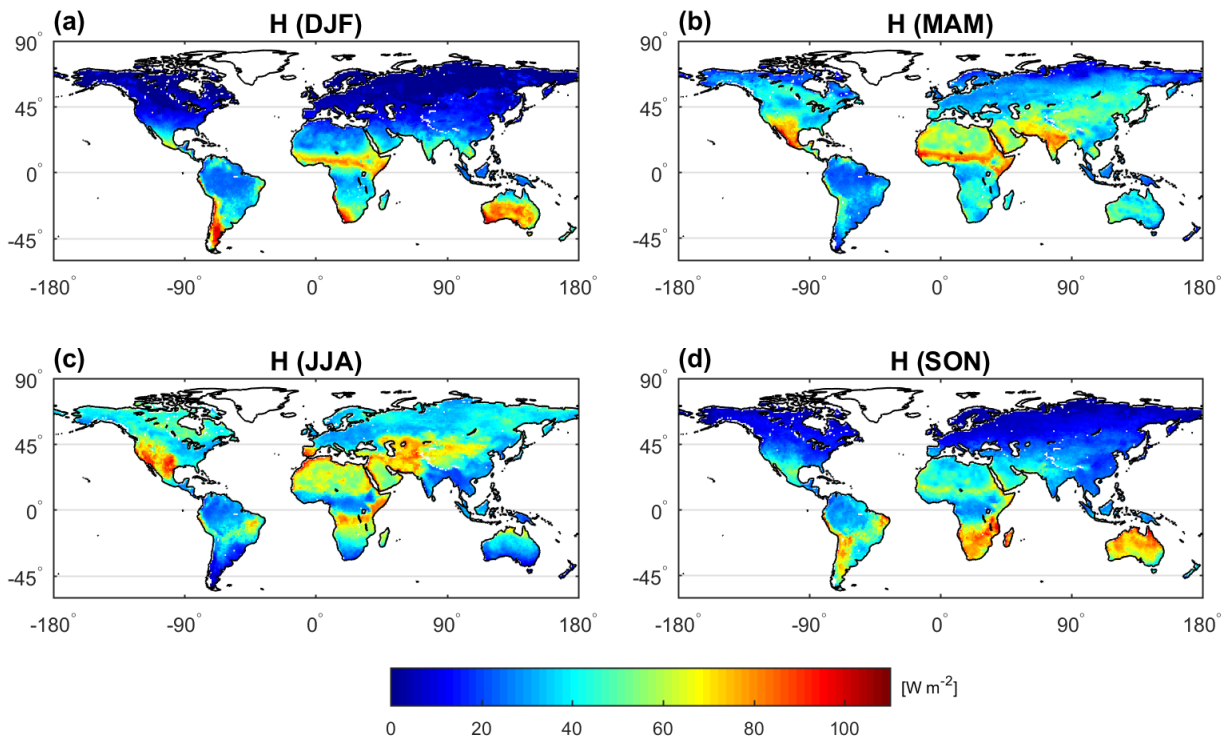
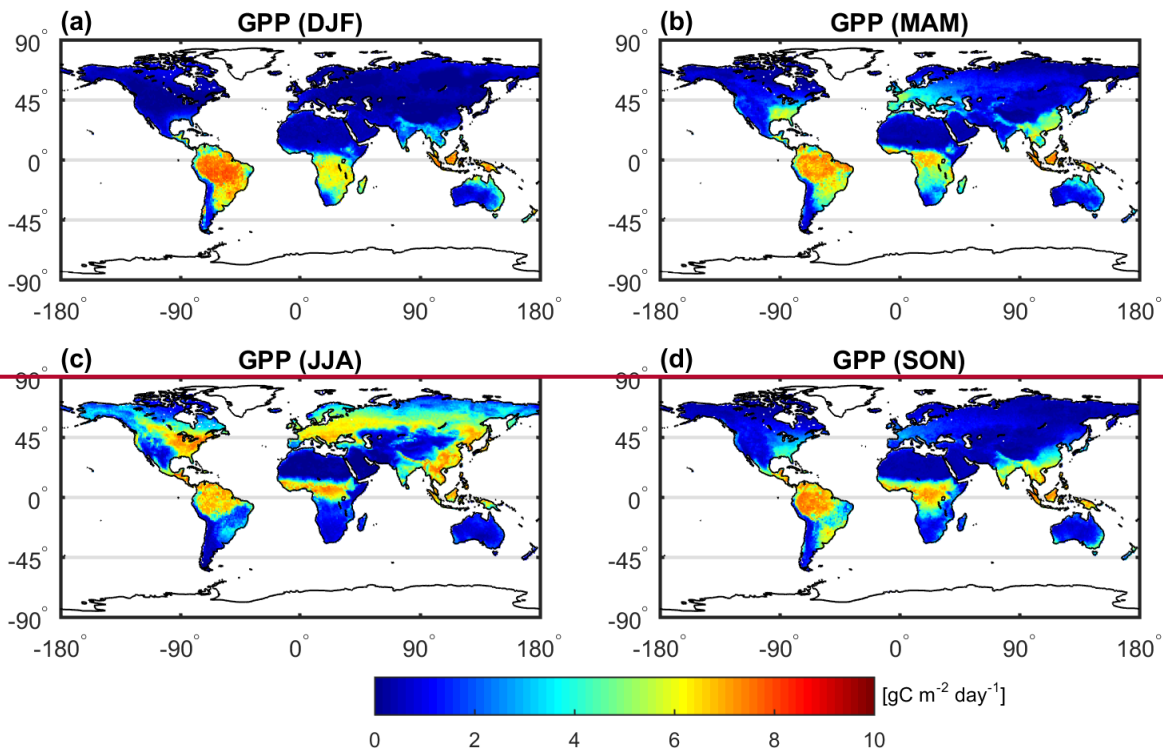


Figure 4: Similar to Figure 3 but for H instead of LE



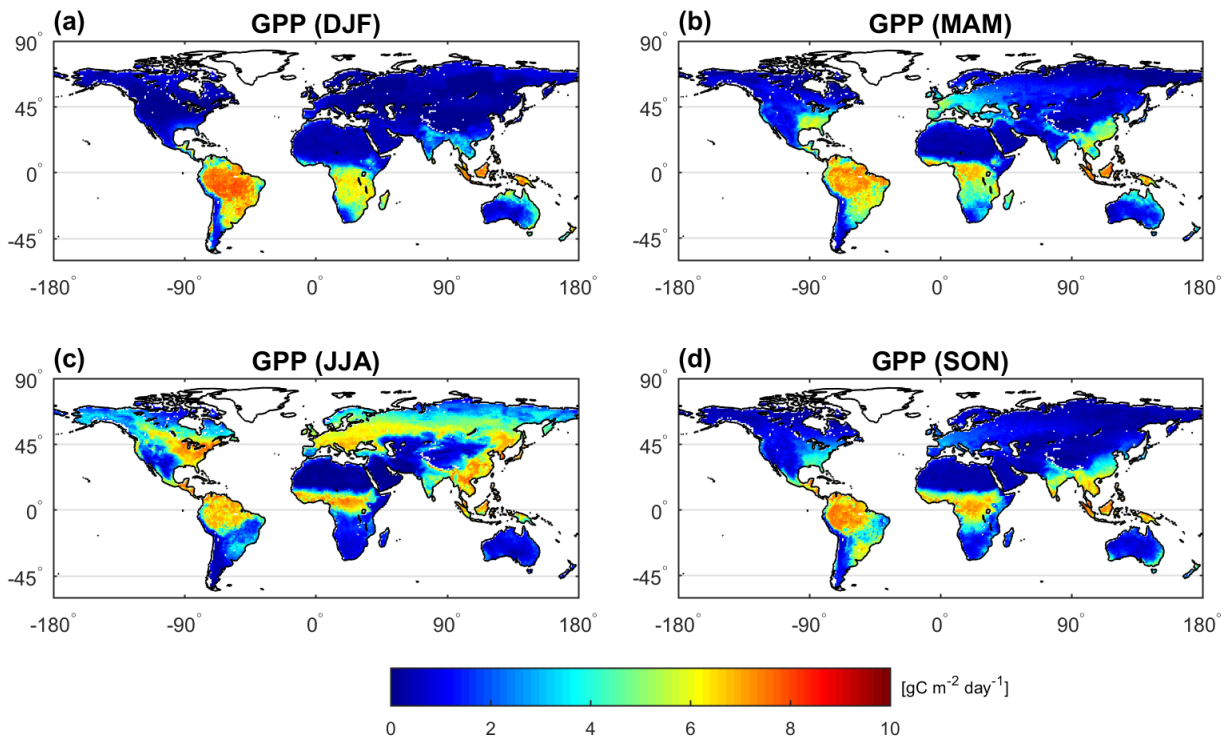
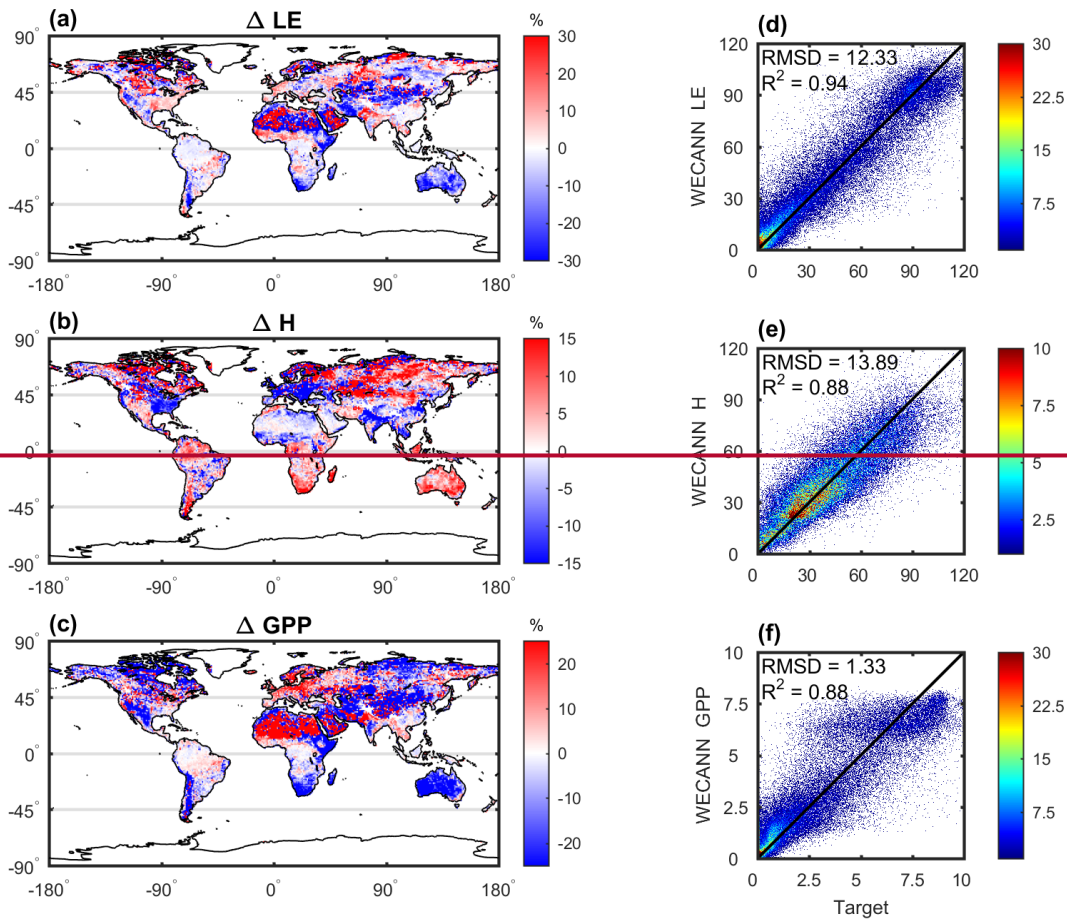


Figure 5: Similar to Figure 3 but for GPP instead of LE



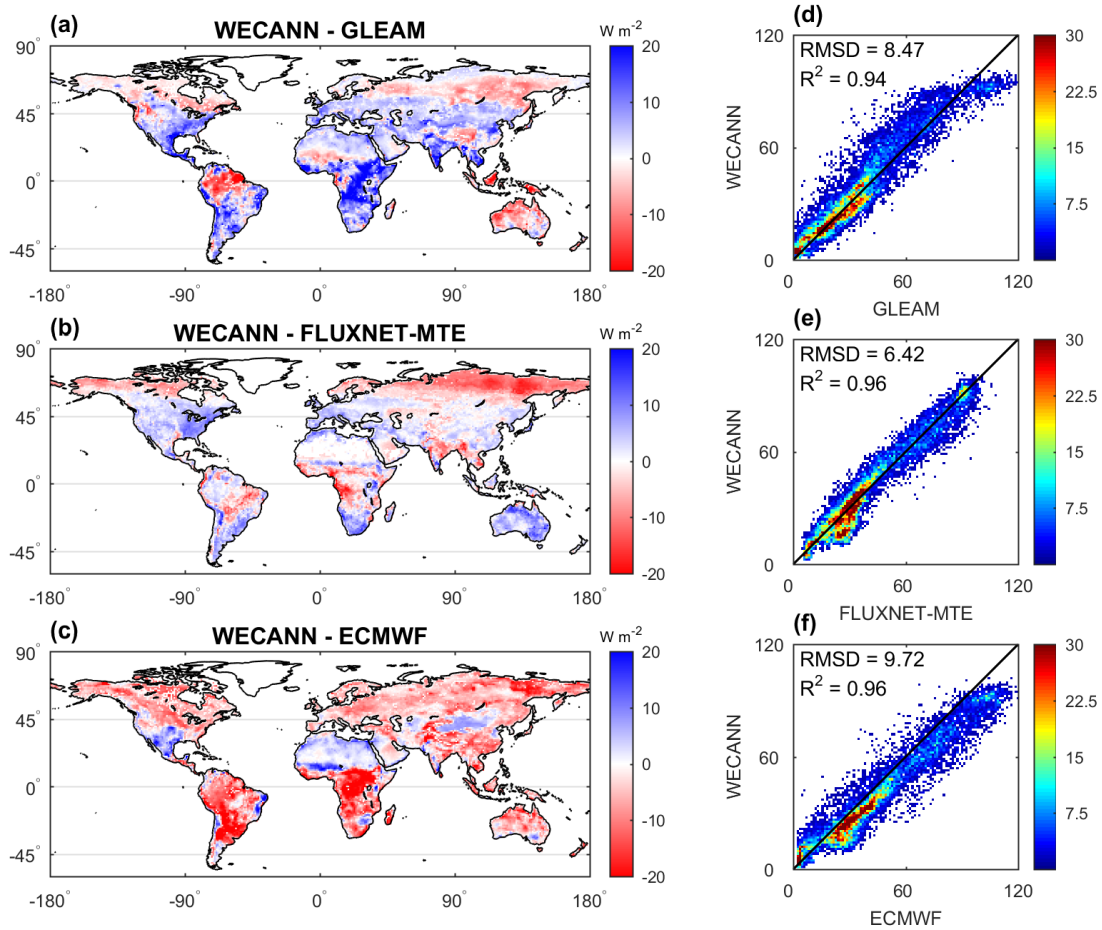


Figure 6: Comparison of WECANN retrievals with the retrievals from an ANN without using SIF data. (a)–(c) shows the WECANN retrieval minus the retrieval without SIF normalized by the WECANN retrievals for LE, H, and GPP during 2011, respectively. (d)–(f) show the scatter plots of WECANN retrievals vs target data.

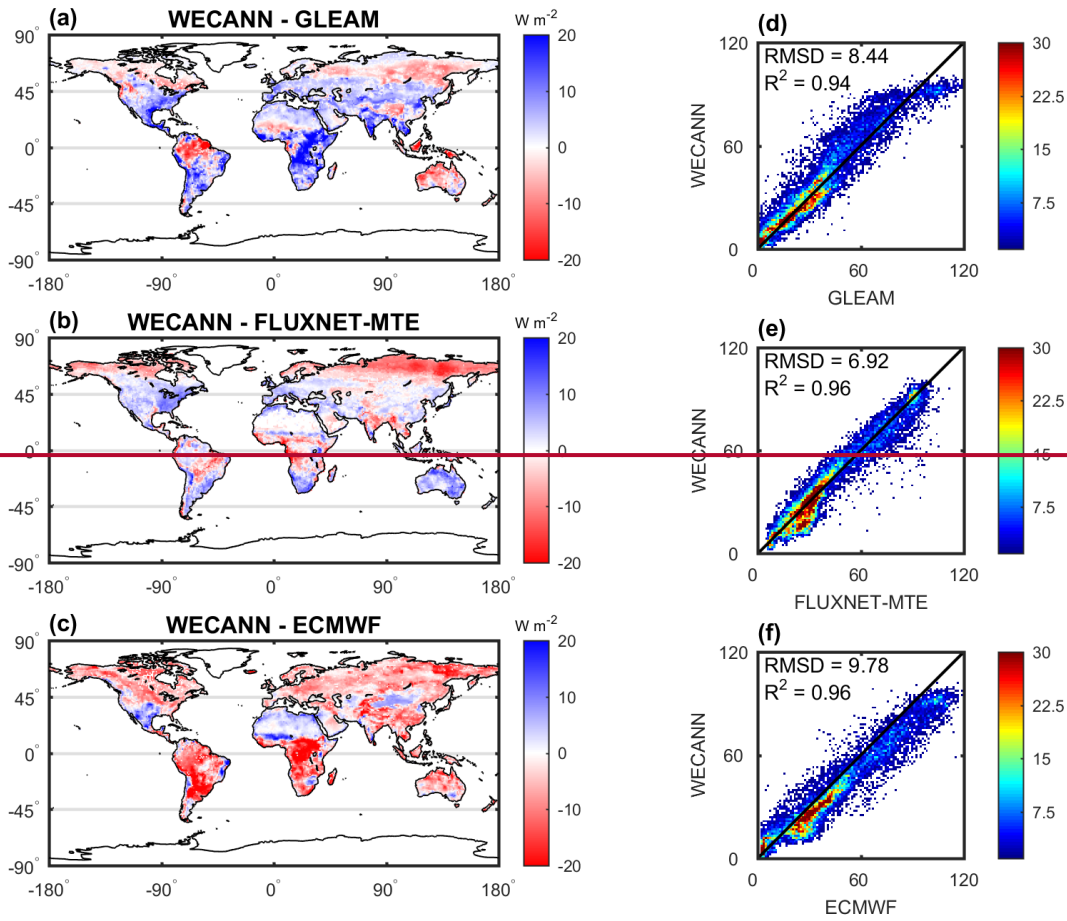
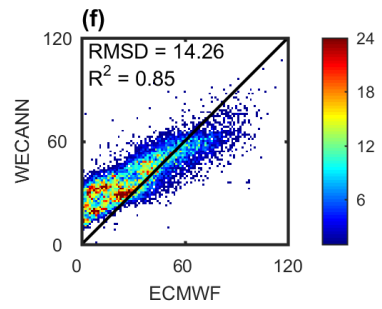
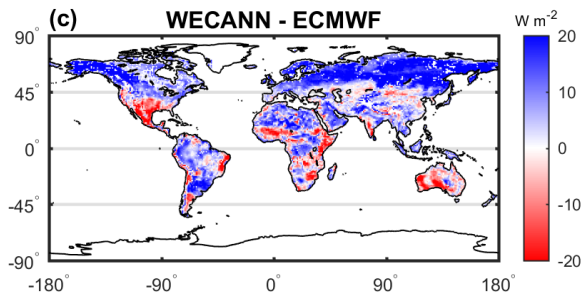
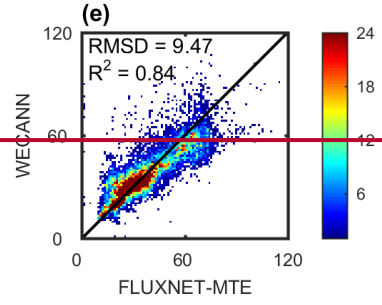
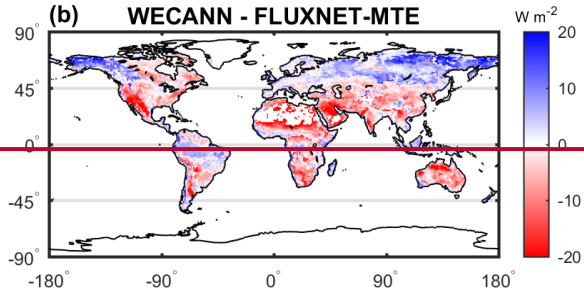
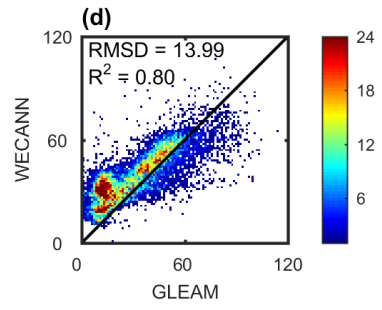
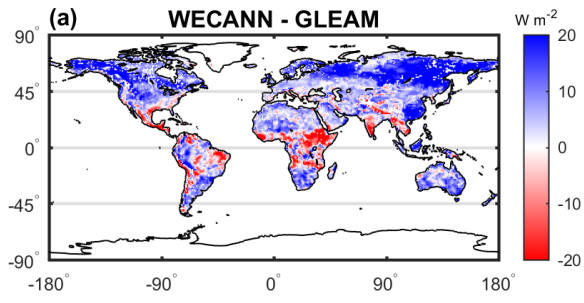


Figure 7: Difference between annual mean LE retrieved by WECANN and the three target datasets (a-c). Scatter plots of LE retrieved from WECANN vs. from each of the target datasets (d-f). Data used are from 2011.



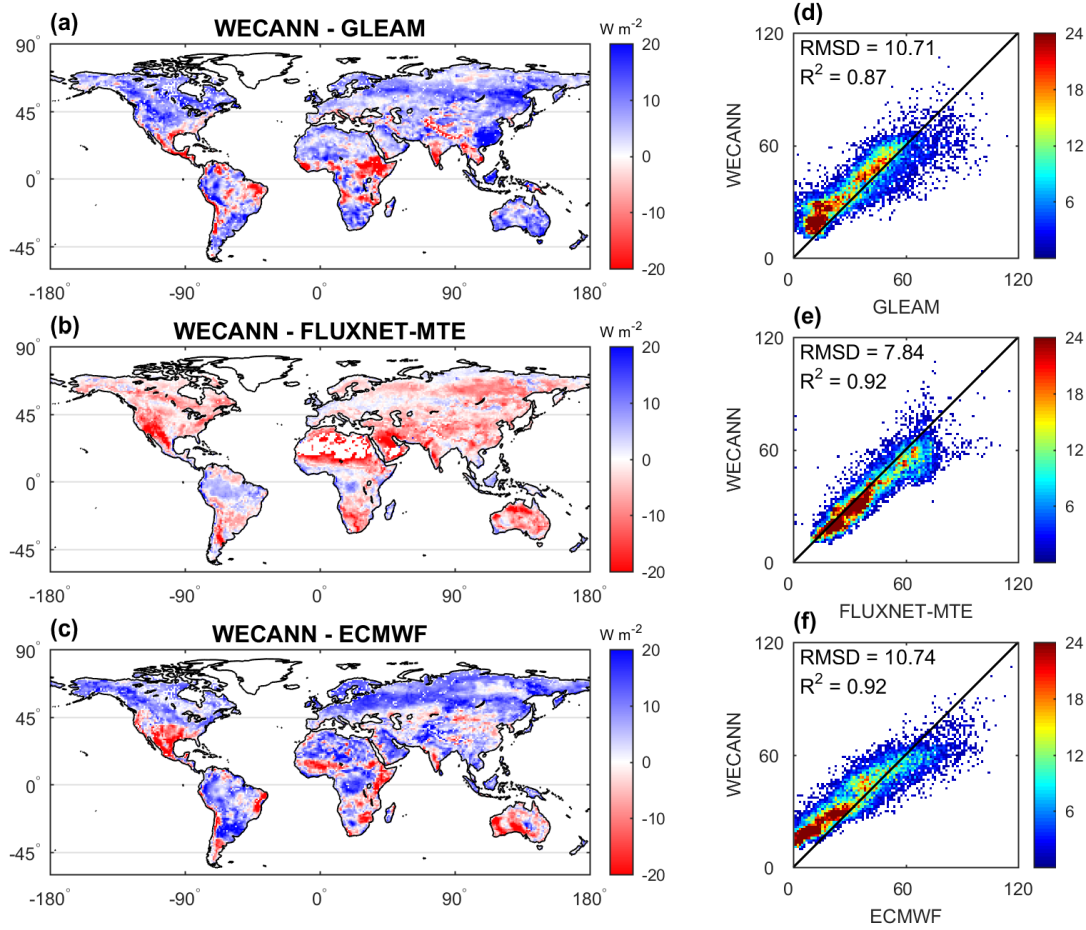


Figure 7: Similar to Figure 6 but for H instead of LE

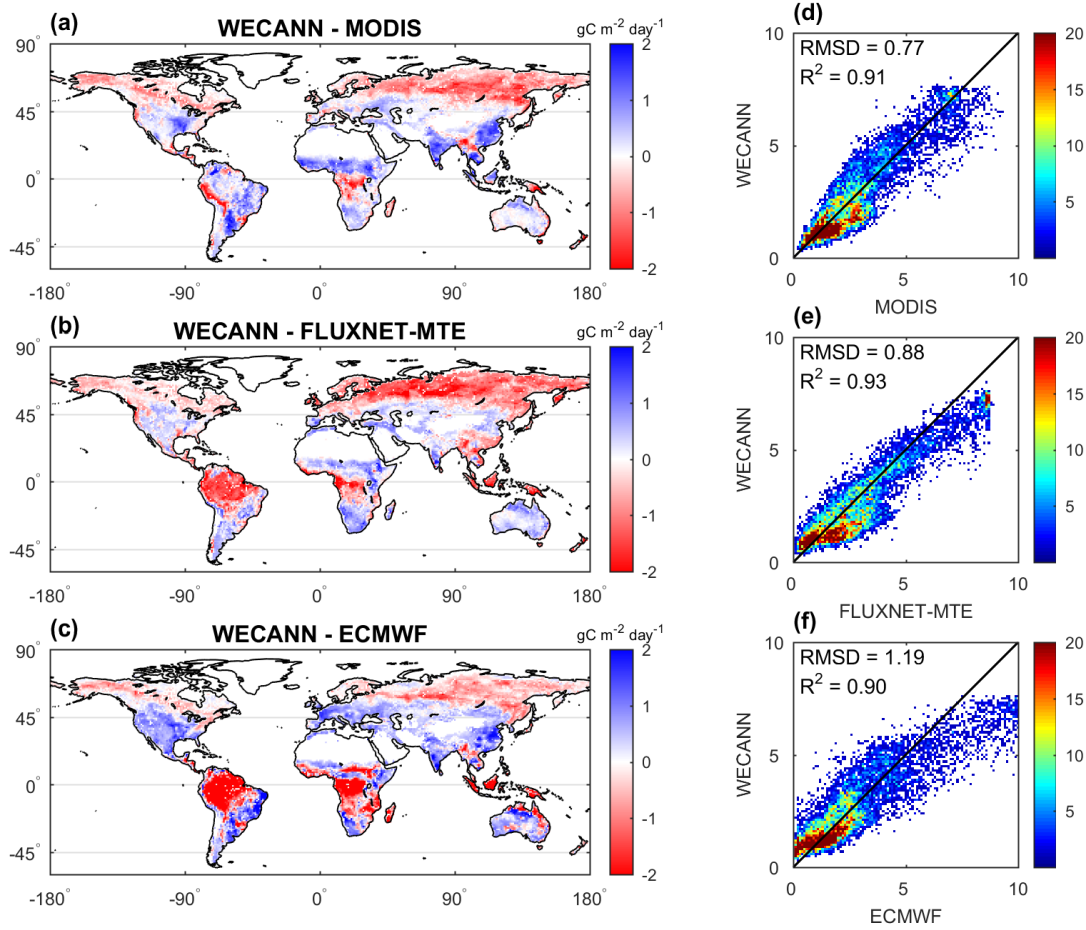


Figure 8: Similar to Figure 6 but for H instead of LE

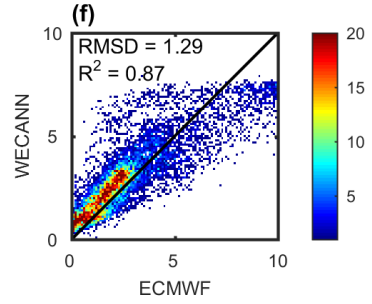
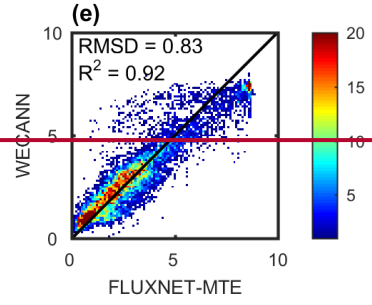
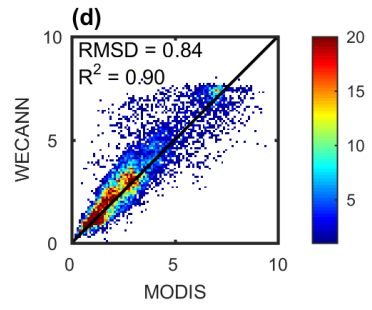
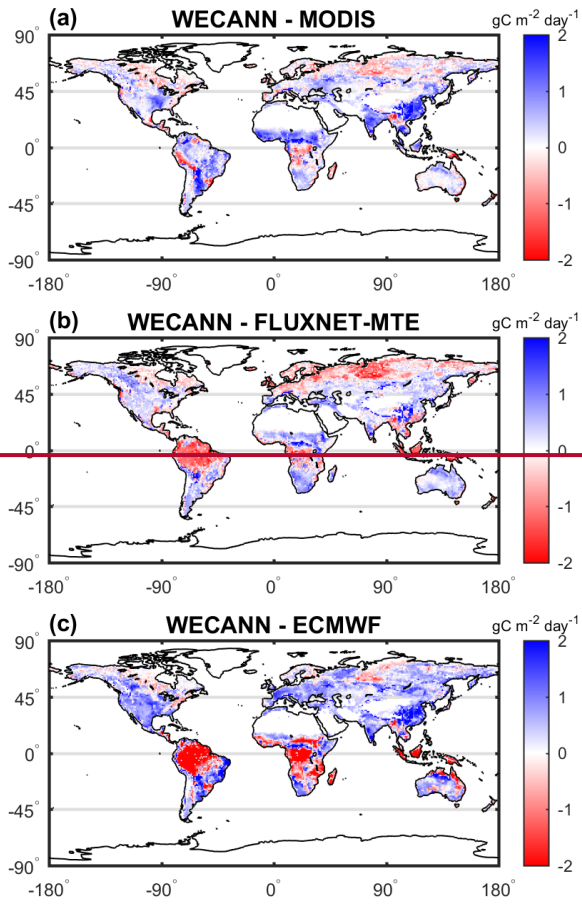
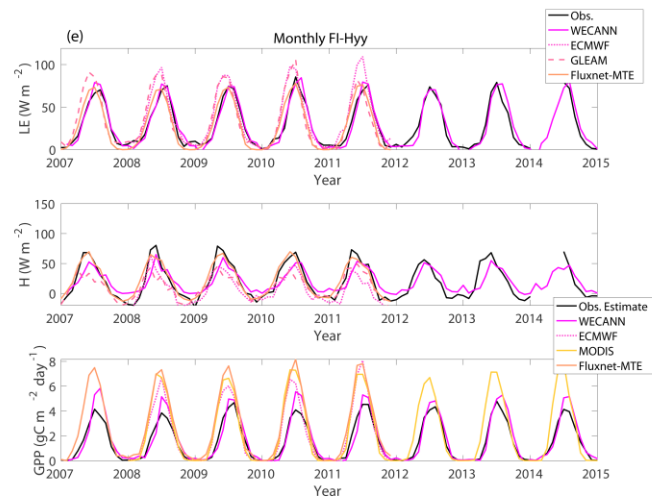
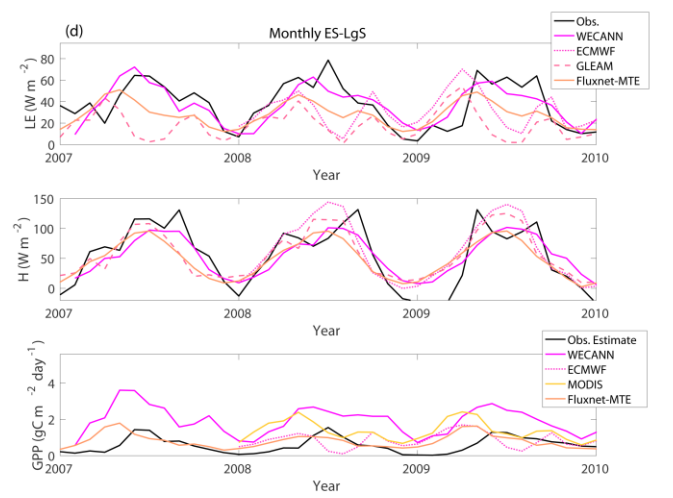
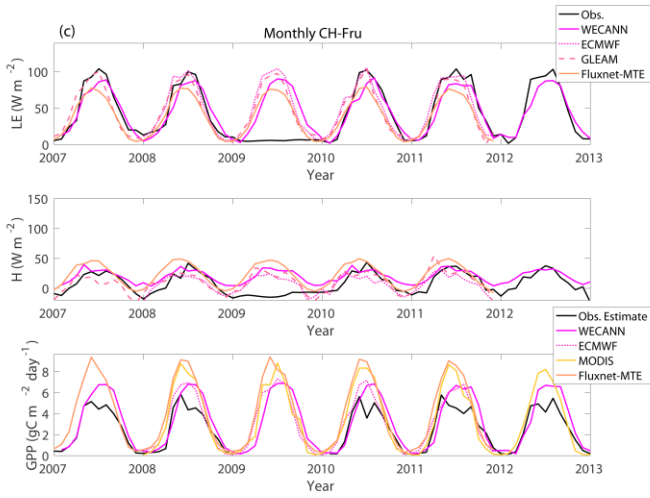
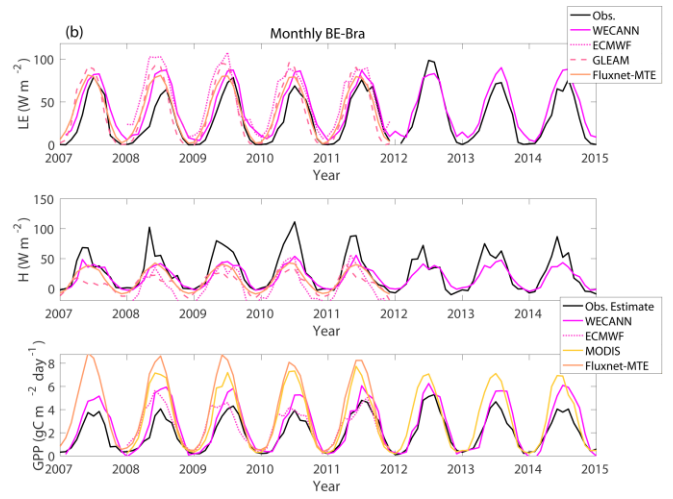
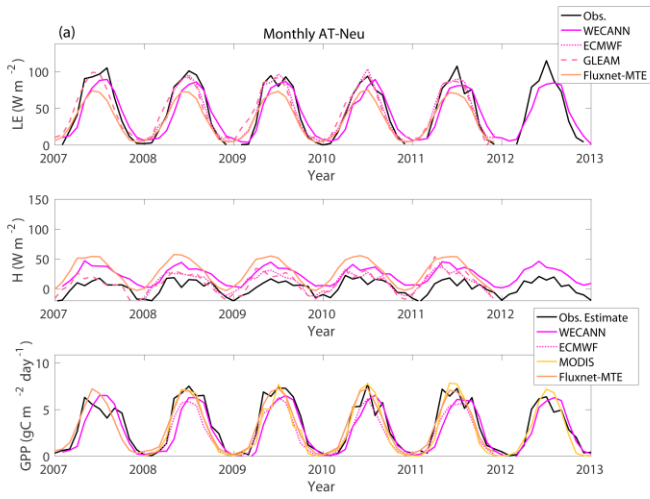


Figure 9: Similar to Figure 7 but for GPP instead of LE



Locations of Flux towers

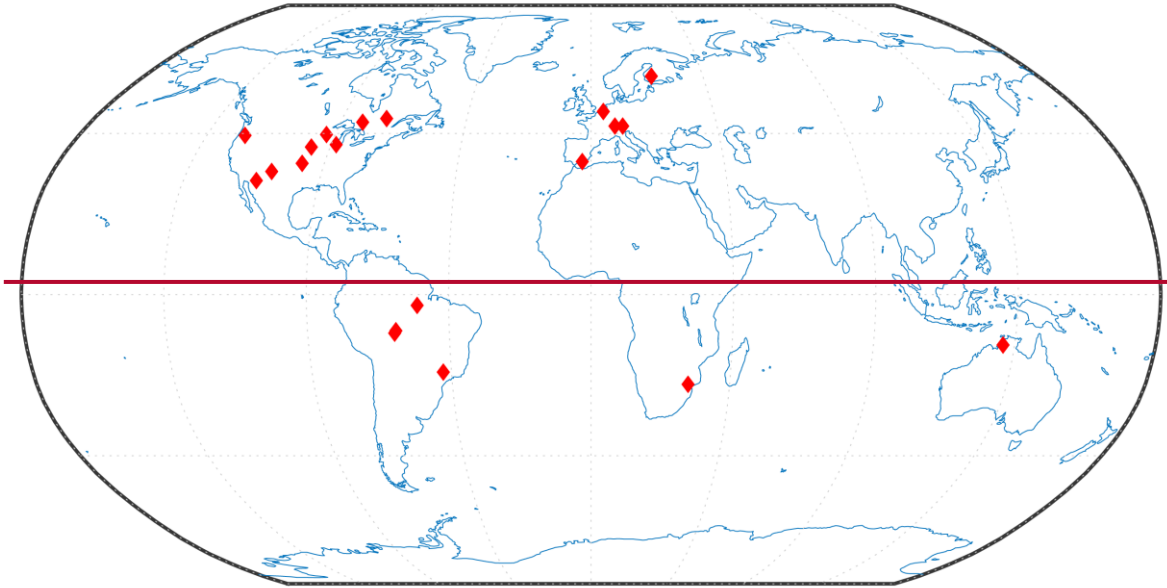


Figure 9: Geographical positions of the eddy covariance sites used for comparison of the flux retrievals.

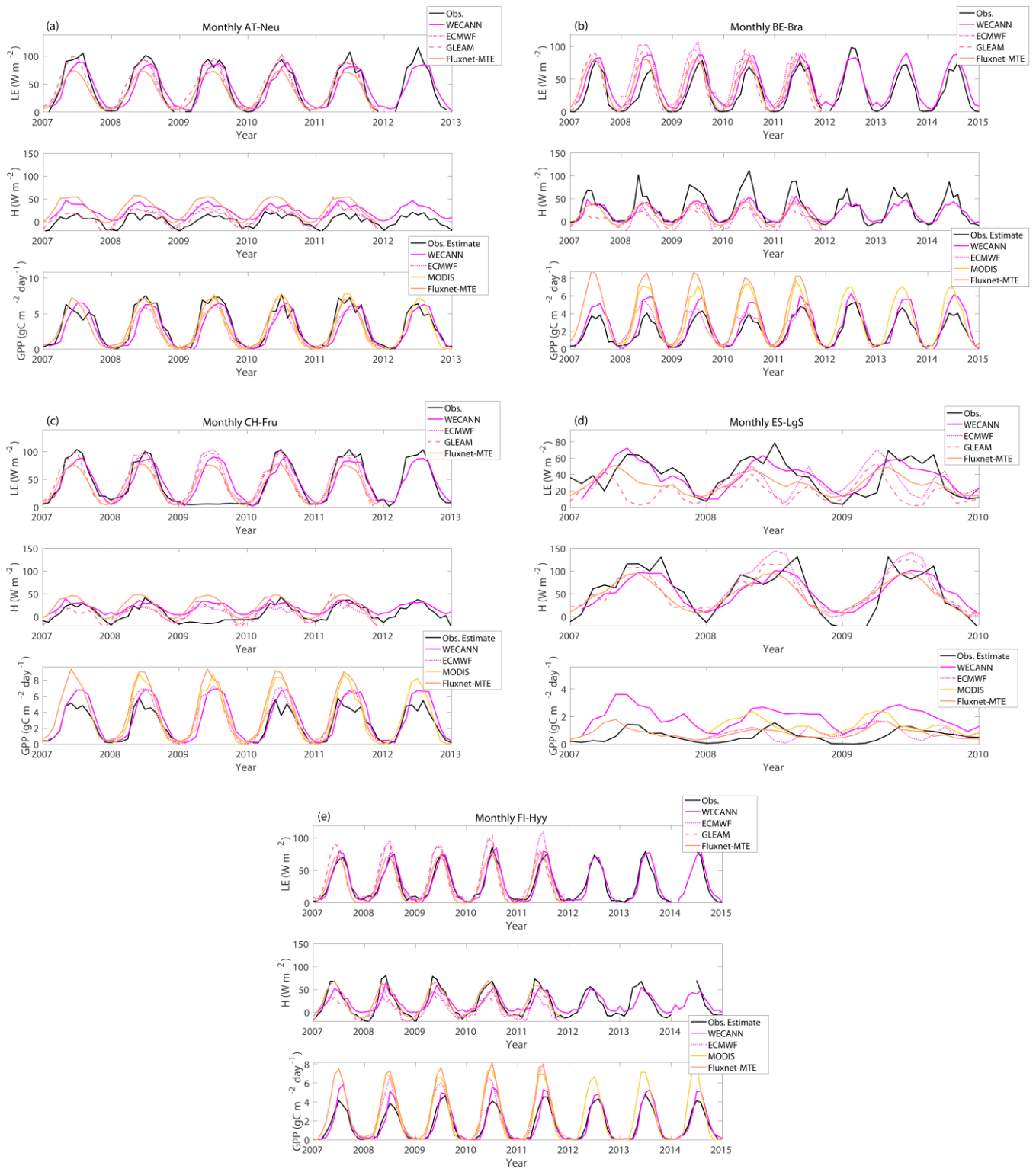
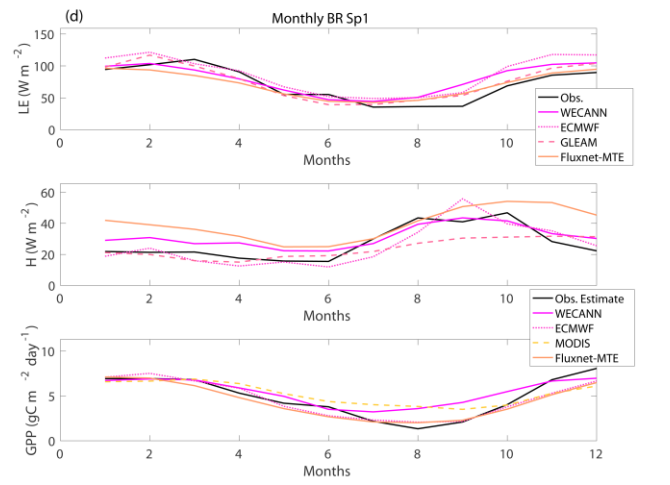
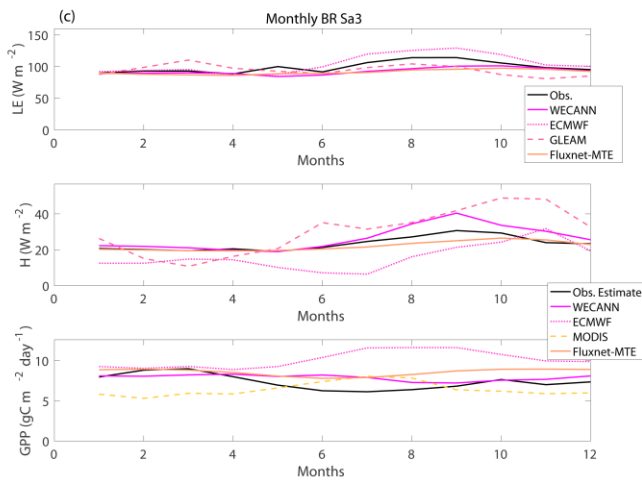
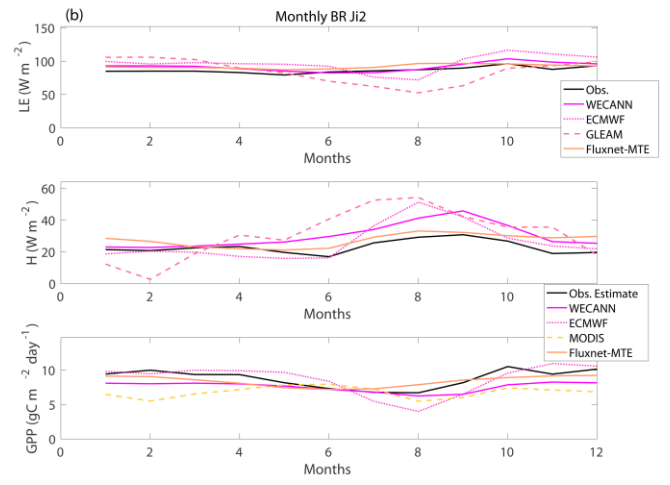
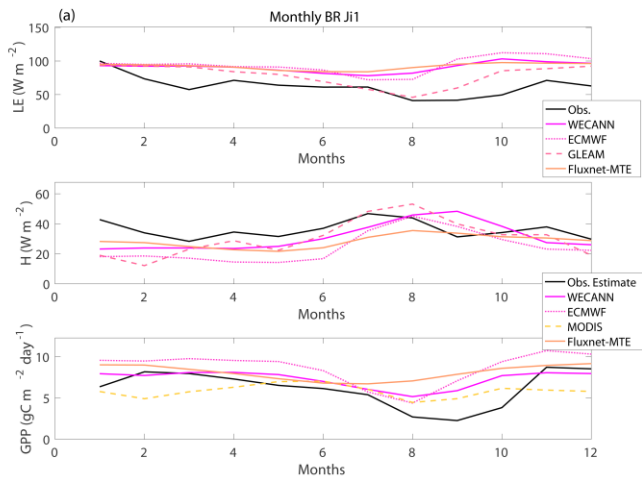


Figure 11: Comparison of the flux retrievals with eddy covariance observations of LE, H and GPP across European sites (a) AT-Neu site, Austria, (b) BE-Bra site, Belgium, (c) CH-Fru site, Switzerland, (d) ES-LgS site, Spain, and (e) FI-Hyy site, Finland



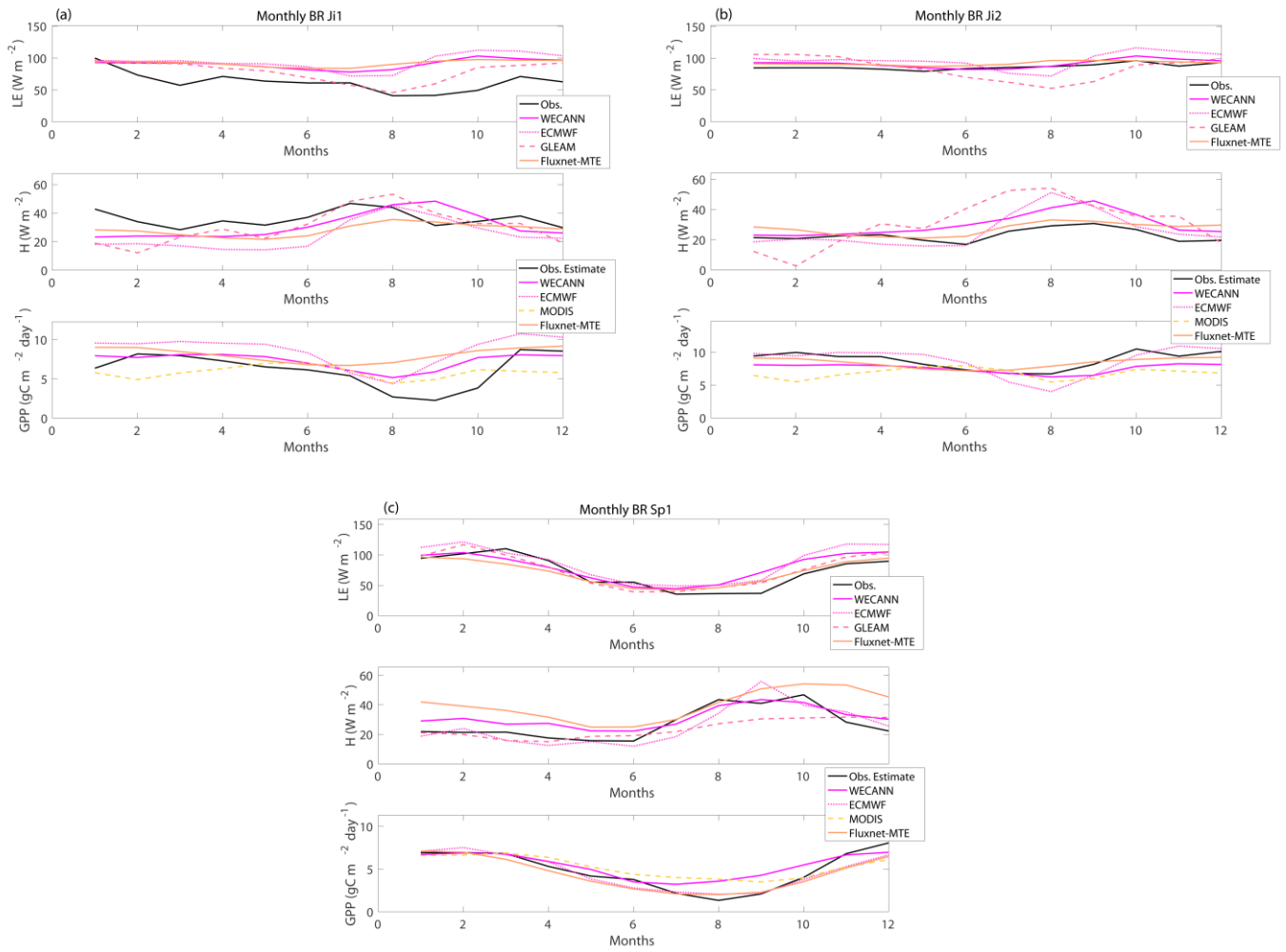


Figure 10: Same as Figure 9 but for Brazilian sites (a) BR-Ji1, (b) BR-Ji2, (c) BR-Sa3, and (d) BR-Sp1.

5

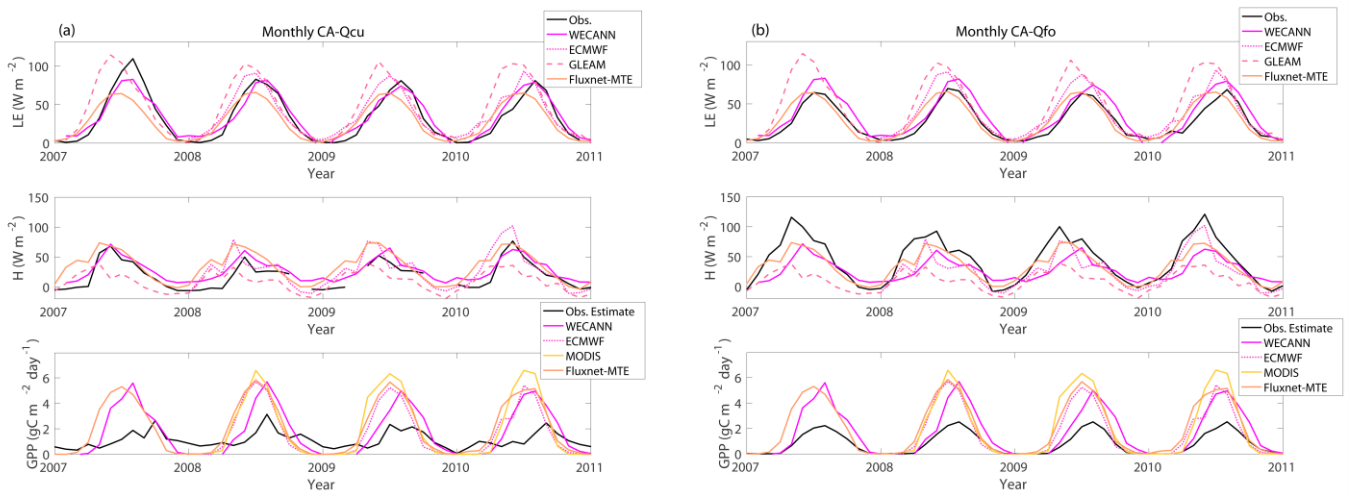
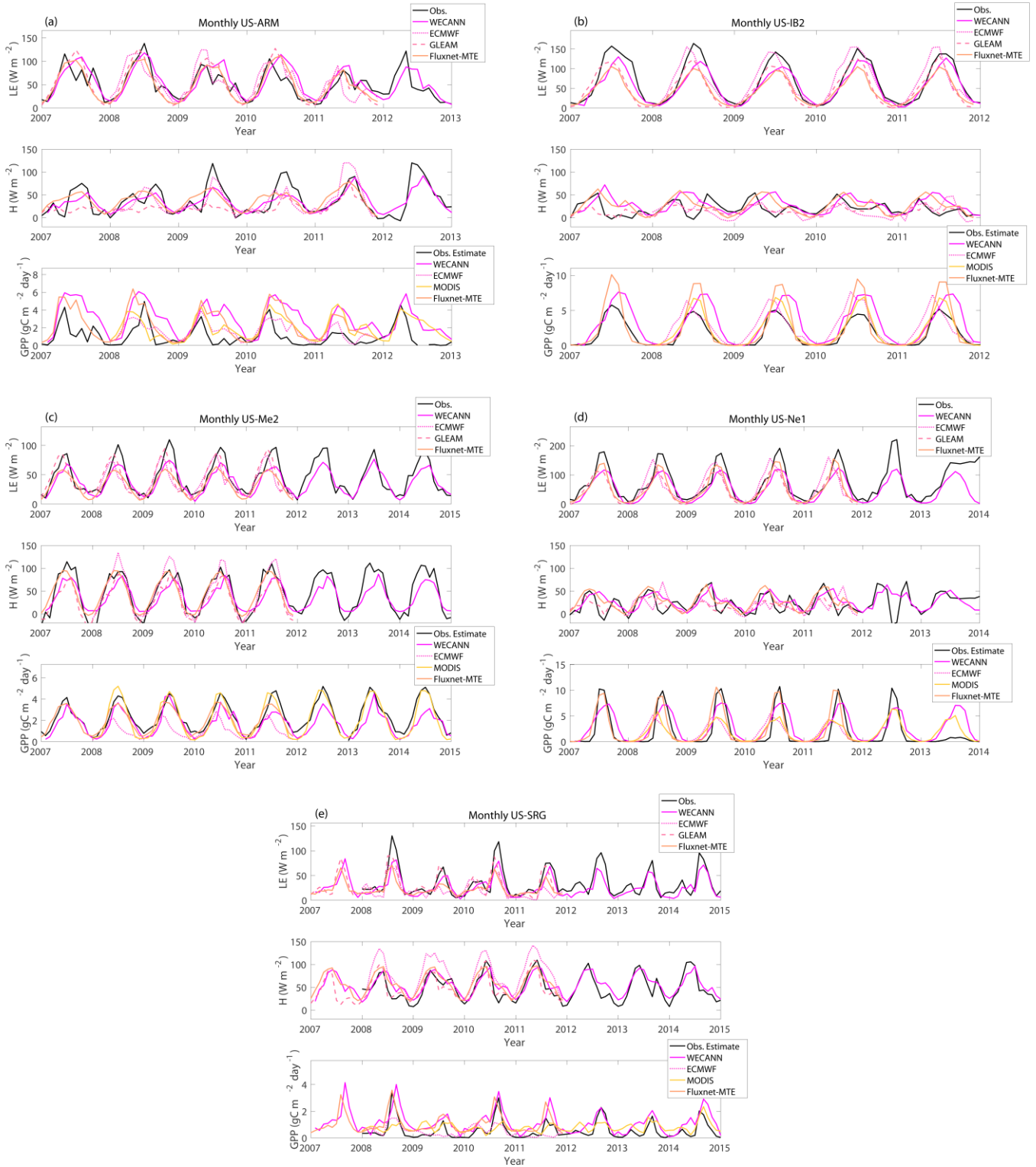


Figure 11: Same as Figure 9 but for Canadian sites (a) CA-Qfo, and (b) CA-Qcu.



5

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.

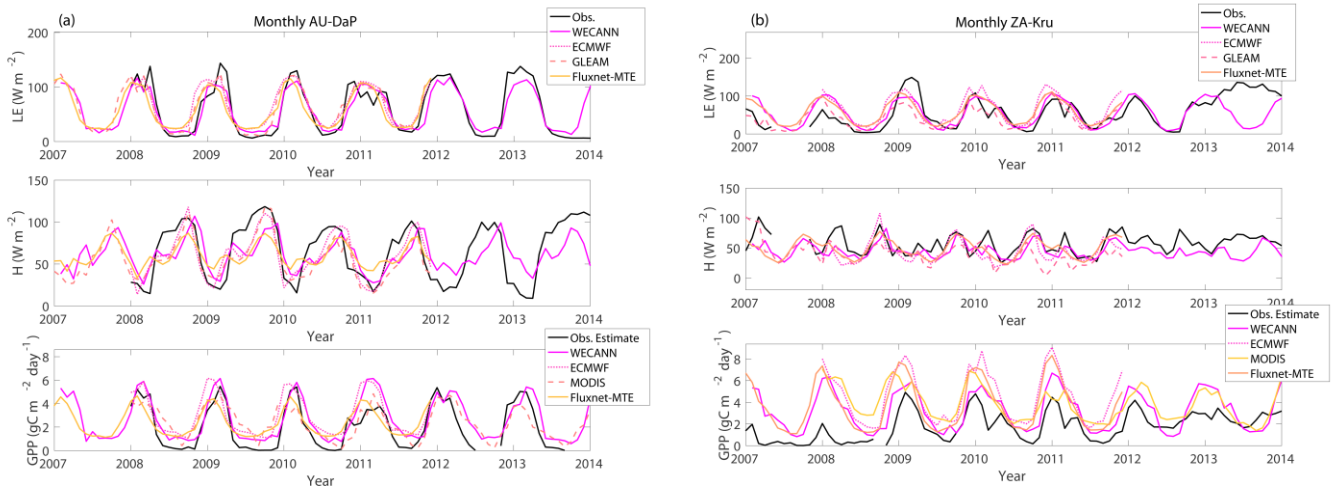


Figure 13: Same as Figure 9 but for (a) AU-DaP, Australia, (b) ZA-Kru, South Africa

5

10

15

20

25

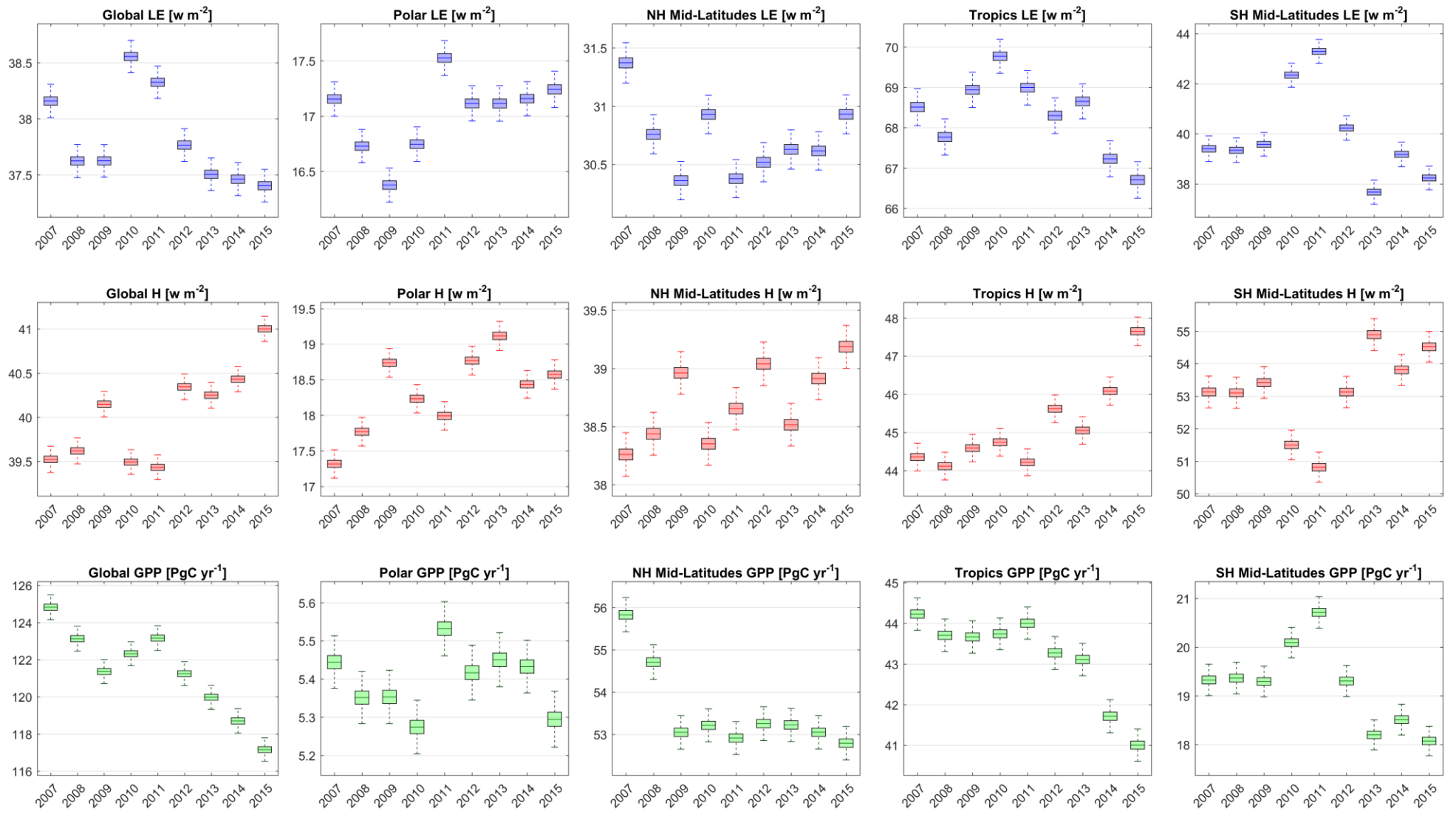


Figure 14: Uncertainty estimates of LE (top row), H (middle row) and GPP (bottom row) retrievals at global (left column) and regional (four right columns) scales between 2007 and 2015.

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

5 **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	Luoju et al., 2013

Table 3: Comparison of WECANN retrievals with retrievals from an ANN without SIF as an input

	LE		H		GPP	
	RMSD [W·m ⁻²]	R ²	RMSD [W·m ⁻²]	R ²	RMSD [gC·m ⁻² ·day ⁻¹]	R ²
WECANN	11.13	0.95	13.35	0.89	1.23	0.90
ANN w/o SIF	12.33	0.94	13.89	0.88	1.33	0.88

10

15

Table 4: Statistics of LE retrievals compared to eddy covariance measurements. Bold fonts represent best performing dataset statistics.

SiteID	Correlation Coefficient				RMSE [$W\ m^{-2}$]				Bias [$W\ m^{-2}$]			
	WECANN	FLUXNET-MTE	GLEAM	ECMWF	WECANN	FLUXNET-MTE	GLEAM	ECMWF	WECANN	FLUXNET-MTE	GLEAM	ECMWF
AT-Neu	0.91	0.94	0.93	0.96	16.25	15.96	14.04	10.11	-2.30	-7.96	0.87	0.60
AU-DaP	0.89	0.80	0.87	0.82	22.25	26.85	21.42	25.42	-0.56	1.44	5.47	8.72
BE-Bra	0.96	0.78	0.80	0.79	8.58	18.20	20.89	19.86	11.28	14.02	9.46	23.34
BR-Ji1	0.13	0.09	0.65	0.13	16.67	16.21	13.41	19.06	27.83	29.08	15.52	31.35
BR-Ji2	0.77	0.77	0.04	0.49	4.16	2.93	18.88	11.24	4.85	5.32	-2.02	10.14
BR-Sa3	0.70	0.73	0.24	0.91	6.52	6.74	10.87	7.84	-6.61	-7.85	-4.79	5.14
BR-Sp1	0.84	0.90	0.92	0.90	14.74	12.50	11.10	12.70	7.63	-0.69	3.67	15.00
CA-Gro	0.94	0.88	0.79	0.87	11.61	14.45	22.27	17.40	9.96	3.20	17.79	13.94
CA-Que	0.96	0.84	0.78	0.89	9.58	15.67	23.77	13.89	1.64	-1.64	12.18	6.52
CH-Fru	0.81	0.76	0.69	0.66	21.35	24.09	27.85	29.96	1.63	-3.52	4.96	11.22
ES-LgS	0.76	0.58	-0.06	0.07	14.09	17.24	25.79	27.71	-1.21	-9.17	-19.21	-4.45
FI-Hyy	0.97	0.90	0.82	0.88	7.05	12.21	18.52	16.16	1.01	-0.48	4.49	7.93
US-ARM	0.79	0.85	0.85	0.77	19.67	16.59	21.22	22.37	9.27	4.82	6.77	9.79
US-IB2	0.95	0.86	0.86	0.84	19.31	26.15	26.79	28.50	-11.49	-17.56	-16.27	2.71
US-Me2	0.91	0.87	0.77	0.62	13.08	15.25	18.43	23.27	-7.49	-11.82	-3.11	-7.61
US-Ne1	0.85	0.90	0.85	0.65	35.71	25.59	31.03	45.20	-28.02	-15.06	-23.50	-9.52
US-SRG	0.90	0.84	0.81	0.71	13.82	19.90	17.55	21.53	-8.38	-12.03	-4.91	-13.81
ZA-Kru	0.50	0.69	0.59	0.65	38.81	28.79	30.22	31.82	-5.09	14.70	-1.83	17.81
Average	0.81	0.78	0.68	0.70								

Table 5: Statistics of H retrievals compared to eddy-covariance measurements. Bold fonts represent best performing dataset statistics.

SiteID	Correlation Coefficient				RMSE [$W\ m^{-2}$]				Bias [$W\ m^{-2}$]			
	WECANN	FLUXNET-MTE	GLEAM	ECMWF	WECANN	FLUXNET-MTE	GLEAM	ECMWF	WECANN	FLUXNET-MTE	GLEAM	ECMWF
AT-Neu	0.88	0.78	0.71	0.73	6.43	17.44	12.56	10.65	20.33	34.34	7.41	9.26
AU-DaP	0.68	-0.85	0.82	0.82	25.68	63.64	18.93	18.91	-4.03	-5.20	-9.21	-4.61
BE-Bra	0.87	0.78	0.78	0.90	18.01	18.97	22.51	15.06	-8.45	18.71	-21.85	-28.21
BR-Ji1	0.29	0.42	0.57	0.45	9.27	5.67	10.09	9.27	-4.88	-7.66	-5.70	-11.56
BR-Ji2	0.83	0.64	0.51	0.85	4.79	3.57	14.10	8.09	7.00	4.28	7.92	3.02
BR-Sa3	0.97	0.88	0.79	0.50	3.21	2.04	9.84	6.33	3.03	-1.34	6.84	-7.45
BR-Sp1	0.92	0.69	0.70	0.83	5.31	8.40	7.99	7.43	4.05	12.42	-3.47	-1.50
CA-Gro	0.71	0.67	0.88	0.92	16.54	22.42	11.02	9.03	-7.45	4.03	-15.91	-8.35
CA-Que	0.92	0.89	0.54	0.75	7.85	10.77	19.73	19.60	10.10	12.77	-8.76	10.87
CH-Fru	0.75	0.70	0.50	0.53	12.03	19.43	17.41	16.06	11.41	31.45	3.09	1.25
ES-LgS	0.87	0.59	0.75	0.81	26.93	46.40	34.49	32.83	0.47	-23.91	4.07	17.25
FI-Hyy	0.85	0.91	0.86	0.93	16.51	12.76	16.67	12.45	0.01	5.78	-9.44	-14.33
US-ARM	0.80	0.52	0.59	0.70	18.47	28.31	21.24	22.11	1.44	14.82	-11.65	1.16
US-IB2	0.11	-0.03	0.58	0.35	21.69	37.07	12.36	16.71	8.22	22.49	-5.68	-5.97
US-Me2	0.89	0.87	0.91	0.92	22.87	33.33	18.97	18.65	-9.12	-13.89	-14.65	2.49
US-Ne1	0.47	-0.07	0.67	0.08	20.76	54.58	15.70	25.40	3.07	27.08	-5.09	-1.80
US-Ro1	0.81	0.63	0.42	0.33	10.00	27.64	16.87	18.35	9.54	27.42	-0.28	-1.14
US-Ses	0.92	0.74	0.82	0.92	9.23	17.57	14.08	13.14	-14.94	-41.03	-20.87	-6.25
US-SRG	0.87	0.02	0.86	0.88	13.98	30.95	14.80	18.01	6.53	-25.52	0.81	28.58
ZA-Kfu	0.59	0.18	0.47	0.59	12.97	32.57	20.12	17.27	-12.78	-1.55	-16.42	-10.82
Average	0.75	0.50	0.69	0.69								

Table 6: Statistics of GPP retrievals compared to eddy-covariance measurements. Bold fonts represent best performing dataset statistics.

SiteID	Correlation-Coefficient				RMSE [$\text{gC}\cdot\text{m}^{-2}\cdot\text{day}^{-1}$]				Bias [$\text{gC}\cdot\text{m}^{-2}\cdot\text{day}^{-1}$]			
	WECANN	FLUXNET-MTE	MODIS	ECMWF	WECANN	FLUXNET-MTE	MODIS	ECMWF	WECANN	FLUXNET-MTE	MODIS	ECMWF
AT-Neu	0.86	0.92	0.92	0.94	1.35	1.05	1.08	1.04	-0.52	0.14	-0.38	-0.93
BE-Bra	0.94	0.88	0.90	0.79	0.86	1.89	1.38	1.00	0.65	2.28	1.25	0.65
BR-Ji1	0.80	0.51	0.34	0.76	1.53	1.92	2.09	1.47	1.14	1.92	-0.32	2.49
BR-Ji2	0.84	0.83	-0.03	0.82	0.82	0.79	1.57	1.34	-1.20	-0.40	-1.98	-0.08
BR-Sa3	0.44	0.70	-0.83	-0.73	0.85	0.71	1.74	1.86	0.53	1.20	-0.92	2.77
BR-Sp1	0.94	0.95	0.87	0.94	1.04	0.72	1.31	0.75	0.54	-0.47	0.36	-0.20
CA-Gro	0.60	0.41	0.41	0.45	1.98	2.46	2.66	2.56	0.64	0.92	1.20	0.87
CA-Qeu	0.74	0.38	0.51	0.54	1.53	1.95	2.19	1.73	0.63	1.09	0.91	0.55
CA-Qfo	0.98	0.88	0.90	0.91	1.10	1.39	1.69	1.23	0.93	1.17	1.21	0.86
CH-Fru	0.91	0.91	0.93	0.90	1.13	1.63	1.51	1.12	0.70	1.61	0.97	0.60
ES-LgS	0.66	0.29	-0.03	-0.28	0.60	0.50	0.71	0.70	1.36	0.22	0.78	0.26
FI-Hyy	0.97	0.89	0.91	0.91	0.56	1.52	1.41	1.21	0.23	1.22	0.89	0.71
US-IB2	0.85	0.86	0.96	0.84	1.42	1.94	0.80	1.27	1.41	0.98	0.57	0.97
US-Me2	0.89	0.91	0.94	0.65	0.67	0.53	0.60	1.02	-0.64	-0.08	-0.23	-1.22
US-Ne1	0.77	0.88	0.69	0.54	2.33	1.69	2.43	3.04	0.80	0.30	-0.03	-0.42
US-Ses	0.63	0.76	0.42	0.58	0.53	0.40	0.62	0.60	0.38	0.03	0.15	-0.51
US-SRG	0.76	0.87	0.52	0.59	0.51	0.38	0.59	0.59	0.55	0.50	0.23	-0.07
ZA-Kru	0.69	0.67	0.40	0.70	1.31	1.61	1.58	1.66	1.27	2.33	1.87	2.93
Average	0.80	0.75	0.54	0.60								