Predicting carbon and energy fluxes across global FLUXNET sites with regression algorithms

G. Tramontana[1], M. Jung[2], G. Camps-Valls[3], K. Ichii[4,5], B. Raduly[1,6], M. Reichstein[2], C. R. Schwalm[7], M. A. Arain[8], A. Cescatti[9], G. Kiely[10], L. Merbold[11], P. Serrano-Ortiz[12], S. Sickert[13], S. Wolf[14] and D. Papale[1].

[1]Department for Innovation in Biological, Agro-food and Forest systems (DIBAF), Univeristy of Tuscia, Viterbo, 01100, Italy,

[2]Max Planck Institute for Biogeochemistry, Jena, 07745, Germany;

[3]Image Processing Laboratory (IPL), Paterna (València), 46980, Spain.

[4]Department of Environmental Geochemical Cycle Research, Japan Agency for Marine-Earth Science and Technology, Yokohama, 236-0001, Japan.

[5]Center for Global Environmental Research, National Institute for Environmental Studies, Tsukuba, 305-8506, Japan.

[6]Department of Bioengineering Sapientia Hungarian University of Transylvania, Miercurea Ciuc, 530104, Romania.

[7]Woods Hole Research Center, Falmouth MA, 02540, USA.

[8]School of Geography and Earth Sciences, McMaster University, Hamilton (Ontario), L8S4L8, Canada

[9]Institute European Commission, Joint Research Centre, Institute for Environment and Sustainability, Ispra, Ispra, 21027, Italy

[10]Civil & Environmental Engineering and Environmental Research Institute, University College, Cork, T12 YN60, Ireland.

[11]Department of Environmental Systems Science, Institute of Agricultural Sciences, ETH Zurich, Zurich, 8092, Switzerland.

[12]Department of Ecology, University of Granada, Granada, 18071, Spain.

[13]Computer Vision Group, Friedrich Schiller University Jena, 07743 Jena, Germany

[14]Department of Environmental Systems Science, ETH Zurich, Zurich, 8092, Switzerland.

Correspondence to: G. Tramontana (g.tramontana@unitus.it)

Use present time wherever possible!

Abstract

Spat~~ial-~~io-temporal fields of land-atmosphere fluxes simulated with ~~derived from~~ data-driven models can complement simulations from ~~by~~ process-based Land Surface Models. While a number of strategies for empirical models with eddy covariance flux data have been applied, a systematic intercomparison of these methods is ~~has been~~ missing so far. In this study, we perform a cross-validation experiment to ~~for~~ predict~~ing~~ carbon dioxide, latent heat, sensible heat and net radiation fluxes, for ~~in~~ different ecosystem types with eleven machine learning (ML) methods from four different classes (kernel methods, neural network, tree methods, and regression splines). We applied ~~employ~~ two complementary setups: (1) eight days average fluxes based on remotely sensed data, and (2) daily mean fluxes based on meteorological data and mean seasonal cycles of remotely sensed variables. The pattern of predictions from different ML and setups a~~we~~re very consistent. There a~~we~~re systematic differences in performance between ~~among~~ the fluxes, in ~~with~~ the following ascending order: net ecosystem exchange (R^2<0.5), ecosystem respiration (R^2>0.6), gross primary production (R^2>0.7), latent heat (R^2>0.7), sensible heat (R^2>0.7), net radiation (R^2>0.8). ML methods predicted very well the across sites variability and the seasonal cycle (R^2> 0.7) of the observed fluxes, while the weekly deviations from the mean seasonal cycle were not well predicted (R^2< 0.5). Fluxes were better predicted at forested and temperate climate sites than at ones growing in extreme climates or less represent~~at~~ed in training data (e.g. the tropics). The large ensemble of ML based models evaluated will be the basis of new global flux products.

Keywords: Machine learning, carbon fluxes, energy fluxes, FLUXNET, remote sensing, FLUXCOM

1. Introduction

Improving our knowledge of the carbon, water, and energy exchang~~es~~ between terrestrial ecosystems and the atmosphere is essential to better understand and model the Earth's climate system (IPCC, 2007; Reich, 2010). In situ observations of these exchanges are obtained by the eddy covariance technique, which directly measures turbulent fluxes between ecosystems and the atmosphere (Aubinet at al., 2012; Baldocchi et al., 2014). The large-scale measurement network, FLUXNET integrates site observations of these fluxes globally and provides detailed time series of carbon and energy fluxes across biomes and climates (Baldocchi et al., 2008). However, eddy-covariance measurements are site-level observations, and spatial upscaling is required to estimate these fluxes at regional to global scales.

The increasing number of eddy covariance sites across the globe has encouraged the application of data-driven models by machine learning (ML) methods such as Artificial Neural Networks (ANNs, Papale et al., 2003), Random Forest (RF, Tramontana et al., 2015), Model Trees (MTE, Jung et al., 2009; Xiao et al., 2008, 2010) or Support Vector Regression (SVR, Yang et al., 2006, 2007) to estimate surface-atmosphere fluxes from site level to regional or global scales (e.g. Beer et al., 2010, Jung et al., 2010, 2011; Kondo et al., 2015; Schwalm et al., 2010; Yang et al., 2007; Xiao et al., 2008, 2010). The ML upscaled outputs are also increasingly used to evaluate land surface models (e.g., Anav et al., 2013; Bonan et al., 2011; Ichii et al., 2009; Piao et al., 2013).

The key characteristic of data-driven models compared the process-based ones are the former's intrinsic observational nature, and the fact that functional relationships are not prescribed but rather emerge from ~~the~~ patterns found in the measurements. In this context, data-driven models extract multivariate functional relationships between the in - situ measured fluxes of the network and explanatory variables in an empirical way. The explanatory variables ~~are~~ generally ~~coming~~ originate from satellite remote sensing, providing (VI's are only partially descriptive for vegetation state!) information on vegetation state (e.g., vegetation indices) and other land surface properties (e.g., surface temperature), along with continuous measurements of meteorological variables at flux towers.

While ML-based upscaling provides a systematic approach to move from point-based flux estimates to spatially explicit gridded fields, various sources of uncertainty exist. For example, individual ML methods ~~might~~ can have different responses, especially when these models are applied beyond the conditions ~~sampled~~ represented by the training dataset (Jung et al., 2009; Papale et al., 2015). The information content of the driving input variables may not be sufficient to capture the variability of the fluxes in all conditions (Tramontana et al., 2015). Moreover, remotely sensed and meteorological gridded datasets are affected by uncertainties themselves. Remote sensing data contain noise, bias~~es~~ and gaps, and can be perturbed by atmospheric effects or by the presence of snow. Meteorological gridded datasets are known to contain product specific biases (Garnaud et al., 2014; Tramontana et al., 2015; Zhao et al., 2012).

Thorough experiments using multiple data-driven models and explanatory variables are an essential step to identify and assess limitations and sources of uncertainty in the empirical upscaling approach. In this study, we present and evaluate an ensemble of ML based empirical models to predict carbon and energy fluxes across FLUXNET sites. The ~~participating~~ included in this study models were selected in the context of the FLUXCOM initiative, forming the basis of subsequent global flux products. We performed a consistent cross-validation for two complementary experimental setups using: (1) eight days average fluxes based on remotely sensed data, and (2) daily mean fluxes based on remotely sensed and meteorological data. The ML tools span the full range of commonly applied algorithms: from model tree ensembles, multiple adaptive regression splines, artificial neural networks, to kernel methods, with several representatives of each family. The different ML algorithms were trained with consistent sets of predictor variables. Our overarching aim is to understand how well fluxes of carbon (gross primary production (GPP), terrestrial ecosystem respiration (TER) and net ecosystem exchange (NEE)), and energy (latent heat (LH), sensible heat (H) and net radiation (Rn)) can be predicted by an ensemble of ML methods. More specifically, we address the following questions:

1. Are the patterns of predicted fluxes consistent between the two experimental setups?
2. How different are the predictions for ~~of~~ the various ML algorithms?
3. How does ~~the~~ performance differ between ~~among~~ capturing the across-sites, seasonal and the deviations from the mean seasonal cycle variability?
4. How does the performance differ among climate zones or ecosystem types?

2. Material and methods

2.1 Data

2.1.1 Eddy covariance study sites

We used eddy covariance data from 224 flux-tower sites (supplementary material, Sect. S1), which originate from the FLUXNET La Thuile synthesis dataset and CarboAfrica network (Valentini et al., 2014). The study sites are distributed globally and cover most plant functional types (PFT) and biomes over the globe (Table 1, Giri et al., 2005).

2.1.2 Observation-based carbon and energy fluxes

All flux measurements were post-processed using standardized procedures of quality (Reference?) control and gap-filled according to ~~following~~ Reichstein et al. (2005) and Papale et al. (2006). Estimates of GPP and TER were derived from half-hourly NEE measurements using two independent flux partitioning methods: (1) According to ~~following~~ Reichstein et al. (2005), where the temperature sensitivity of ecosystem respiration is initially estimated from night-time NEE data and then extrapolated to daytime to estimate TER and GPP. This is done ~~,~~ by subtracting NEE (negatively signed for the carbon uptake) from TER; (2) According to ~~following~~ Lasslop et al. (2010), where daytime NEE data are ~~is~~ used to constrain a hyperbolic light response curve to directly estimate GPP and TER. In the following we refer to ~~reference~~ GPP and TER derived by Reichstein et al. (2005) as as GPP_R and TER_R~~;~~ E ~~whereas e~~stimates based on the Lasslop et al. (2010) method are referred to as GPP_L and TER_L.

H ~~The h~~alf-hourly data a ~~w~~ere aggregated to daily values and screened according to multiple quality criteria, as follows~~:~~

1) we excluded data when more than 20% of the data were based on gap-filling with low confidence (Reichstein et al., 2005);

2) we identified and removed obviously erroneous periods due to non-flagged instrument or flux partitioning failures based on visual interpretation;

3) we excluded data-points where the two flux-partitioning methods provided extremely different patterns. Specifically, we computed for each site a robust linear regression between (a) TER_R – GPP_L and NEE, and (b) GPP_R and GPP_L. Data points with a residual outside the range of ± 3 times of the inter-quartile range were removed. This criterion removed only the extreme residuals, systematic differences between methods were not removed;

4) we removed the 5% of data-points with the largest friction velocity (u*) uncertainty, defined as data points above the 95th percentile of daily u* uncertainty (measured as inter-quartile range of 100 bootstrap samples).

Similarly to the carbon fluxes, we applied the same criteria 1) and 2) for the energy fluxes. Additionally, we removed data with inconsistent energy fluxes, i.e. when the residual of a robust linear regression between LE + H and Rn for each site was outside three-times the inter-quartile range of the residuals.

2.1.2 Remote sensing data

We collected data coming from Moderate Resolution Imaging Spectroradiometer (MODIS) which provides data at a spatial resolution of 1km or better (Justice et al., 2002). We used the cutout surroundings of 3×3 km pixels centered on the towers. We collected and processed the following MODIS products: MOD11A2 Land Surface Temperature (LST; Wan et al., 2002), MOD13A2 Vegetation Index (Normalized Difference Vegetation Index (NDVI) and Enhanced Vegetation Index (EVI); Huete et al., 2002), MOD15A2 Leaf Area Index (LAI) and Fraction of Absorbed Photosynthetic Active Radiation (~~FPAR~~ fAPAR; Myneni et al., 2002), and MCD43A2 and MCD43A4 Bidirectional Reflectance Distribution Function

(BRDF) corrected surface reflectances (Schaaf et al., 2002). The BRDF-corrected surface reflectance data were further processed to calculate the Normalized Difference Water Index (NDWI) (Gao, 1996) and the Land Surface Water Index (LSWI) (Xiao et al., 2002). These data were obtained from http://daac.ornl.gov/MODIS/.

The remote sensing data were further processed to improve data quality and data gaps which were filled to create continuous time-series data, and to minimize non-land surface signals. We adopted the following processing scheme: we identified good quality pixels by ~~the~~ using the quality assurance/quality control criteria (QA/QC) included in the MODIS product. If more than 25% of pixels had good quality at the time of snapshot, the average of good quality pixels were assigned as the actual value (Why was the Maximum Value Composite not used?). Otherwise, the data at the time snapshot were marked as blank (no data). Then, we created the mean seasonal variations from 2000-2012 using only good pixels data and the data gaps in the processed data were filled using the mean seasonal variation. Only MOD13 is provided with 16 days composite (Which type of composite?), and eight days data were created by assigning the 16 days composite value to the corresponding two eight days periods.

2.1.3 Meteorological data

The in situ measured air temperature (Tair), global radiation (Rg), VPD (VPD being?), and precipitation were used after data screening according to the criteria 1) and 2) applied as for the measured fluxes (see Sect. 2.1.2). We also used long-term time series of these variables from the ERA-Interim dataset (Reference?) for the period 1989-2010 (Dee et al., 2011)~~, which~~. These data were bias-corrected for each site based on the overlapping period of in situ measurements (see http://www.bgc-jena.mpg.de/~MDIwork/meteo/). These long-term meteorological data are primarily used to calculate consistent metrics of climatological variables (e.g. mean annual temperature) ~~across~~ for all sites given the temporal coverage of data of the different sites. In addition, we use a composite of these ERA-Interim data based and in situ measured data , to obtain a gap-free time series for calculating a simple soil Water Availability Index (WAI, see Sect. 2.3.2 and supplementary material, Sect. S3).

2.2 ~~Participating~~ Applied ML methods

We chose 11 ML algorithms for regression from four broad families: tree-based methods, regression splines, neural networks, kernel methods. Moreover a comprehensive review of ML algorithms in biophysical parameter estimation can be found in Verrelst et al. (2015).

<u>Tree based methods</u>

These methods construct hierarchical binary decision trees. The inner nodes of the tree hold decision rules according to explanatory variables (e.g. less/greater than X1), recursively splitting the data into subspaces. The leaf nodes at the end of the decision tree contain models for the response variable. Because a single tree is generally not effective enough to cope with strong non-linear multivariate relationships, ensembles of trees are often used. We applied two different tree ensemble methods: (1) Random Forests (RF) which combines regression trees grown from different bootstrap samples and randomly selected features at each split node (Breiman, 2001; Ho, 1998); and (2) Model Tree Ensembles (MTE) which combine model trees (Jung et al., 2009). The main difference between regression and model trees is the prediction model in the leaf node: a simple mean of the target values from training in regression trees and a parametric function (here a multiple linear regression) in model trees. In this study, we used three different variants of MTE, which differ mainly with respect to different cost functions ~~for~~ to determin~~ing~~e the splits~~,~~ and the technique to create an ensemble of model trees. Further details are described in the supplementary material (Sect. S2).

<u>Regression splines</u>

Multivariate regression splines (MARS) are an extension of simple linear regression that can adapt to non-linear response surfaces—using piecewise (local) functions—by which a target variable is predicted by the sum of regression splines and a constant value (Alonso Fernández, 2013; Friedman et al., 1991).

<u>Neural networks</u>

Neural networks are based on nonlinear and nonparametric regressions. Their base unit is the neuron, where nonlinear regression functions are applied. The neurons are interconnected and organized in layers. The output of *m* neurons in the current layer are the inputs for *n* neurons of the next layer. We used two types of neural networks: the artificial neural network (ANN) and the group of method for data handle (GMDH). In an ANN, each neuron performs a linear regression followed by a non-linear function. Neurons of different layers are interconnected by weights ~~that are~~ adjusted during the training (Haykin et al., 1999; Papale et al., 2003). The GMDH is a self-organizing inductive method (Ungaro et al., 2005) building polynomials of polynomials; the neurons are pairwise connected through a quadratic polynomial to produce new neurons in the next layer (Shirmohammadi et al., 2015).

Kernel methods

Kernel methods (Shawe-Taylor and Cristianini, 2004) owe their name to the use of kernel functions, which measure ~~measures~~ similarities between input data examples. Among the available kernel methods we used: (1) support vector regression (SVR) (Vapnik et al., 1997), (2) kernel ridge regression (KRR) (Shawe-Taylor and Cristianini, 2004), and (3) Gaussian process regression (GPR) (Rasmussen, 2006; Verrelst et al., 2012). The SVR defines a linear prediction model over mapped samples to a much higher dimensional space, which is non-linearly related to the original input (Verrelst et al., 2012, Yang et al., 2007). The KRR is considered as the kernel version of the regularized least squares linear regression (Shawe-Taylor and Cristianini, 2004, Verrelst et al., 2012). The GPR is a probabilistic approximation to nonparametric kernel-based regression, and both a predictive mean (point-wise estimates) and predictive variance (error bars for the predictions) can be derived. We also used a hybrid approach combining RF with simple decision stumps in the inner nodes and GPR for prediction in the leaf nodes (Fröhlich et al., 2012).

2.3 Experimental design

2.3.1. Experiment setups

We defined two complementary experimental setups, which differ in the choice of explanatory variables, and the temporal resolution of the target fluxes: 1) at eight days temporal resolution using exclusively remote sensing data (hereafter RS); and 2) at daily temporal resolution using meteorological data together with the mean seasonal cycle (MSC) of remote sensing data (hereafter RS+METEO). In the latter case, the MSC of remote sensing data were smoothed and interpolated to a daily time step. Each setup has advantages and disadvantages. While RS ~~can~~ provides products with ~~high~~ a spatial resolution ~~(up to 1km)~~ of 1 km or less, data are limited to the MODIS era (2000-present) and has a coarse (weekly) temporal resolution. The uncertainties of remote sensing data at tower locations due to finer scale spatial heterogeneity may also degrade the performance of the ML methods. RS+METEO can take advantage of information from meteorological variables, and is resistant to the noise of remote sensing time series because only mean seasonal cycle of data from satellite ~~were~~ are used. RS+METEO allows for upscaled products over a longer time period (because not constrained by the availability of MODIS data) and finer time scale (daily). However the predictive ~~skill~~ capacity of this setting ~~was~~ is conditioned by ~~the~~ missing ~~of~~ information regarding the interannual variability of vegetation greenness (phenology). In addition the use of meteorological gridded datasets introduces another source of uncertainty ~~coming from~~ e.g., potential dataset specific biases and ~~by~~ the~~ir~~ typical~~ly~~ coarse spatial resolution (0.5 degrees or larger).

2.3.2. Variable selection

Combining remote sensing and meteorological data (see Sect. 2.1.2 and 2.1.3) we created additional model input variables~~ for model input~~s. In the case of RS+METEO setup we derived the Water Availability Index (WAI) ~~that it~~ is based on a simple soil water balance model (for more details see supplementary material, Sect. S3) as an attempt to better represent ~~water stressed~~ waterstress conditions. For both setups we derived proxies for absorbed radiation as the product between vegetation greenness (e.g. EVI, NDVI, fAPAR ~~FPAR~~) and available energy (e.g. daytime LST, Rg, and potential radiation). Other derived variables include the mean seasonal cycle (phenology) (MSC) of dynamic variables and associated metrics (minimum, maximum, amplitude, and mean). For remote sensing predictors, the MSC and associated metrics are based on the period 2001-2012 while ~~for~~ climate variables are based on the bias corrected daily long-term ERA-Interim data reference period (1989-2010). In total, 216 potential explanatory variables were created for RS and 231 for RS+METEO (see supplementary material S4 for details).

For each setup we selected a small subset of variables optimally suitable ~~best suited~~ to predict ~~the~~ target fluxes using a variable selection search algorithm. Variable selection is an important component in ~~the~~ spatial upscaling since it improves ~~because~~ the accuracy of predictions while ~~improves and the~~ computational costs of ~~the~~ global predictions are minimized. We used the Guided Hybrid Genetic Algorithm (GHGA) published by Jung and Zscheischler (2013), which was designed for variable selection problems with many candidate predictor variables and computationally expensive cost functions. The GHGA requires the training of a regression algorithm (here RF) to estimate the cost associated with selected variable subsets (see S5 for details).

Instead of doing a computationally demanding variable selection for each individual flux, variable selection runs were performed for the RS and RS+METEO setups and separately for carbon and energy fluxes. This procedure has the advantage that the resulting global products originate from a consistent set of predictor variables. The selected variables for the prediction of carbon and energy fluxes are listed in Table 2.

2.3.3. Model training

The capability of ML methods to spatially extrapolate carbon and energy fluxes ~~was~~ has been evaluated by a 10-fold cross-validation strategy. The training datasets were stratified into 10-folds, each one roughly containing 10% of the data. Entire sites a~~we~~re assigned to each fold (Jung et al., 2011). The target values for each fold ~~were~~ have been predicted based on the training using the remaining nine folds. Due to computational expense of the RS+METEO setup, only one method representing each “family” – multiple regressions, RF, MARS, ANN and KRR – was trained.

ML methods base settings ~~were~~ have been tuned before the training (for further details, see supplementary material S6). These hyper-parameters account for regularization ~~in order~~ to avoid overfitting, as well as for the shape and smoothness constraints. Instead, the model parameters were estimated for each ML every time in each fold.

2.3.4. Model evaluation

~~In order to~~ To highlight the differences between the RS and RS+METEO setups, the daily output from RS+METEO ~~were~~ has been aggregated to eight day~~s~~ time steps; the same periods and sites ~~were~~ have been used for ~~the~~ comparison. Besides the statistical analysis of the individual ML cross-validation results, we ~~focus~~ focussed on the ensemble median estimate, ~~here~~ defined as the median predicted value across all ML for a given setup and time step. The advantages of the median ensemble estimate is the robustness of the predictions for the contribution of many ML that reduce~~d~~ the risk of ~~outlier~~ outliers in the extrapolation exercise.

We used different metrics to evaluate the ML performance such as the Nash and Sutcliffe model efficiency (MEF), the root mean square error (MSE), the empirical BIAS, the

Pearson's linear correlation coefficient (ρ), the coefficient of determination (R^2) and the ratio of variance (ROV).

MEF (Nash and Sutcliffe, 1970), if the capability of a model to estimate a target variable~~s~~ is better than a reference model. If the reference model is the mean value of the target, MEF can be calculated as:

$$MEF = 1 - \frac{\sum_{i=1}^{n}(x_i - y_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \qquad (1)$$

where x_i and y_i are ~~were~~ the predicted and the observed values respectively and $\bar{y}$ is the mean value of the observations (here the reference model). MEF can vary between -inf to 1; if MEF > 0 the predictive capacity ~~skill~~ of the model is better than the mean (MEF = 1 for the ideal model), if MEF=0 the predictive capacity ~~skill~~ of the model is equivalent to the mean, finally if MEF < 0, the predictive capacity ~~skill~~ of the mean value of the target is better than the model.

The RMSE is ~~was~~ estimated ~~estimates~~ as the root square of the mean ~~mean~~ values of the squared residuals:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(x_i - y_i)^2}{n}} \qquad (2)$$

The BIAS is ~~was~~ evaluated as the difference~~s~~ between the mean values of model's residuals

$$BIAS = \frac{\sum_{i=1}^{n}(x_i - y_{i)}}{n} \qquad (3)$$

Following Gupta et al. (2009) the importance of bias on the overall uncertainty is ~~was~~ evaluated as the ratio between the square of BIAS and Mean Square Error, the latter estimated as the square value of RMSE.

The Pearson's linear correlation coefficient (ρ) is ~~was~~ the ratio between the covariance between the modeled and observed values (σ_{xy}) and the product of the standard deviation of modeled (σ_x) and observed (σ_y) values:

$$\rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \tag{4}$$

R^2 is ~~was~~ estimated as the squared value of ρ; finally ROV is ~~was~~ evaluated as the ratio between predicted and observed standard deviation.

We evaluated the overall predictive capacity ~~skill~~ of the models, evaluating the consistency among trained ML approaches and across the experimental setup. Then we evaluated the capability of the regression models to predict site-specific mean fluxes, mean seasonal cycle (MSC), and anomalies (Jung et al., 2011). The MSC per site was calculated using the average~~d~~ values for each eight day~~s~~ period across all available years, but only when at least two values (i.e., years) for each eight days period are ~~were~~ available. To assess the mean values of the study sites, we calculated the mean of the MSC if at least 50% of the 46 eight daily values are ~~were~~ present, whereas the weekly anomalies have been ~~were~~ calculated as the deviation of a flux value from the MSC. Finally, the mean site value was removed from the MSC to disentangle the seasonal variation from the mean, thus making the MSC and mean complementary.

We also analyzed the performance for the different Köppen climate zones and IGBP plant functional types (PFT's). In particular. we computed for each flux, setup and tower site the performances of ML median estimate. Then, for each setup, we estimated the median value of the site-by-site statistics per PFT and climate zone.

3. Results and Discussions

3.1 Overview

The ensemble median estimate always outperform~~ed~~s the median performance of ML-specific methods (the median value of metrics calculated for individual ML) (Table 3; Appendix A). (what does skill mean here? Unclear phrasing) Individual ML methods also exhibit~~ed~~ higher skill than multiple linear regressions (higher MEF and lower RMSE; Fig. 1). This highlights the added value of ML methods as these are able to account for nonlinearities in either explanatory variables or fluxes. Overall, using the ensemble median estimate gives a representative overview of ML-based flux predictions.

3.2 Predictive ~~skill~~ capacity of carbon and energy fluxes

Predictive ~~skill~~ capacity of the ensemble median estimate clustered into tiers whereby energy fluxes are uniformly better predicted than carbon fluxes: Rn > H/LE/GPP > TER > NEE (Table 3). The highest ~~skill~~ predicitve capacity levels as exhibited by net radiation shows near perfect agreement; Rn displays a model efficiency (MEF) of 0.91-0.92 and a correlation of 0.96. The decline in ~~skill~~ predicitve capacity for the second tier fluxes is ca. 15% to 20%; MEF for H, LE, and GPP is

0.79, 0.75-0.76, and 0.71 respectively. The lowest two tiers exhibits 20% and 40% declines in MEF (0.57-0.64 and 0.43-0.46 for TER and NEE respectively). These relative rankings are unchanged regardless of skill metric used—apart from RMSE where the difference in fluxes units and magnitude, confounds a direct comparison (Table 3)—suggesting that accuracy and precision scale linearly.

There were only minor performance differences between the two carbon fluxes partitioning methods (Table 3), although for the RS setup, the performance of TER_L were comparatively lower than TER_R (lower MEF, ρ and ROV). A similar trend was not found in the case of RS+METEO setup.

The overall predicitve capacity ~~skill~~ profile in this study confirms previous upscaling efforts (Jung et al., 2011; Yuan et al., 2010). This relatively stable cross-study predicitve capacity ~~skill~~ gradient reflects the information content of the available predictor variables. The spatiotemporal variability of remotely sensed land surface properties is well-suited to predict the top tier fluxes (Rn, H, LE, and GPP) (Jung et al., 2008; Tramontana et al., 2015; Xiao et al., 2010;.Yang et al., 2007)

The higher predicitve capacity ~~skill~~ associated with energy fluxes suggests that these fluxes are more easily predictable using the drivers selected in particular respect to NEE. In fact NEE is strongly controlled by external factors such as management and disturbances (Amiro et al., 2010; Thornton et al., 2002) and by lag and memory effects (Bell et al., 2012; Frank et al., 2015, Papale et al., 2015; Paruelo et al., 2005), which are both poorly captured by predictor variables typically used in upscaling and poorly constrained in general, i.e., data limited. Another reason for the low performances in NEE simulation can be in the uncertainties in the measurements that are larger compared to H and LE and have an important effect being NEE the difference between two large components (GPP and TER).

Among the carbon fluxes, GPP is ~~the best~~ the best predicted probably because the seasonal cycle and canopy properties, which are ~~were~~ strongly related to GPP, are ~~were~~ well represented by the

ML drivers. The intermediate skill of TER, relative to carbon fluxes only, is supported by its tight coupling to the well-predicted GPP and the availability of predictor variables that capture the temperature dependency of respiration. However specific drivers for TER could be still missing. In fact in contrast to GPP, the canopy properties are less important drivers of TER, while the soil properties, carbon pools and their turnover rates are key for respiratory processes (Amiro et al., 2010) but not available to be used as drivers. This likely explains the poor performance for TER in comparison with GPP.

3.3 Are ~~the~~ flux predictions consistent between RS and RS+METEO?

predicitve capacity ~~Skills~~, in terms of both performance tiers and absolute value of skill metrics, are similar for both RS and RS+METEO approaches with some differences, in particular: (1) RS and RS+METEO diverge~~d~~ more for those fluxes and sho~~wed~~ lower overall predicitve capacity ~~skill~~ levels, in particular for NEE (Fig. 1, Table 3); (2) MEF and correlation values are ~~were~~ slightly larger for RS than RS+METEO, excluding TER_L where the opposite was found, indicating an important role of the meteorological data for this version of the ecosystem respiration. It should be considered that the differences in performance~~s~~ can ~~could~~ be ~~be~~ also due to a different ensemble size, with the RS composed of 11 individual ML-based ensemble members, whereas RS+METEO is based on only four. The overall good performance of the RS setup implies that carbon and energy fluxes can be mapped exclusively based on remotely sensed inputs allowing high-spatial resolution products and reduction of uncertainty due to the meteorological drivers spatialization (Tramontana et al., 2015). Nonetheless, the differences between the experimental setups were less appreciable.

3.4 How different are the predictions of the various ML algorithms?

Pair-wise R^2 values for model outputs (Table 4) are ~~were~~ close to unity ($R^2 \geq 0.90$), regardless of the experimental setup, with NEE showing a slightly lower value ($R^2 = 0.84$). Between ~~Among~~ corresponding model residuals (Table 4), R^2 values ranged from 0.79 (Rn) to 0.89 (TER_L). Comparing the same ML technique but using different experimental setups (Table 4, RS vs. RS+METEO) show~~ed~~s similarly high, albeit a somewhat diminished level of consistency (R^2 range range~~d~~s from 0.71 to 0.80 for model residuals). These finding highligh~~ts~~t that the ML methods a~~we~~re mapping between explanatory variables and target fluxes both reliably and robustly. Between ~~Across the~~ all three consistency checks there i~~wa~~s also a tendency for better predicted fluxes (e.g., H) to exhibit higher pair-wise R^2 values than poorly predicted fluxes (e.g., NEE). This is expected as more robust patterns—and therefore those that lead to greater predictive power ~~skill~~—are easier to extract regardless of ML algorithm and experimental setup ~~in this study;~~ thus increasing consistency. ~~While~~ T~~t~~his broad consistency confirms that the extracted patterns are robust~~,~~. The ~~the~~ decline in R^2 when comparing the same ML trained with different drivers (RS vs RS+METEO) with respect to the correlation among ML methods with the same drivers, suggests that the choice of the explanatory variable ha~~d~~s a higher impact than the choice of the ML technique for the pattern of predictions.

3.5 How does ~~the~~ performance differ between ~~among~~ capturing ~~the~~ across-sites, seasonal and ~~the~~ deviations from ~~the~~ mean seasonal cycle variability?

Decomposing FLUXNET data into across-sites variability, mean seasonal cycle, and interannual variability components (Sect. 2.3.4) reveal~~ed~~s clear gradients in predictive ~~skill~~ capacity (Table 5 and Fig. 2). Across-site~~s~~ variability is ~~was~~ in general well-captured by the ML (R^2 range: 0.61 to 0.81 except for NEE) and the best predicted pattern for GPP and TER. This suggests that ML is ~~are~~ suitable to reproduce the spatial pattern of mean annual fluxes.

The variability in the mean seasonal cycle (at weekly time scale) was also uniformly well predicted (R^2 between 0.67-0.77 for GPP and TER, and between 0.86-0.98 for the energy fluxes) and the best predicted pattern for energy fluxes in particular for LE and Rn.

The importance of phenology ~~seasonality~~ in carbon and energy annual fluxes is known (Joiner et al., 2014; Jung et al., 2011; Merbold et al., 2009; Wolf et al., 2011). ~~and b~~ Biases in its phenological dynamics (e.g. in the growing season length) can ~~could~~ lead to biases in~~to the~~ predicted fluxes (Ichii et al., 2010). A clear benefit of the ML upscaling approach ~~used here~~ is that none of the parameters controlling phenology ~~seasonality~~ are prescribed, reducing the possibility of biase~~s~~.

In contrast, interannual variability is generally poorly captured by all the ML approaches ~~used with o~~ Only H and Rn show~~ing~~ a~~n~~ R^2 greater than 0.4. This low predictive capacity ~~skill~~ holds regardless ~~of~~ whether weekly, monthly (Jung et al., 2011), or annual time steps are used (data not shown). This is likely due to a combination of missing predictor variables (e.g. disturbances, management, legacy effects) and the noise/uncertainty in both predictor and target variables ~~that~~ play~~s~~ing a major role when small differences (like in the interannual variability) are predicted. The slightly better performances when sensible heat flux is estimated could be due to the lower uncertainty for ~~in~~ this flux with respect to the others (only one sensor used, the sonic anemometer, in contract with the other fluxes where also the gas analyzer is used) but also to the fact that it is strongly and directly related to ~~the~~ LST a ~~used~~ as driver. In any case, predicting interannual variability remains one of the largest challenges in the context of empirical upscaling.

NEE is confirmed to be the most difficult and consequently poorest predicted flux (Table 3). Predicitive capacity for NEE is ~~shows~~ considerably lower ~~skill~~ relative to the other fluxes for between ~~across~~-sites variability (R^2 = 0.46), the mean seasonal cycle (R^2 = 0.59), and interannual variability (R^2 = 0.13, TER_L is lowest at 0.10).

3.6 How does the performance differ among climate zones or ecosystem types?

Using climate zone and plant functional type (PFT) to disaggregate ML methods performances we find that in general energy fluxes are better predicted than carbon fluxes between ~~among the~~ different climate zones and PFTs (Fig. 3 and Appendix C). The median R^2 between simulation and observation for carbon fluxes (excluding NEE) is larger ~~greater~~ than 0.6 for more than 75% of the PFT's and climate zones, while for ~~in~~ the energy fluxes ~~an~~ R^2 greater than 0.7 is found for more than the 85% of the PFT's and climate zones (in all sites for Rn). NEE is again consistently poorly predicted (low R^2 and high relative RMSE; Fig. 3), apart from deciduous broadleaf forests (DBF) and MF where a marginal improvement is elicited ~~evident~~. The better performance for ~~in~~ these two vegetation types can ~~could~~ be related to the higher seasonal variance of NEE in comparison with the other PFTs that in ~~is~~ a pattern more consistent with the seasonal variance of the used drivers. Overall the ML methods show poor prediction capability in tropical ~~tropics~~ and evergreen broadleaf forests (EBF). Possible explanations are the absence of a clear seasonal cycle traceable by ~~the~~ remote sensing ~~signal~~ (evergreen vegetation) and a low variance in the~~ir~~ seasonal cycle ~~that is~~ challenging to explain~~ed~~ mathematically and to capture with a model. (Sims et al., 2008; Yebra et al., 2015; Yuan et al., 2010). In addition, the difficulty in acquiring cloud-free remotely sensed data introduces additional uncertainty in the drivers.

Reason for low performance~~s~~ can ~~be~~ also be the lack of important driving variables, that are probably the main explanation for cropland where management information ~~are~~ is missing (e.g. irrigation, fertilization, tillage) and also for cold and dry environments where the extreme nature of the environments characterized by water limitation (for dry sites) and the extreme low temperature (for cold sites) would require probably more targeted drivers such direct estimations of soil water content. In addition, cold and dry sites are characterized by both low magnitude and low variance of fluxes, making it difficult to explain the fluxes in these ecosystems types by empirical models.

4. Conclusions

The ML methods presented and evaluated in this study ~~have~~ show~~n~~ high capability to predict carbon and energy fluxes, in particular the between-site variability and the seasonal variations, with a general tendency of increasing performance in the following order: NEE, TER, GPP, LE, H, and Rn. The relatively poor performance for NEE likely results from factors that cannot be easily accounted for in ML based modelling approaches, such as legacies of site history (e.g. disturbances, management, age and stocks). Future progress in this direction requires the reconstruction of the relevant management and disturbance history, trying to integrate information from forest inventories, high resolution satellites such LANDSAT and high resolution biomass data from radar and LIDAR with the aim to improve model performance. The better results obtained for the energy fluxes (LE and H) in comparison to the carbon fluxes (GPP and TER) could be related to more complex mechanisms driving the carbon cycle that are also not represented in the drivers used, in addition to ~~a~~ relatively higher uncertainties in ~~the~~ GPP and TER due to the use of flux partitioning methods based on NEE measurements.

We found no substantial bias in the predictions of the ML models for most vegetation types or biomes. However, the predictions ~~have~~ deviated more from the observations for evergreen broadleaf forests, croplands, the tropics and extreme climates. The growing number of eddy-covariance sites, in particular new sites in poorly represented regions will likely improve~~s~~ the predictive capacity ~~skill~~ in the future. This is particularly relevant since tropical areas account for a disproportionate share of the global water and carbon cycle (Beer et al., 2010).

The deviations from the mean seasonal cycle (weekly anomalies) are still poorly captured by the ML methods. We expected the RS+METEO approach to perform better regarding the prediction of anomalies as meteorological drivers were included and the noise in remote sensing time series was greatly reduced by using smoothed seasonal cycle. However, this was not the case, which indicates that either, the weekly anomalies of both, the flux data, and the drivers, are strongly affected by the uncertainties, or that the anomalies are dominated by management and disturbance (or other factors) not accounted for in the predictors. Hence, the prediction of interannual anomalies remains a major unresolved research topic.

The predictions for ecosystem fluxes across FLUXNET by different ML techniques and by different explanatory variable sets (RS vs RS+METEO) are highly consistent, indicating that the extracted patterns by the trained models are robust, realistic and not subject to severe overfitting. The differences in predictions between ~~among~~ the RS and RS+METEO setups are slightly larger than between ~~among~~ different ML methods, suggesting that future activities should concentrate on identifying new driver variables to further improve the performance of flux~~es~~ prediction~~s~~. Nevertheless, we recommend using the ensemble median estimate of multiple ML techniques analyze ~~for analyzing~~ global flux products because extrapolation beyond the FLUXNET-sampled conditions can generate large~~r~~ differences between ~~among~~ methods than discernible from our cross-validation comparison.

The ML based models presented and extensively evaluated here are ~~form~~ the basis of an extensive archive of global gridded flux products, which is currently under development. The thorough cross-validation experiment presented in this paper helps users understanding the products' strengths and weaknesses. The good performance of the ML methods, the availability of an ensemble of them, and the detailed analysis of their uncertainties will make this archive an unprecedented data stream to study the global land-atmosphere exchange of carbon, water and energy.

Appendix A: Median performance of the methods.

In table A1 we reported, for both setups, the median value of skill metrics (MEF, RMSE, and absolute value of BIAS) realized by singular ML and their related variability such estimated as the median absolute deviation (MAD) from the median multiplied per 1.4826 (see Jung et al., 2009 for details)

Appendix B: scatterplot between the observations and the predictions by the median ensemble of ML.

In Fig. B1 and B2 we reported the scatterplots between eddy covariance observations and the modeled median ensemble estimates respetively for RS and RS+METEO setup. We reported the overall eight days time series, and the comparison for the across site variability, the mean seasonal cycle and the weekly deviations (or anomalies) from the mean seasonal cycle.

Appendix C Median value of site-by-site performance per vegetation and climate type.

At follow we reported the median estimate of site-by-site R^2, RMSE, and absolute bias per PFT and climate zones.

Acknowledgments

G. Tramontana was supported by the GEOCARBON EU FP7 project (GA 283080). D. Papale, M. Jung and M. Reichstein thank the support of the BACI H2020 (GA 640176) EU project. G. Camps-Valls wants to acknowledge the support by an ERC Consolidator Grant with grant agreement 647423 (SEDAL). K. Ichii was supported by Environment Research and Technology Development Funds (2-1401) from the Ministry of the Environment of Japan and the JAXA Global Change Observation Mission (GCOM) project (#115). C. R. Schwalm was supported by National Aeronautics and Space Administration (NASA) Grants #NNX12AP74G, #NNX10AG01A, and #NNX11AO08A. M. A. Arain thanks the support of Natural Sciences and Engineering Research Council (NSREC) of Canada. P. Serrano Ortiz was partially supported by the Spanish Ministry of Economy and Competitiveness though the project CGL2014-52838-C2-R(GEISpain). S. Wolf acknowledges support from a Marie Curie International Outgoing Fellowship (European Commission, grant 300083). This work used Eddy Covariance data acquired by the FLUXNET community and in particular by the following networks: AmeriFlux (U.S. Department of Energy, Biological and Environmental Research, Terrestrial Carbon Program (DE-FG02-04ER63917 and DE-FG02-04ER63911)), AfriFlux, AsiaFlux, CarboAfrica, CarboEuropeIP, CarboItaly, CarboMont,ChinaFlux, Fluxnet-Canada (supported by CFCAS, NSERC, BIOCAP, Environment Canada, and NRCan), GreenGrass, KoFlux, LBA, NECC, OzFlux, TCOS-Siberia, USCCC. We acknowledge the financial support to the eddy covariance data harmonization provided by CarboEuropeIP, FAO-GTOS-TCO, iLEAPS, Max Planck Institute for Biogeochemistry, National Science Foundation, University of Tuscia and US Department of Energy and the databasing and technical support from Berkeley Water Center, Lawrence Berkeley National Laboratory, Microsoft Research eScience, Oak Ridge National Laboratory, University of California - Berkeley, University of Virginia.

5. References

Amiro, B. D., Barr, A. G., Barr, J. G.,Black, T. A., Gracho, R., Brown, M., Chen, J., Clark, K. L., Davis, K. J., Desai, A. R., Dore, S., Engel, V., Fuentes, J. D., Goldstein, A. H., Goulden, M. L., Kolb, T. E., Lavigne, M. B., Law, B. E., Margolis, H. A., Martin, T., McCaughey, J. H., Misson, L., Montes, Helu, M., Noormets, A., Randerson, J. T., Starr, G. and Xiao, J.: Ecosystem carbon dioxide fluxes after disturbance in forests of North America. J Geophys Res-Biogeo, 115, G00K02, doi:10.1029/2010JG001390, 2010.

Alonso Fernández, J.R., DíazMuñiza, C., Garcia, Nieto, P.J., de Cos, Juez, F.J, Sánchez, Lasheras, F. and Roqueñíc, M.N.: Forecasting the cyanotoxins presence in fresh waters: A new model based on genetic algorithms combined with the MARS technique, Ecol Eng, 53, 68–78. doi:10.1016/j.ecoleng.2012.12.015, 2013.

Anav, A., Friedlingstein, P., Kidston, M., Bopp, L., Ciais, P., Cox, P., Jones, C., Jung, M., Myneni, R. and Zhu Z.: Evaluating the land and ocean components of the global carbon cycle in the cmip5 earth system models, J Climate, 26, 6801–6843, doi: http://dx.doi.org/10.1175/JCLI-D-12-00417.1, 2013.

Aubinet, M., Vesala, T. and Papale, D.: Eddy Covariance: A Practical Guide to Measurement and Data Analysis , Springer, Dordrecht Heidelberg London New York, 460, 2012.

Baldocchi, D.: Breathing of the terrestrial biosphere: lessons learned from a global network of carbon dioxide flux measurement systems, Aust J Bot, 56, 1–26, http://dx.doi.org/10.1071/BT07151, 2008.

Baldocchi, D.: Measuring fluxes of trace gases and energy between ecosystems and the atmosphere – the state and future of the eddy covariance method, Global Change Biol, 20, 3600-3609, DOI: 10.1111/gcb.12649, 2014.

Beer, C., Reichstein, M., Tomelleri, E., Ciais, P., Jung, M., Carvalhais, N., Rödenbeck, C., Arain, A., M., Baldocchi, D., Bonan, B., G., Bondeau, A., Cescatti, A., Lasslop, G., Lindroth, A., Lomas, M., Luyssaert, S., Margolis, H., Oleson, W. K., Roupsard, O., Veenendaal, E., Viovy, N., Woodward, I. F. and Papale, D.: Terrestrial Gross Carbon Dioxide Uptake: Global

Distribution and Covariation with Climate, Science, 329, 834-838, doi: 10.1126/science.1184984, 2010.

Bell, T. W., Menzer, O., Troyo-Diéquez, E. and Oechel, W.: Carbon dioxide exchange over multiple temporal scales in an arid shrub ecosystem near La Paz, Baja California Sur, Mexico, Global Change Biol, 18, 2570–2582, doi:10.1111/j.1365-2486.2012.02720.x., 2012.

Bonan, G. B., Lawrence, P. J., Oleson, K. W., Levis, S., Jung, M., Reichstein, M., Lawrence, D. M. and Swenson, S. C.: Improving canopy processes in the Community Land Model version 4 (CLM4) using global flux fields empirically inferred from FLUXNET data, J Geophys Res-Biogeo, 116, G02014, doi:10.1029/2010JG001593, 2011.

Breiman, L.: Random Forests, Mach. Learn., 45 (1), 5–32, doi:10.1023/A:1010933404324, 2001.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B., M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N. and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Q.J.R. Meteorol Soc, 137, 553–597, doi: 10.1002/qj.828, 2011.

Frank, D., Reichstein, M., Bahn, M., Thonicke, K., Frank, D., Mahecha, M. D., Smith, P., Van Der Velde, M., Vicca, S., Babst, F., Beer, C., Buchmann, N., Canadell, J. G., Ciais, P., Cramer, W., Ibrom, A., Miglietta, F., Poulter, B., Ramming, A., Seneviratne, S. I., Walz, A., Wattenbach, M., Zavala, M. A. and Zscheischler, J.: Effects of climate extremes on the terrestrial carbon cycle: concepts, processes and potential future impacts, Global Change Biol, 21, 2861–2880, doi: 10.1111/gcb.12916, 2015.

Friedman, J. H.: Multivariate Adaptive Regression Splines, Ann. Statist., 19, 1-67, doi:10.1214/aos/1176347963, 1991.

Fröhlich, B., Rodner, E., Kemmler, M. and Denzler, J.: Large-scale gaussian process classification using random decision forests, S. Mach. Perc., 22 (1), 113–120, DOI 10.1007/s00138-012-0480-y, 2012.

Gao, B. C.: NDWI-A Normalized difference water index for remote sensing of vegetation liquid water from space, Remote Sens Environ, 58, 257-266, doi:10.1016/S0034-4257(96)00067-3, 1996.

Garnaud, C., Sushama, L. and Arorab, V. K.: The effect of driving climate data on the simulated terrestrial carbon pools and fluxes over North America, Int J Climatol, 34, 1098-1110, DOI: 10.1002/joc.3748, 2014.

Giri, C., Zhu, Z. and Reed B.: A comparative analysis of the Global Land Cover 2000 and MODIS land cover data sets, Remote Sens Environ, 94, 123–132, doi:10.1016/j.rse.2004.09.005, 2005

Gupta, H. V., Kling, H., Yilmaz, K. K. and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, J Hydrol, 20, 80-91, doi:10.1016/j.jhydrol.2009.08.003, 2009.

Haykin, S.: Neural Networks – A Comprehensive Foundation (2nd ed.)., Prentice Hall., 1999.

Ho, T. K.: The Random Subspace Method for Constructing Decision Forests, IEEE T Pattern Anal, 20 (8), 832–844, doi:10.1109/34.709601, 1998.

Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X. and Ferreira, L.G.: Overview of the radiometric and biophysical performance of the MODIS vegetation indices, Remote Sens Environ, 83, 195–213, doi:10.1016/S0034-4257(02)00096-2, 2002.

Ichii, K., Wang, W., Hashimoto, H., Yang, F., Votava, P., Michaelis, A. R. and Nemani, R. R.: Refinement of rooting depths using satellite-based evapotranspiration seasonality for ecosystem modeling in California, Agr Forest Meteorol, 149, 1907-1918, doi:10.1016/j.agrformet.2009.06.019, 2009.

Ichii, K., Suzuki, T., Kato, T., Ito, A., Hajima, T., Ueyama, M., Sasai, T., Hirata, R., Saigusa, N., Ohtani, Y. and Takagi, K.: Multi-model analysis of terrestrial carbon cycles in Japan: limitations and implications of model calibration using eddy flux observations, Biogeosciences, 7, 2061–2080, doi:10.5194/bg-7-2061-2010, 2010.

IPCC: Climate Change 2007: Synthesis Report. Contribution of Working Groups I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, Pachauri, R.K and Reisinger, A. (eds.)]. IPCC, Geneva, Switzerland, 104, 2007.

Joiner, J., Yoshida, Y., Vasilkov, A. P., Schaefer, K., Jung, M., Guanter, L., Zhang, Y., Garrity, S., Middleton, E. M., Huemmrich, K. F., Guh, L. and Belelli Marchesini, L.: The seasonal cycle of satellite chlorophyll fluorescence observations and its relationship to vegetation phenology and ecosystem atmosphere carbon exchange, Remote Sens Environ, 152, 375–391, http://dx.doi.org/10.1016/j.rse.2014.06.022, 2014.

Jung, M., and Zscheischler, J.: A Guided Hybrid Genetic Algorithm for Feature Selection with Expensive Cost Functions, Procedia Computer Science, 18, 2337-2346, doi: 10.1016/j.procs.2013.05.405, 2013.

Jung, M., Verstraete, M., Gobronz, N., Reichstein, M., Papale, D., Bondeau, A., Robustelli, M. and Pinty, R.: Diagnostic assessment of European gross primary production, Global Change Biol, 14, 2349–2364, doi: 10.1111/j.1365-2486.2008.01647.x, 2008.

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

Jung, M., Reichstein, M., Ciais, P., Seneviratne, S. I., Sheffield, J., Goulden, M. L., Bonan, G., Cescatti, A., Chen, J., de Jeu, R., Dolman, A. J., Eugster, W., Gerten, D., Gianelle, D., Gobron, N., Heinke, J., Kimball, J., Law, B. E., Montagnani, L., Mu, Q., Mueller, B., Oleson, K., Papale, D., Richardson, A. D., Roupsard, O., Running, S., Tomelleri, E., Viovy, N., Weber, U., Williams, C., Wood, E., Zaehle, S. and Zhang, K.: Recent decline in the global land evapotranspiration trend due to limited moisture supply, Nature Letter, 467, 951-953, 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, Journal of geophys res-Biogeo, 116, G00J07, doi:10.1029/2010JG001566, 2011.

Justice, C. O., Townshend, J. R. G., Vermote, E. F., Masuoka, E., Wolfe, R. E., Saleous, N., Roy, D. P. and Morisette, J. T.: An overview of MODIS Land data processing and product status, Remote Sens Environ, 83, 3–15, doi:10.1016/S0034-4257(02)00084-6, 2002.

Kondo, M., Ichii, K., Takagi, H. and Sasakawa, M.: Comparison of the data-driven top-down and bottom-up global terrestrial CO2 exchanges: GOSAT CO2 inversion and empirical eddy flux upscaling, Journal of geophys res-Biogeo, 120, 1226–1245, doi:10.1002/2014JG002866, 2015.

Lasslop, G., Reichstein, M., Papale, D., Richardson, A. D., Arneth, A., Barr, A., Stoy, P. and Wohlfahrt, G.: Separation of net ecosystem exchange into assimilation and respiration using a light response curve approach: critical issues and global evaluation, Global Change Biol, 16, 187-208, doi:10.1111/j.1365-2486.2009.02041.x, 2010.

Merbold, L., Ardö, J., Arneth, A., Scholes, R. J., Nouvellon, Y., de Grandcourt, A., Archibald, S., Bonnefond, J. M., Boulain, N., Brueggemann, N., Bruemmer, C., Cappelaere, B., Ceschia, E., El-Khidir, H. A. M., El-Tahir, B. A., Falk, U., Lloyd, J., Kergoat, L., Le Dantec, V., Mougin, E., Muchinda, M., Mukelabai, M. M., Ramier, D., Roupsard, O., Timouk, F., Veenendaal,E. M. and Kutsch, W. L.: Precipitation as driver of carbon fluxes in 11 African ecosystems, Biogeosciences, 6, 1027–1041, doi:10.5194/bg-6-1027-2009, 2009, 2009.

Myneni, R.B., Hoffman, S., Knyazikhin, Y., Privette, J.L., Glassy, J., Tian, Y., Wang, Y., Song, X., Zhang, Y., Smith, G.R., Lotsch, A., Friedl, M., Morisette, J.T., Votava, P., Nemani, R.R., and Running, S.W.: Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data, Remote Sens Environ, 83, 214–231, doi:10.1016/S0034-4257(02)00074-3, 2002.

Nash, J. E. and Sutcliffe J. V.: River flow forecasting through conceptual models part I: A discussion of principles, Journal Hydrol, 10, 282–290, doi:10.1016/0022-1694(70)90255-6, 1970.

Papale, D. and Valentini, R.: A new assessment of European forests carbon exchanges by eddy fluxes and artificial neural network spatialization, Global Change Biol, 9, 525–535, doi: 10.1046/j.1365-2486.2003.00609.x, 2003.

Papale, D., Reichstein, M., Aubinet, M., Canfora, E., Bernhofer, C., Kutsch, W., Longdoz, B., Rambal, S., Valentini, R., Vesala, T. and Yakir, D.: Towards a standardized processing of Net Ecosystem Exchange measured with eddy covariance technique: Algorithms and uncertainty estimation, Biogeosciences, 3, 571–583, doi:10.5194/bg-3-571-2006, 2006.

Papale, D., Black, T. A., Carvalhais, N., Cescatti, A., Chen, J., Jung, M., Kiely, G., Lasslop, G., Mahecha, D. M., Margolis, H., Merbold, L., Montagnani, L., Moors, E., Olesen, J. E., Reichstein, M., Tramontana, G., van Gorsel, E., Wohlfahrt, G. and Ráduly, B.: Effect of spatial sampling from European flux towers for estimating carbon and water fluxes with artificial neural networks, Journal of geophys res-Biogeo, 120, 1941-1957,doi: 10.1002/2015JG00299, 2015.

Paruelo, J. M., Piñero, G., Oyonarte, C., Alcaraz, D., Cabello, J. and Escribano, P.: Temporal and spatial patterns of ecosystem functioning in protected arid areas in southeastern Spain, Appl Veg Sci, 8, 93–102, doi: http://dx.doi.org/10.1658/1402-2001(2005)008[0093:TASPOE]2.0.CO;2, 2005.

Piao, S., Sitch, S., Ciais, P., Friedlingstein, P., Peylin, P., Wang, X., Ahlström, A., Anav, A., Canadell, J. G., Cong, N., Huntingford, C., Jung, M., Levis, S., Levy, P. E., Li, J., Lin, X., Lomas, M. R., Lu, M., Luo, Y., Ma, Y., Myneni, R. B., Poulter, B., Sun, Z., Wang, T., Viovy, N., Zaehle, S. and Zeng, N.: Evaluation of terrestrial carbon cycle models for their response to climate variability and to CO2 trends, Glob Change Biol., 19, 2117–2132, doi:10.1111/gcb.12187, 2013.

Rasmussen C. E. and Williams C. K. I.: Gaussian Processes for Machine Learning, the MIT Press, ISBN 026218253X, 2006.

Reichstein, M., Falge, E., Baldocchi, D., Papale, D., Aubinet, M., Berbigier, P., Bernhofer, C., Buchmann, N., Gilmanov, T., Granier, A., Grünwald, T., Havránková, K., Ilvesniemi, 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, Global Change Biol, 11, 1424–1439, doi:10.1111/j.1365-2486.2005.001002.x, 2005.

Reich, P. B.: The carbon dioxide exchange, Science, 329, 774-775, http://dx.doi.org/10.1126/science.1194353, 2010.

Schwalm C. R., Williams C. A., Schaefer K., Anderson R., Arain M. A., Baker I., Barr A., Black T. A., Chen G., Chen J. M., Ciais P., Davis K. J., Desai A., Dietze M., Dragoni D., Fischer M. L., Flanagan L. B., Grant R., Gu L., Hollinger D., Izaurralde R. C., Kucharik C., Lafleur P., Law B. E., Li L., Li Z., Liu S., Lokupitiya E., Luo Y., Ma S., Margolis H., Matamala R., McCaughey H., Monson R. K., Oechel W. C., Peng C., Poulter B., Price D. T., Riciutto D. M., Riley W., Sahoo A. K., Sprintsin M., Sun J., TI. H., Tonitto C., Verbeeck H., Verma S. B.: A model-data intercomparison of CO2 exchange across North America: Results from the North American Carbon Program site synthesis, Journal of geophys res-Biogeo, 115, G00H05, doi:10.1029/2009JG001229, 2010.

Schaaf, C. B., Gao, F., Strahler, A. H., Lucht, W., Li, X., Tsang, T., Strugnell, N. C., Zhang, X., Jin, Y., Muller, J-P., Lewis, P., Barnsley, M., Hobson, P., Disney, M., Roberts, G., Dunderdale, M., Doll, C., d'Entremont, R. P., Hu, B., Liang, S., Privette, J. L. and Roy, D.: First operational BRDF, albedo nadir reflectance products from MODIS, Remote Sens Environ, 83, 135–148, doi:10.1016/S0034-4257(02)00091-3, 2002.

Shawe-Taylor, J. and Cristianini, N.: Kernel Methods for Pattern Analysis, Cambridge University Press, 2004.

Shirmohammadi, R., Ghorbani, B., Hamedi, M., Hamedi, M. H., and Romeo, L. M.: Optimization of mixed refrigerant systems in low temperature applications by means of group method of data handling (GMDH), Journal of Natural Gas Science and Engineering, 26, 303-312, doi:10.1016/j.jngse.2015.06.028, 2015.

Sims, D. A., Rahman, A. F., Cordova, V. D., El-Masri, B. Z., Baldocchi, D. D., Bolstad, P. V., Flanagan, L. B., Goldstein, A. H., Hollinger, D. Y., Misson, L., Monson, R. K., Oechel, W. C., Schmid, H. P., Wofsy, S. C. and Xu, L.: A new model of gross primary productivity for North American ecosystems based solely on the enhanced vegetation index and land surface temperature from MODIS, Remote Sens Environ, 12, 1633–1646, doi:10.1016/j.rse.2007.08.004, 2008.

Thornton, P. E., Law, B. E., Gholz, H. L., Clark, K. L., Falge, E., Ellsworth, D. S., Goldstein, A. H., Monson, R. K., Hollinger, D., Falk, M., Chen, J., Sparks, J. P.: Modeling and measuring the effects of disturbance history and climate on carbon and water budgets in evergreen needleleaf forests, Agr Forest Meteorol, 113, 185-222, doi:10.1016/S0168-1923(02)00108-9, 2002.

Tramontana, G., Ichii, K., Camps-Valls, G., Tomelleri, E. and Papale, D.: Uncertainty analysis of gross primary production upscaling using Random Forests, remote sensing and eddy covariance data, Remote Sens Environ, 168, 360–373, doi:10.1016/j.rse.2015.07.015, 2015.

Ungaro, F., Calzolari, C. and Busoni, E: Development of pedotransfer functions using a group method of data handling for the soil of the Pianura Padano–Veneta region of North Italy: water retention properties, Geoderma, 124, 293–317, doi:10.1016/j.geoderma.2004.05.007, 2005.

Valentini, R., Arneth, A., Bombelli, A., Castaldi, S., Cazzolla Gatti, R., Chevallier, F., Ciais, P., Grieco, E., Hartmann, J., Henry, M., Houghton, R. A., Jung, M., Kutsch, W. L., Malhi, Y., Mayorga, E., Merbold, L., Murray-Tortarolo, G., Papale, D., Peylin, P., Poulter, B., Raymond, P. A., Santini, M., Sitch, S., Vaglio Laurin, G., van der Werf, G. R., Williams, C. A. and Scholes, R. J.: A full greenhouse gases budget of Africa: synthesis, uncertainties, and vulnerabilities, Biogeosciences, 11, 381-407, doi:10.5194/bg-11-381-2014., 2014.

Vapnik, V., Golowich, S. and Smola, A.: Support vector method for function approximation, regression estimation, and signal processing, Adv Neur In, 9, 281–287, 1997.

Verrelst, J., Muñoz, J., Alonso, L., Delegido, J., Rivera, J., Camps-Valls, G., and Moreno, J.: Machine learning regression algorithms for biophysical parameter retrieval: Opportunities for Sentinel-2 and -3, Remote Sens Environ, 118, 127–139, 2012

Verrelst, J., Camps-Valls, G., Muñoz, J., Rivera, J. P., Veroustraete, F., Clevers, J. P. G. W. and Moreno, J.: Optical remote sensing and the retrieval of terrestrial vegetation bio-geophysical properties – A review, ISPRS J Photogramm, doi:10.1016/j.isprsjprs.2015.05.005, 108, 273-290, 2015

Wan, Z., Zhang, Y., Zhang, Q. and Li, Z. L.: Validation of the land-surface temperature products retrieved from Terra Moderate Resolution Imaging Spectroradiometer data, Remote Sens Environ, 83, 163−180, doi: 10.1016/S0034-4257(02)00093-7, 2002.

Wolf, S., Eugster, W., Potvin, C. and Buchmann, N.: Strong seasonal variations in net ecosystem CO2 exchange of a tropical pasture and afforestation in Panama, Agr Forest Meteorol, 151, 1139-1151, doi:10.1016/j.agrformet.2011.04.002, 2011.

Xiao, J., Zhuang, Q., Baldocchi, D. D., Law, B. E., Richardson, A. D., Chen, J., Oren, R., Starr, G., Noormets, A., Ma, S., Verma, S. B., Wharton, S., Wofsy, S. C., Bolstad, P. V., Burns, S. P., Cook, D. R., Curtis, P. S., Drake, B. G., Falk, M., Fischer, M. L., Foster, D. R., Gu, L., Hadley, J. L., Hollinger, D. Y., Katul, G. G., Litvak, M., Martin, T. A., Matamala, R., McNulty, S., Meyers, T. P., Monson, R. K., Munger, J. W., Oechel, W. C., Paw U, K. T., Schmid, H. P., Scott, R. L., Sun, G., Suyker, A. E., Torn, M. S.: Estimation of net ecosystem carbon exchange for the conterminous United States by combining MODIS and AmeriFlux data, Agr Forest Meteorol, 148, 1827-1847, doi:10.1016/j.agrformet.2008.06.015, 2008.

Xiao, J., Zhuang, Q., Law, B. E., Chen, J., Baldocchi, D. D., Cook, D. R., Oren, R., Richardson, A. D., Wharton, S., Ma, S., Martin, T. A., Verma, S. B., Suyker, A. E., Scott, R. L., Monson, R. K., Litvak, M., Hollinger, D. Y., Sun, G., Davis, K. J., Bolstad, P. V., Burns, S. P., Curtis, P. S., Drake, B. G., Falk, M., Fischer, M. L., Foster, D. R., Gu, L., Hadley, J. L., Katul, G. G., Matamala, R., McNulty, S., Meyers, T., P., Munger, J. W., Noormets, A., Oechel, W. C., Paw, K. T., Schmid, H. P., Starr, G. Torn, M. S. and Wofsy, S. C.: A continuous measure of gross primary production for the conterminous United States derived from MODIS and AmeriFlux data, Remote Sens Environ, 114, 576–591, doi: 10.1016/j.rse.2009.10.013, 2010.

Xiao, X., Boles, S., Liu, J. Y., Zhuang, D. F. and Liu, M. L.: Characterization of forest types in Northeastern China, using multi-temporal SPOT-4 VEGETATION sensor data, Remote Sens Environ, 82, 335–348, doi:10.1016/S0034-4257(02)00051-2, 2002.

Yang, F., White, M. A., Michaelis, A. R., Ichii, K., Hashimoto, H., Votava, P., Zhu, A. X. and Nemani, R. R.: Prediction of continental-scale evapotranspiration by combining MODIS and AmeriFlux data through support Vector machine, IEEE T. Geosci Remote, 44, 3452-3461, doi: 10.1109/TGRS.2006.876297, 2006.

Yang, F., Ichii, K., White, M. A., Hashimoto, H., Michaelis, A. R., Votava, P., Zhu, A-X., Huete, A., Running, S. W., and Nemani, R. R.: Developing a continental-scale measure of gross primary production by combining MODIS and AmeriFlux data through Support Vector Machine approach, Remote Sens Environ, 110, 109–122, doi:10.1016/j.rse.2007.02.016, 2007.

Yebra, M., Van Dijk, A. I. J. M., Leuning, R. and Guerschman, J. P.: Global vegetation gross primary production estimation using satellite-derived light-use efficiency and canopy conductance, Remote Sens Environ, 163, 206–216, doi:10.1016/j.rse.2015.03.016, 2015.

Yuan, W., Liu, S., Yu, G., Bonnefond, J-M., Chen, J., Davis, K., Desai, A. R., Goldstein, A. H., Gianelle, D., Rossi, F., Suyker, A. E. and Verma, S. B.: Global estimates of evapotranspiration and gross primary production based on MODIS and global meteorology data, Remote Sens Environ, 114, 1416–1431, doi:10.1016/j.rse.2010.01.022, 2010.

Zhao, Y., Ciais, P., Pylin, P., Viovy, N., Longdoz, B., Bonnefond, J. M., Rambal, S., Klumpp, K., Olioso, A., Cellier, P., Maigna, F., Eglin, T. and Calvet, J. C.: How errors on meteorological variables impact simulated ecosystem fluxes: a case study for six French sites, Biogeosciences, 9, 2537-2564, doi:10.5194/bg-9-2537-2012, 2012.

Table 1. Distribution of flux tower sites across plant functional types (PFT) and climate zones.

PFT	**N° of sites**	**Climate zone**	**N° of sites**
Evergreen needleleaf forest	66	Temperate	111
Grassland	38	Subtropical - Mediterranean	47
Cropland	27	Boreal	34
Deciduous broadleaf forest	24	Tropical	14
Evergreen broadleaf forest	19	Dry	13
Wetland	17	Artic	5
Shrubland	12		
Mixed forest	11		
Savannah	10		

Table 2. Selected predictors for both setup for carbon fluxes (GPP, TER and NEE) and energy fluxes (H, LE and Rn). List of acronyms: Enhanced Vegetation Index (EVI), fraction of photosynthetically active radiation absorbed by a canopy (fPAR), Leaf Area Index (LAI), daytime Land Surface Temperature (LST_{Day}) and nighttime Land Surface Temperature (LST_{Night}), Middle Infrared Reflectance (band 7) (MIR[(1)]), Normalized Difference Vegetation Index (NDVI), Normalized Difference Water Index (NDWI), Plant Functional Type (PFT), incoming global Radiation (Rg), top of atmosphere potential Radiation (Rpot), Index of Water Availability (IWA), Relative humidity (Rh), Water Availability Index lower (WAI_L), and upper (WAI_U) (for details see supplementary material, Sect. S3), Mean Seasonal Cycle (MSC). Interaction between A and B is shown as (A, B)

Setup	Type of variability	Carbon fluxes	Energy fluxes
RS	Spatial	PFT	PFT
		Amplitude of MSC of EVI	Maximum of MSC of (fPAR, Rg)
		Amplitude of MSC of MIR[(1)]	Minimum of MSC of (NDVI, Rg)
		Maximum of MSC of LST_{Day}	
	Spatial & seasonal	MSC LAI	MSC of (EVI, LST)
			Rpot
	Spatial, seasonal & interannual	NDWI	Rg
		LST_{Day}	LST_{Day}
		LST_{Night}	Anomalies of LST_{Night}
		(NDVI, Rg)	Anomalies of (EVI, LST)
RS+METEO	Spatial	PFT	PFT
		Amplitude of MSC of NDVI	Maximum of MSC of (NDVI, WAI_U)
		Amplitude of MSC of band	Mean of MSC of band 6

		4 BRDF reflectance[2]	BRDF reflectance[2]
		Minimum of MSC of NDWI	
		Amplitude of MSC of (NDVI, WAI_L)	
	Spatial & Seasonal	MSC of LST_{Night}	Rpot
		MSC of (FPAR, LST)	MSC of NDWI
		MSC of (NDVI, Rg)	MSC of LST_{Night}
			MSC of (NDVI, Rg, IWA)
	Spatial & Seasonal & Interannual	Tair	Rain
			Rg
			Rh

[1] derived from the MOD13 product; [2] derived from MCD43 product.

Table 3. Statistics of the accuracy of predictions of carbon and energy fluxes made by the ensemble median estimate based on RS and RS+METEO. For RMSE and BIAS, the reference units are $gCm^{-2}d^{-1}$ and $MJm^{-2}d^{-1}$, in the case of carbon fluxes (GPP, TER and NEE) and energy fluxes (H, LE and Rn) respectively.

Flux	RS					RS+METEO				
	MEF	RMSE	P	ROV	BIAS	MEF	RMSE	ρ	ROV	BIAS
GPP_R	0.71	1.56	0.85	0.69	-0.02	0.70	1.59	0.84	0.73	0.09
GPP_L	0.71	1.53	0.84	0.68	-0.02	0.71	1.54	0.84	0.74	0.09
TER_R	0.64	1.14	0.80	0.61	-0.01	0.64	1.15	0.80	0.69	0.09
TER_L	0.60	1.18	0.77	0.56	-0.01	0.63	1.14	0.79	0.66	0.08
NEE	0.46	1.24	0.68	0.39	0.04	0.43	1.28	0.65	0.40	-0.02
H	0.79	1.36	0.89	0.71	-0.02	0.79	1.37	0.89	0.75	0.02
LE	0.76	1.37	0.87	0.71	-0.07	0.75	1.39	0.87	0.73	-0.01
Rn	0.92	1.51	0.96	0.90	-0.01	0.91	1.55	0.96	0.93	0.08

Table 4: Mean values of the determination coefficient (R^2) by the pair-wise comparison of the models output and their residuals. We compared different ML and same drivers (RS and RS+METEO respectively) or the same ML and different drivers (RS vs RS+METEO). Numbers in brackets are the standard deviation of R^2. All correlations are statistically significant ($p < 0.001$).

Fluxes	Correlation ~~among~~ between model~~s~~ output s			Correlation ~~among~~ between model~~s~~ residuals		
	RS	RS+METEO	RS vs RS+METEO	RS	RS+METEO	RS vs RS+METEO
GPP_R	0.95 (0.02)	0.95 (0.02)	0.89 (0.02)	0.88 (0.04)	0.87 (0.04)	0.74 (0.04)
GPP_L	0.95 (0.02)	0.94 (0.02)	0.88 (0.02)	0.88 (0.04)	0.86 (0.04)	0.72 (0.04)
TER_R	0.91 (0.03)	0.94 (0.03)	0.86 (0.04)	0.86 (0.05)	0.88 (0.05)	0.75 (0.06)
TER_L	0.92 (0.03)	0.93 (0.03)	0.85 (0.03)	0.89 (0.04)	0.88 (0.05)	0.77 (0.05)
NEE	0.84 (0.06)	0.84 (0.07)	0.75 (0.08)	0.88 (0.05)	0.87 (0.06)	0.80 (0.06)
H	0.94 (0.02)	0.96 (0.02)	0.93 (0.03)	0.80 (0.06)	0.87 (0.05)	0.76 (0.08)
LE	0.94 (0.02)	0.96 (0.01)	0.90 (0.02)	0.83 (0.05)	0.88 (0.04)	0.73 (0.04)
Rn	0.98 (0.01)	0.99 (0.00)	0.97 (0.01)	0.79 (0.08)	0.86 (0.03)	0.71 (0.12)

Table 5: R^2 and RMSE for the comparison ~~among~~ between sites, mean seasonal cycle and anomalies. The last two columns show the consistency between the median estimates of the two setups. For RMSE, the reference units are $gCm^{-2}d^{-1}$ and $MJm^{-2}d^{-1}$, in the case of carbon fluxes (GPP, TER and NEE) and energy fluxes (H, LE and Rn) respectively.

Fluxes	RS *vs*. OBS		RS+METEO *vs*. OBS		RS *vs*. RS+METEO	
	R^2	RMSE	R^2	RMSE	R^2	RMSE
Across-sites						
GPP_R	0.78	0.80	0.77	0.82	0.95	0.34
GPP_L	0.78	0.77	0.79	0.75	0.94	0.36
TER_R	0.68	0.73	0.61	0.81	0.92	0.32
TER_L	0.72	0.60	0.71	0.61	0.92	0.27
NEE	0.48	0.61	0.46	0.61	0.83	0.22
H	0.81	0.68	0.81	0.68	0.97	0.25
LE	0.79	0.74	0.75	0.80	0.93	0.33
Rn	0.80	0.93	0.79	0.96	0.96	0.38
Mean Seasonal Cycle						
GPP_R	0.76	1.03	0.77	1.02	0.93	0.48
GPP_L	0.77	1.00	0.77	0.99	0.93	0.50
TER_R	0.71	0.62	0.71	0.62	0.92	0.29
TER_L	0.67	0.64	0.68	0.63	0.92	0.29
NEE	0.61	0.83	0.59	0.84	0.93	0.24
H	0.86	0.89	0.86	0.87	0.97	0.36
LE	0.87	0.79	0.87	0.79	0.95	0.45
Rn	0.98	0.74	0.98	0.74	0.99	0.43
Anomalies						
GPP_R	0.18	0.67	0.12	0.68	0.38	0.32

GPP_L	0.16	0.67	0.11	0.68	0.37	0.31
TER_R	0.14	0.48	0.15	0.48	0.36	0.17
TER_L	0.10	0.58	0.13	0.57	0.35	0.18
NEE	0.13	0.56	0.13	0.55	0.43	0.20
H	0.43	0.81	0.41	0.81	0.77	0.34
LE	0.21	0.78	0.21	0.77	0.46	0.32
Rn	0.57	0.81	0.54	0.83	0.84	0.41

Table A1: Accuracy of carbon and energy fluxes predicted by machine learning methods based on RS and RS+METEO dataset. The median value of methods and the variability (in brackets) estimated as the median absolute deviation (MAD) from the median multiplied per 1.4826 (as reported in Jung et al., 2009) were reported.

FLUXES	RS			RS+METEO		
	MEF	RMSE	Abs BIAS	MEF	RMSE	Abs BIAS
GPP	0.698 (±0.012)	1.604 (±0.033)	0.022 (±0.019)	0.694 (±0.012)	1.614 (±0.032)	0.073 (±0.011)
GPP_{HB}	0.700 (±0.009)	1.564 (±0.024)	0.023 (±0.024)	0.701 (±0.008)	1.561 (±0.020)	0.083 (±0.011)
TER	0.612 (±0.022)	1.183 (±0.033)	0.026 (±0.025)	0.623 (±0.005)	1.166 (±0.008)	0.089 (±0.033)
TER_{HB}	0.571 (±0.016)	1.218 (±0.023)	0.019 (±0.017)	0.609 (±0.001)	1.163 (±0.002)	0.079 (±0.017)
NEE	0.433 (±0.017)	1.270 (±0.019)	0.024 (±0.021)	0.407 (±0.029)	1.298 (±0.032)	0.014 (±0.003)
H	0.767 (±0.015)	1.426 (±0.047)	0.014 (±0.005)	0.776 (±0.008)	1.397 (±0.025)	0.022 (±0.009)
LE	0.739 (±0.015)	1.418 (±0.042)	0.052 (±0.046)	0.734 (±0.003)	1.434 (±0.009)	0.023 (±0.008)
Rn	0.909 (±0.009)	1.589 (±0.082)	0.030 (±0.025)	0.908 (±0.008)	1.600 (±0.070)	0.073 (±0.015)

Table C1. Median site-by-site R^2 of the carbon fluxes per PFT and climate zones. ENF is vvergreen needle leaf forest, DBF deciduous broadleaf forest, EBF Evergreen broadleaf forest, MF mixed forest, SHR shrubland, SAV Savannah, GRA Grassland, CRO cropland, WET Wetland, Trop Tropical, SubTrop Subtropical, Dry Dryland, Tmp Temperate, TmpCont Temperate-continental, Bor Boreal, Cold cold environment or Iceland covered by ice.

CAT	GPP_R		GPP_L		TER_R		TER_L		NEE	
	RS	RS+METEO	RS	RS+METEO	RS	RS+METEO	RS	RS+METEO	RS	RS+METEO
ENF	0.87 (0.10)	0.86 (0.10)	0.85 (0.12)	0.86 (0.12)	0.81 (0.15)	0.85 (0.11)	0.75 (0.24)	0.76 (0.20)	0.50 (0.34)	0.55 (0.30)
DBF	0.89 (0.07)	0.87 (0.09)	0.87 (0.07)	0.88 (0.08)	0.81 (0.12)	0.83 (0.13)	0.76 (0.14)	0.76 (0.14)	0.72 (0.16)	0.68 (0.17)
EBF	0.50 (0.29)	0.48 (0.20)	0.48 (0.29)	0.44 (0.28)	0.34 (0.34)	0.49 (0.35)	0.15 (0.18)	0.29 (0.20)	0.26 (0.23)	0.24 (0.26)
MF	0.91 (0.06)	0.95 (0.02)	0.91 (0.03)	0.95 (0.04)	0.85 (0.10)	0.90 (0.07)	0.84 (0.10)	0.86 (0.15)	0.73 (0.10)	0.75 (0.09)
SHR	0.67 (0.30)	0.71 (0.28)	0.67 (0.36)	0.72 (0.23)	0.80 (0.13)	0.78 (0.24)	0.68 (0.18)	0.66 (0.38)	0.37 (0.38)	0.41 (0.31)
SAV	0.75 (0.13)	0.70 (0.13)	0.72 (0.05)	0.67 (0.17)	0.65 (0.07)	0.72 (0.11)	0.55 (0.16)	0.61 (0.10)	0.38 (0.20)	0.34 (0.29)
GRA	0.69 (0.27)	0.62 (0.33)	0.69 (0.25)	0.60 (0.32)	0.70 (0.25)	0.73 (0.25)	0.66 (0.20)	0.72 (0.21)	0.40 (0.29)	0.36 (0.30)
CRO	0.58 (0.41)	0.44 (0.36)	0.56 (0.41)	0.45 (0.31)	0.78 (0.17)	0.76 (0.15)	0.68 (0.22)	0.65 (0.23)	0.35 (0.46)	0.33 (0.43)
WET	0.87 (0.11)	0.91 (0.07)	0.85 (0.12)	0.87 (0.09)	0.78 (0.19)	0.83 (0.14)	0.65 (0.17)	0.74 (0.20)	0.64 (0.16)	0.61 (0.24)
Trop	0.32 (0.46)	0.40 (0.39)	0.63 (0.23)	0.31 (0.32)	0.25 (0.23)	0.34 (0.47)	0.11 (0.13)	0.26 (0.14)	0.28 (0.35)	0.21 (0.30)
SubTrop	0.64 (0.26)	0.66 (0.28)	0.65 (0.26)	0.65 (0.24)	0.64 (0.25)	0.66 (0.26)	0.52 (0.24)	0.55 (0.28)	0.39 (0.37)	0.39 (0.26)
Dry	0.47 (0.27)	0.40 (0.33)	0.50 (0.25)	0.38 (0.30)	0.62 (0.25)	0.62 (0.38)	0.55 (0.19)	0.55 (0.39)	0.21 (0.29)	0.11 (0.14)
Tmp	0.81 (0.19)	0.74 (0.24)	0.83 (0.14)	0.78 (0.22)	0.78 (0.13)	0.77 (0.18)	0.68 (0.20)	0.72 (0.17)	0.56 (0.28)	0.47 (0.34)
TmpCont	0.86 (0.09)	0.82 (0.16)	0.84 (0.11)	0.80 (0.17)	0.81 (0.12)	0.78 (0.14)	0.75 (0.17)	0.76 (0.15)	0.54 (0.42)	0.53 (0.36)
Bor	0.90 (0.07)	0.90 (0.07)	0.92 (0.06)	0.89 (0.07)	0.90 (0.05)	0.91 (0.04)	0.86 (0.08)	0.89 (0.06)	0.59 (0.31)	0.59 (0.25)
Cold	0.56 (0.57)	0.50 (0.56)	0.49 (0.62)	0.46 (0.59)	0.84 (0.20)	0.86 (0.13)	0.50 (0.38)	0.55 (0.23)	0.47 (0.56)	0.45 (0.57)

Table C2. Median site-by-site RMSE of the carbon fluxes per PFT and climate zones. ENF is svergreen needle leaf forest, DBF deciduous broadleaf forest, EBF Evergreen broadleaf forest, MF mixed forest, SHR shrubland, SAV Savannah, GRA Grassland, CRO cropland, WET Wetland, Trop Tropical, SubTrop Subtropical, Dry Dryland, Tmp Temperate, TmpCont Temperate-continental, Bor Boreal, Cold cold environment or Iceland covered by ice.

CAT	GPP_R (gCm^{-2}d^{-1})		GPP_L (gCm^{-2}d^{-1})		TER_R (gCm^{-2}d^{-1})		TER_L (gCm^{-2}d^{-1})		NEE (gCm^{-2}d^{-1})	
	RS	RS+METEO	RS	RS+METEO	RS	RS+METEO	RS	RS+METEO	RS	RS+METEO
ENF	1.05 (0.60)	1.12 (0.60)	1.04 (0.59)	1.14 (0.66)	0.82 (0.50)	0.80 (0.52)	0.87 (0.60)	0.91 (0.68)	0.87 (0.51)	0.86 (0.53)
DBF	1.21 (0.78)	1.35 (0.59)	1.17 (0.68)	1.36 (0.62)	0.68 (0.26)	0.76 (0.33)	0.76 (0.33)	0.93 (0.44)	1.28 (0.39)	1.28 (0.39)
EBF	1.70 (0.55)	1.64 (0.85)	1.65 (0.70)	1.46 (0.51)	1.23 (0.69)	1.48 (0.85)	1.88 (1.23)	1.71 (0.73)	1.15 (0.48)	1.15 (0.45)
MF	0.87 (0.17)	0.76 (0.45)	0.89 (0.27)	0.97 (0.56)	0.65 (0.18)	0.73 (0.42)	0.79 (0.14)	0.79 (0.18)	0.91 (0.47)	0.81 (0.29)
SHR	0.73 (0.47)	0.78 (0.46)	0.69 (0.44)	0.77 (0.37)	0.50 (0.33)	0.70 (0.41)	0.50 (0.34)	0.55 (0.36)	0.57 (0.31)	0.52 (0.15)
SAV	0.83 (0.44)	0.81 (0.18)	0.87 (0.45)	0.84 (0.18)	0.80 (0.53)	0.68 (0.41)	0.86 (0.55)	0.77 (0.38)	0.71 (0.36)	0.69 (0.31)
GRA	1.22 (0.64)	1.22 (0.60)	1.18 (0.68)	1.20 (0.62)	1.00 (0.48)	1.01 (0.54)	0.99 (0.58)	0.95 (0.52)	0.76 (0.61)	0.85 (0.49)
CRO	1.69 (1.38)	2.30 (1.02)	1.57 (1.42)	2.24 (1.10)	0.87 (0.46)	0.90 (0.57)	0.80 (0.51)	0.98 (0.57)	1.42 (0.90)	1.44 (0.70)
WET	1.04 (0.95)	0.93 (0.77)	1.03 (0.96)	0.78 (0.53)	1.04 (0.87)	0.98 (0.82)	1.07 (0.51)	1.02 (0.51)	0.46 (0.19)	0.64 (0.26)
Trop	1.93 (0.46)	1.74 (1.01)	2.24 (0.62)	1.56 (0.78)	2.07 (0.69)	1.55 (0.87)	2.47 (0.74)	2.05 (0.43)	1.28 (0.29)	1.17 (0.46)
SubTrop	1.37 (0.55)	1.40 (0.61)	1.37 (0.56)	1.38 (0.57)	1.03 (0.46)	1.00 (0.41)	1.08 (0.36)	1.11 (0.40)	1.13 (0.63)	1.15 (0.62)
Dry	0.60 (0.24)	0.78 (0.36)	0.63 (0.16)	0.74 (0.30)	0.49 (0.10)	0.54 (0.20)	0.58 (0.26)	0.67 (0.32)	0.41 (0.13)	0.46 (0.15)
Tmp	1.73 (1.02)	1.82 (0.99)	1.73 (0.98)	1.71 (1.03)	1.09 (0.54)	1.17 (0.67)	1.24 (0.57)	1.31 (0.59)	1.43 (0.59)	1.40 (0.61)
TmpCont	1.01 (0.42)	1.29 (0.59)	1.00 (0.45)	1.26 (0.57)	0.71 (0.30)	0.75 (0.38)	0.74 (0.31)	0.79 (0.34)	0.95 (0.39)	1.02 (0.43)
Bor	0.66 (0.27)	0.70 (0.36)	0.66 (0.27)	0.67 (0.33)	0.48 (0.27)	0.47 (0.27)	0.48 (0.16)	0.45 (0.21)	0.50 (0.32)	0.48 (0.22)
Cold	0.44 (0.04)	0.58 (0.42)	0.51 (0.24)	0.46 (0.32)	0.41 (0.06)	0.23 (0.06)	0.57 (0.16)	0.29 (0.12)	0.51 (0.21)	0.54 (0.35)

Table C3. Median site-by-site absolute bias of the carbon fluxes per PFT and climate zones. ENF is evergreen needle leaf forest, DBF deciduous broadleaf forest, EBF Evergreen broadleaf forest, MF mixed forest, SHR shrubland, SAV Savannah, GRA Grassland, CRO cropland, WET Wetland, Trop Tropical, SubTrop Subtropical, Dry Dryland, Tmp Temperate, TmpCont Temperate-continental, Bor Boreal, Cold cold environment or Iceland covered by ice.

CAT	GPP_R (gCm^{-2}d^{-1})		GPP_L (gCm^{-2}d^{-1})		TER_R (gCm^{-2}d^{-1})		TER_L (gCm^{-2}d^{-1})		NEE (gCm^{-2}d^{-1})	
	RS	RS+METEO	RS	RS+METEO	RS	RS+METEO	RS	RS+METEO	RS	RS+METEO
ENF	0.53 (0.46)	0.54 (0.56)	0.45 (0.42)	0.48 (0.50)	0.47 (0.47)	0.50 (0.54)	0.42 (0.40)	0.41 (0.43)	0.39 (0.44)	0.32 (0.36)
DBF	0.43 (0.38)	0.56 (0.59)	0.42 (0.36)	0.50 (0.52)	0.29 (0.32)	0.35 (0.35)	0.39 (0.33)	0.42 (0.34)	0.60 (0.28)	0.55 (0.30)
EBF	0.82 (0.91)	0.77 (0.50)	0.75 (0.81)	0.76 (0.48)	0.88 (0.98)	0.84 (0.72)	0.76 (0.81)	0.93 (0.65)	0.36 (0.45)	0.46 (0.44)
MF	0.47 (0.20)	0.34 (0.38)	0.38 (0.29)	0.57 (0.29)	0.39 (0.28)	0.41 (0.13)	0.37 (0.15)	0.30 (0.35)	0.34 (0.49)	0.32 (0.36)
SHR	0.38 (0.37)	0.54 (0.49)	0.38 (0.44)	0.39 (0.47)	0.36 (0.38)	0.50 (0.43)	0.31 (0.40)	0.32 (0.23)	0.27 (0.27)	0.28 (0.24)
SAV	0.42 (0.40)	0.36 (0.21)	0.35 (0.40)	0.23 (0.15)	0.43 (0.41)	0.35 (0.23)	0.42 (0.37)	0.31 (0.10)	0.23 (0.21)	0.19 (0.10)
GRA	0.60 (0.59)	0.48 (0.49)	0.60 (0.56)	0.52 (0.55)	0.38 (0.29)	0.36 (0.37)	0.44 (0.39)	0.38 (0.38)	0.17 (0.20)	0.31 (0.31)
CRO	0.47 (0.37)	0.66 (0.44)	0.36 (0.33)	0.56 (0.47)	0.29 (0.32)	0.25 (0.22)	0.29 (0.32)	0.30 (0.29)	0.41 (0.31)	0.56 (0.55)
WET	0.54 (0.64)	0.28 (0.41)	0.55 (0.62)	0.29 (0.25)	0.72 (0.35)	0.48 (0.52)	0.69 (0.29)	0.50 (0.51)	0.24 (0.19)	0.30 (0.25)
Trop	1.66 (1.31)	0.67 (0.79)	1.71 (1.23)	0.77 (0.86)	1.73 (0.88)	1.16 (1.19)	1.94 (0.81)	1.21 (0.67)	0.52 (0.57)	0.38 (0.55)
SubTrop	0.54 (0.45)	0.55 (0.43)	0.50 (0.38)	0.52 (0.55)	0.46 (0.44)	0.53 (0.47)	0.47 (0.35)	0.42 (0.37)	0.34 (0.44)	0.37 (0.34)
Dry	0.31 (0.20)	0.33 (0.26)	0.33 (0.38)	0.36 (0.29)	0.24 (0.21)	0.32 (0.35)	0.34 (0.21)	0.43 (0.26)	0.14 (0.08)	0.22 (0.14)
Tmp	0.72 (0.55)	0.77 (0.71)	0.66 (0.59)	0.63 (0.56)	0.50 (0.46)	0.47 (0.50)	0.51 (0.55)	0.41 (0.45)	0.46 (0.43)	0.51 (0.41)
TmpCont	0.45 (0.35)	0.60 (0.52)	0.39 (0.35)	0.57 (0.47)	0.37 (0.28)	0.29 (0.25)	0.37 (0.33)	0.38 (0.37)	0.35 (0.40)	0.55 (0.55)
Bor	0.36 (0.30)	0.32 (0.34)	0.32 (0.24)	0.27 (0.31)	0.32 (0.40)	0.32 (0.33)	0.31 (0.35)	0.26 (0.32)	0.27 (0.26)	0.23 (0.26)
Cold	0.07 (0.00)	0.08 (0.09)	0.08 (0.12)	0.15 (0.06)	0.34 (0.04)	0.12 (0.06)	0.34 (0.06)	0.15 (0.01)	0.37 (0.15)	0.27 (0.27)

Table C4. Median site-by-site R^2 of the energy fluxes per PFT and climate zones. ENF is evergreen needle leaf forest, DBF deciduous broadleaf forest, EBF Evergreen broadleaf forest, MF mixed forest, SHR shrubland, SAV Savannah, GRA Grassland, CRO cropland, WET Wetland, Trop Tropical, SubTrop Subtropical, Dry Dryland, Tmp Temperate, TmpCont Temperate-continental, Bor Boreal, Cold cold environment or Iceland covered by ice.

CAT	H		LE		Rn	
	RS	RS+METEO	RS	RS+METEO	RS	RS+METEO
ENF	0.87 (0.10)	0.86 (0.10)	0.83 (0.10)	0.84 (0.11)	0.97 (0.02)	0.97 (0.02)
DBF	0.76 (0.18)	0.74 (0.12)	0.87 (0.05)	0.87 (0.07)	0.97 (0.01)	0.97 (0.02)
EBF	0.85 (0.13)	0.82 (0.17)	0.56 (0.30)	0.52 (0.42)	0.95 (0.05)	0.96 (0.03)
MF	0.85 (0.06)	0.82 (0.10)	0.91 (0.07)	0.89 (0.06)	0.97 (0.02)	0.96 (0.02)
SHR	0.83 (0.15)	0.83 (0.17)	0.73 (0.29)	0.77 (0.23)	0.98 (0.01)	0.97 (0.01)
SAV	0.74 (0.25)	0.77 (0.26)	0.85 (0.06)	0.78 (0.11)	0.86 (0.05)	0.88 (0.10)
GRA	0.72 (0.22)	0.71 (0.22)	0.85 (0.11)	0.83 (0.16)	0.96 (0.02)	0.96 (0.02)
CRO	0.70 (0.16)	0.66 (0.18)	0.79 (0.14)	0.80 (0.14)	0.97 (0.02)	0.96 (0.02)
WET	0.81 (0.06)	0.78 (0.14)	0.86 (0.10)	0.84 (0.06)	0.94 (0.02)	0.92 (0.06)
Trop	0.52 (0.18)	0.60 (0.32)	0.56 (0.38)	0.50 (0.44)	0.86 (0.14)	0.89 (0.13)
SubTrop	0.81 (0.18)	0.82 (0.18)	0.78 (0.13)	0.80 (0.13)	0.96 (0.03)	0.96 (0.02)
Dry	0.87 (0.07)	0.86 (0.13)	0.80 (0.07)	0.79 (0.14)	0.90 (0.06)	0.93 (0.05)
Tmp	0.78 (0.14)	0.78 (0.13)	0.86 (0.11)	0.83 (0.13)	0.97 (0.02)	0.96 (0.02)
TmpCont	0.72 (0.18)	0.69 (0.18)	0.83 (0.08)	0.84 (0.09)	0.97 (0.02)	0.96 (0.02)
Bor	0.90 (0.07)	0.89 (0.08)	0.92 (0.05)	0.92 (0.03)	0.98 (0.01)	0.97 (0.02)
Cold	0.83 (0.12)	0.57 (0.19)	0.83 (0.08)	0.82 (0.07)	0.94 (0.03)	0.85 (0.13)

Table C5. Median site-by-site RMSE of the energy fluxes per PFT and climate zones. ENF is evergreen needle leaf forest, DBF deciduous broadleaf forest, EBF Evergreen broadleaf forest, MF mixed forest, SHR shrubland, SAV Savannah, GRA Grassland, CRO cropland, WET Wetland, Trop Tropical, SubTrop Subtropical, Dry Dryland, Tmp Temperate, TmpCont Temperate-continental, Bor Boreal, Cold cold environment or Iceland covered by ice.

CAT	H ($MJm^{-2}d^{-1}$)		LE ($MJm^{-2}d^{-1}$)		Rn ($MJm^{-2}d^{-1}$)	
	RS	RS+METEO	RS	RS+METEO	RS	RS+METEO
ENF	1.09 (0.25)	1.16 (0.25)	1.00 (0.56)	1.02 (0.55)	1.27 (0.68)	1.26 (0.57)
DBF	1.30 (0.43)	1.31 (0.38)	1.22 (0.26)	1.14 (0.46)	1.11 (0.42)	1.24 (0.41)
EBF	1.14 (0.60)	1.29 (0.76)	1.55 (0.39)	1.60 (0.46)	1.33 (0.43)	1.14 (0.56)
MF	1.18 (0.44)	1.12 (0.42)	0.82 (0.37)	1.15 (0.54)	1.14 (0.45)	1.09 (0.43)
SHR	1.21 (0.46)	1.14 (0.28)	1.12 (0.41)	1.11 (0.56)	1.37 (0.80)	1.01 (0.43)
SAV	1.23 (0.25)	1.20 (0.22)	1.32 (0.56)	1.35 (0.30)	1.10 (0.33)	1.19 (0.60)
GRA	1.14 (0.35)	1.08 (0.47)	1.09 (0.34)	1.32 (0.54)	1.48 (0.83)	1.48 (0.90)
CRO	1.24 (0.45)	1.36 (0.33)	1.51 (0.61)	1.54 (0.35)	1.24 (0.52)	1.23 (0.26)
WET	0.97 (0.36)	1.22 (0.60)	0.88 (0.13)	0.90 (0.18)	1.42 (0.51)	1.65 (0.71)
Trop	0.98 (0.51)	1.19 (0.63)	1.60 (0.52)	1.62 (0.41)	1.33 (0.73)	1.03 (0.48)
SubTrop	1.28 (0.38)	1.32 (0.46)	1.36 (0.62)	1.36 (0.53)	1.40 (0.40)	1.33 (0.49)
Dry	1.07 (0.24)	1.05 (0.50)	1.21 (0.33)	1.27 (0.52)	1.61 (0.75)	2.02 (0.93)
Tmp	1.18 (0.23)	1.15 (0.33)	1.18 (0.43)	1.17 (0.49)	1.10 (0.36)	1.14 (0.47)
TmpCont	1.30 (0.42)	1.35 (0.37)	1.25 (0.41)	1.47 (0.37)	1.17 (0.65)	1.16 (0.54)
Bor	0.98 (0.23)	1.05 (0.26)	0.70 (0.26)	0.61 (0.20)	0.88 (0.31)	1.08 (0.50)
Cold	1.03 (0.36)	1.50 (0.55)	1.00 (0.23)	1.03 (0.45)	1.47 (0.18)	2.04 (0.19)

Table C6. Median site-by-site absolute bias for energy fluxes. . ENF is evergreen needle leaf forest, DBF deciduous broadleaf forest, EBF Evergreen broadleaf forest, MF mixed forest, SHR shrubland, SAV Savannah, GRA Grassland, CRO cropland, WET Wetland, Trop Tropical, SubTrop Subtropical, Dry Dryland, Tmp Temperate, TmpCont Temperate-continental, Bor Boreal, Cold cold environment or Iceland covered by ice.

CAT	H ($MJm^{-2}d^{-1}$)		LE ($MJm^{-2}d^{-1}$)		Rn ($MJm^{-2}d^{-1}$)	
	RS	RS+METEO	RS	RS+METEO	RS	RS+METEO
ENF	0.44 (0.40)	0.40 (0.33)	0.42 (0.41)	0.44 (0.49)	0.78 (0.63)	0.64 (0.61)
DBF	0.60 (0.35)	0.66 (0.35)	0.57 (0.56)	0.49 (0.50)	0.38 (0.28)	0.61 (0.49)
EBF	0.38 (0.48)	0.55 (0.46)	0.97 (0.79)	0.88 (0.70)	0.88 (0.51)	0.62 (0.43)
MF	0.48 (0.40)	0.26 (0.31)	0.34 (0.40)	0.64 (0.52)	0.56 (0.45)	0.56 (0.57)
SHR	0.34 (0.43)	0.47 (0.52)	0.41 (0.41)	0.50 (0.43)	0.62 (0.76)	0.44 (0.52)
SAV	0.68 (0.35)	0.56 (0.15)	0.63 (0.80)	0.40 (0.15)	0.27 (0.22)	0.63 (0.55)
GRA	0.51 (0.39)	0.40 (0.24)	0.38 (0.38)	0.57 (0.50)	0.97 (0.81)	0.81 (1.03)
CRO	0.23 (0.21)	0.24 (0.24)	0.36 (0.38)	0.41 (0.50)	0.66 (0.58)	0.68 (0.39)
WET	0.47 (0.51)	0.67 (0.37)	0.54 (0.41)	0.38 (0.21)	0.34 (0.34)	0.83 (0.78)
Trop	0.37 (0.51)	0.67 (0.47)	0.97 (0.79)	1.24 (0.82)	0.94 (1.10)	0.63 (0.60)
SubTrop	0.58 (0.59)	0.50 (0.39)	0.62 (0.58)	0.58 (0.56)	0.83 (0.71)	0.70 (0.55)
Dry	0.68 (0.62)	0.55 (0.56)	0.21 (0.14)	0.30 (0.26)	1.06 (0.55)	1.61 (0.91)
Tmp	0.38 (0.23)	0.34 (0.31)	0.49 (0.46)	0.56 (0.54)	0.65 (0.49)	0.68 (0.58)
TmpCont	0.49 (0.41)	0.40 (0.46)	0.44 (0.51)	0.53 (0.50)	0.69 (0.72)	0.61 (0.58)
Bor	0.33 (0.32)	0.38 (0.24)	0.22 (0.16)	0.23 (0.24)	0.38 (0.27)	0.50 (0.47)
Cold	0.43 (0.46)	0.71 (0.11)	0.56 (0.31)	0.39 (0.18)	0.30 (0.29)	0.86 (0.58)

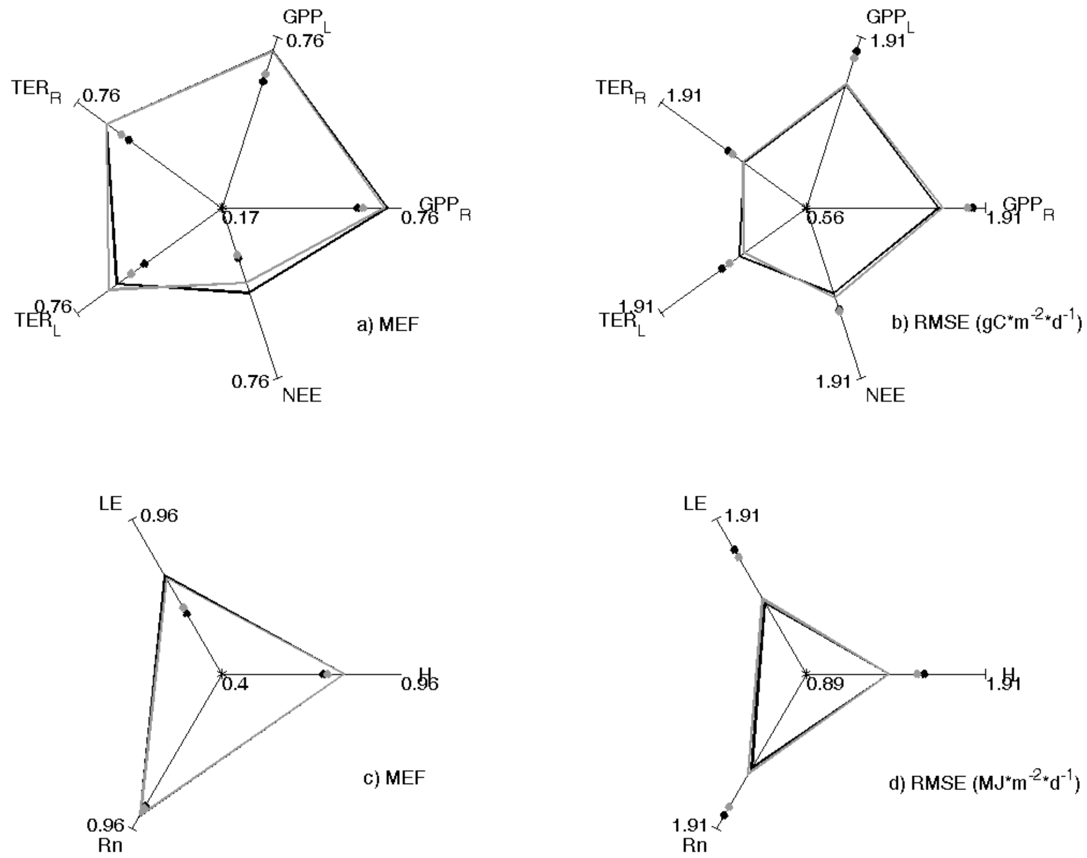

Fig. 1: Spider plot of MEF (first column) and RMSE (second column) for carbon (first row) and energy fluxes (second row) showing the consistency of prediction made RS (Black line) and RS+METEO (grey lines) setups. The lines are the ensemble median estimate; we also show the performance of multiple regressions trained with RS (black point) and RS+METEO (gray points). GPP_R and GPP_L are respectively gross primary production estimated following Reichstein et al. (2005) and Lasslop et al. (2010), TER_R and TER_L the total ecosystem respiration estimated following Reichstein et al. (2005) and Lasslop et al. (2010), NEE net ecosystem exchange, H the sensible heat, LE the latent heat and Rn the net radiation.

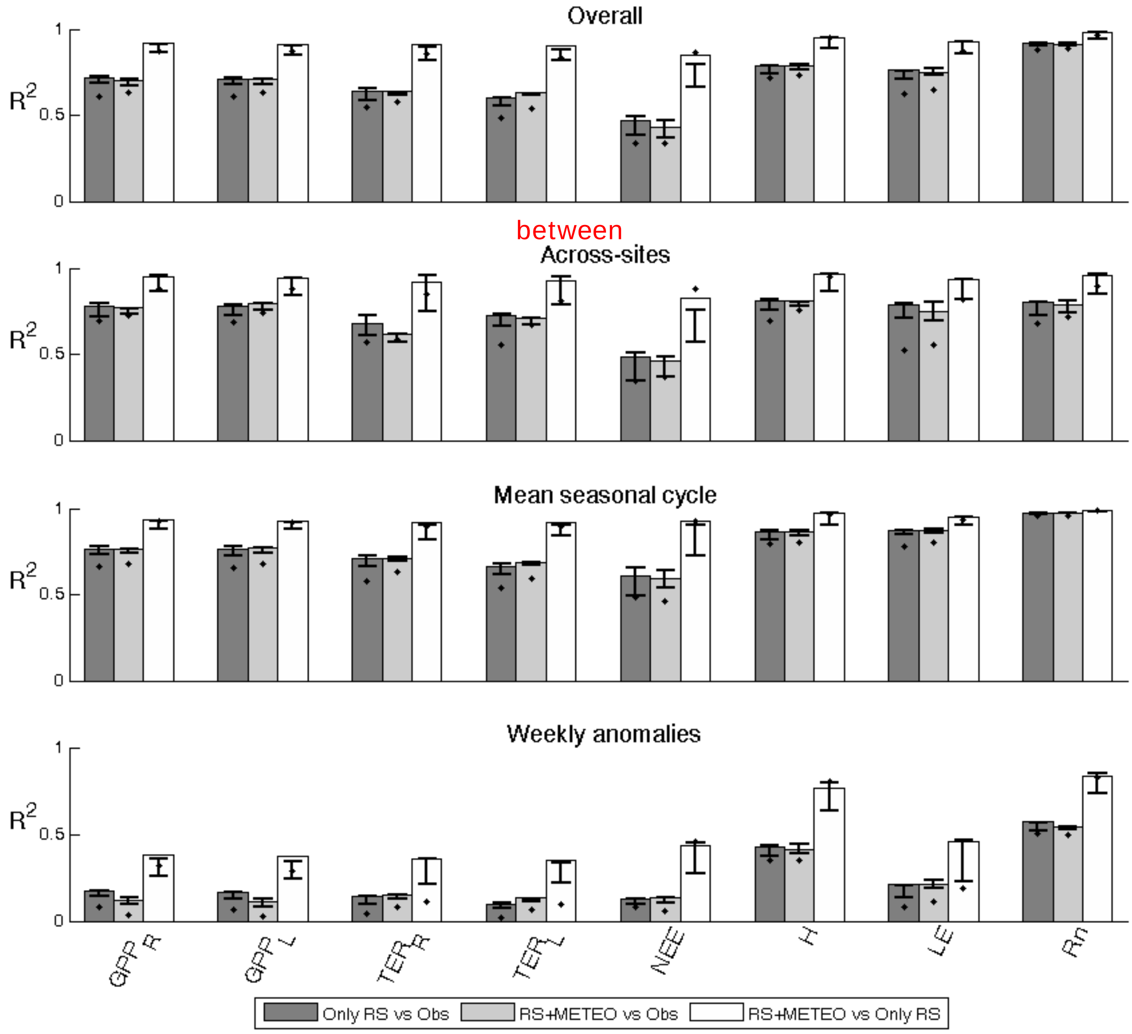

Fig. 2: Coefficients of determination (R^2) from the comparison of overall time series, across-sites, mean seasonal cycle, and the anomalies, in particular: the determination coefficients between predictions by the ensemble median estimate of RS setup and observation (dark grey bars), between predictions by the ensemble median estimate of RS+METEO setup and observation (light grey bars), and between the two ensembles median estimate (white bars). Whiskers are the higher and lower R^2 when the comparisons are made among the singular ML. The comparison of output by the multiple regressions is also shown (black points).

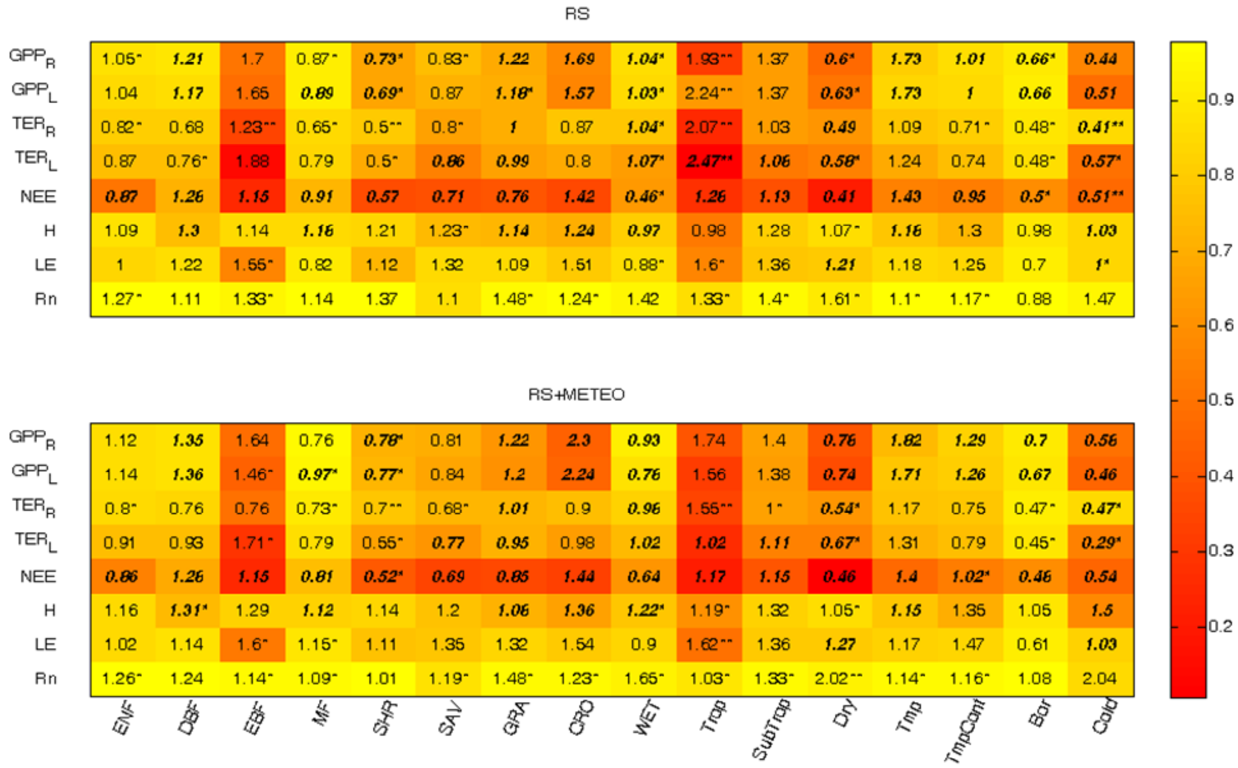

Fig. 3: Performance of FLUXCOM median estimates per climate zone and plant functional type (PFT). The colored matrices show the median values of R^2 (red pixels for low R^2, yellow pixels for high R^2). Number indicate the RMSE (units of carbon fluxes are $gCm^{-2}d^{-1}$ and $MJm^{-2}d^{-1}$ in the case of energy fluxes). Oblique and bold font are used when the relative RMSE (normalized for the mean observed fluxes per PFT and climate zone) is greater than 0.5. The symbols '**' after RMSE are used when the weight of bias (estimated as the ratio between the square of median absolute bias and the MSE) is greater than 0.5, instead '*' are used if the weight of bias is between 0.25 than 0.5. No symbols are used if the weight of bias is lesser than 0.25.

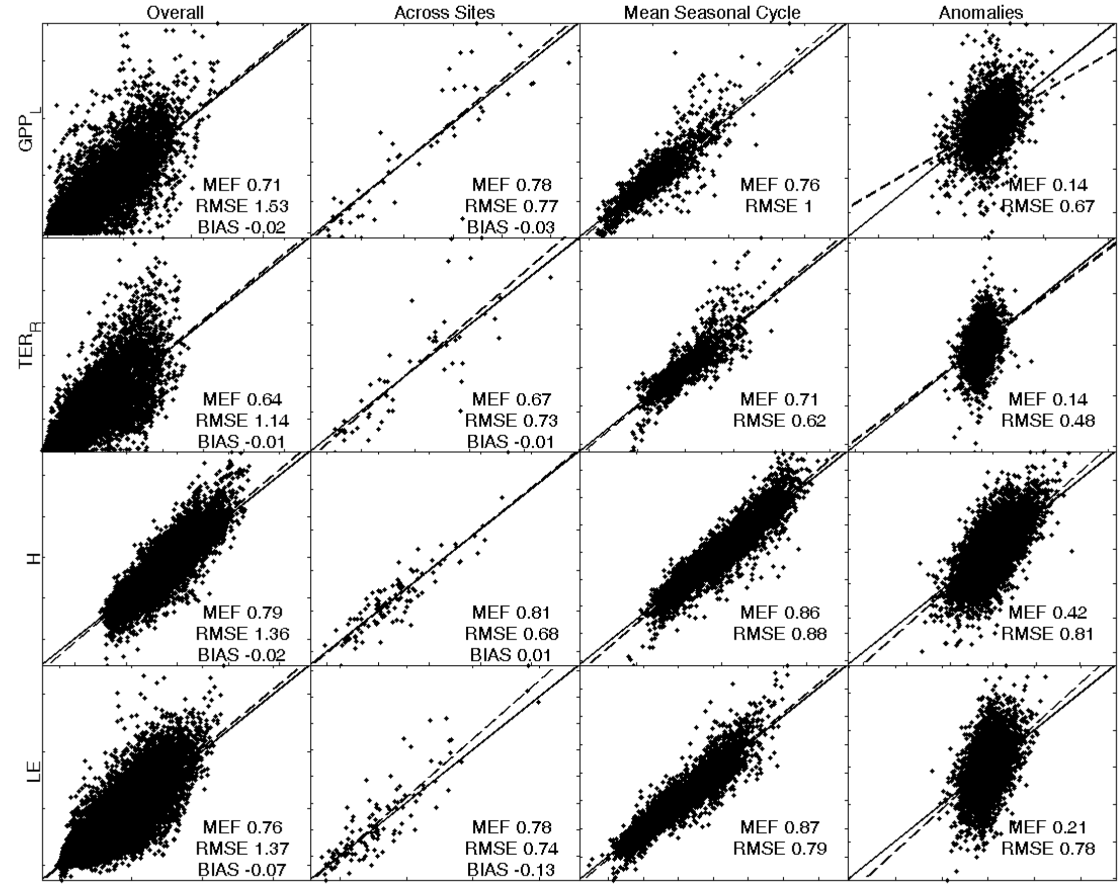

Fig. B1. Scatterplots of observed data by eddy covariance (y-axis) and the median ensemble of modeled fluxes by RS setup (x-axis). The panels from left to right are the eight days predictions, across sites variability, mean seasonal cycle and weekly anomalies. The fluxes considered here are GPP_L (first row), TER_R (second row), H (third row) and LE (fourth row).

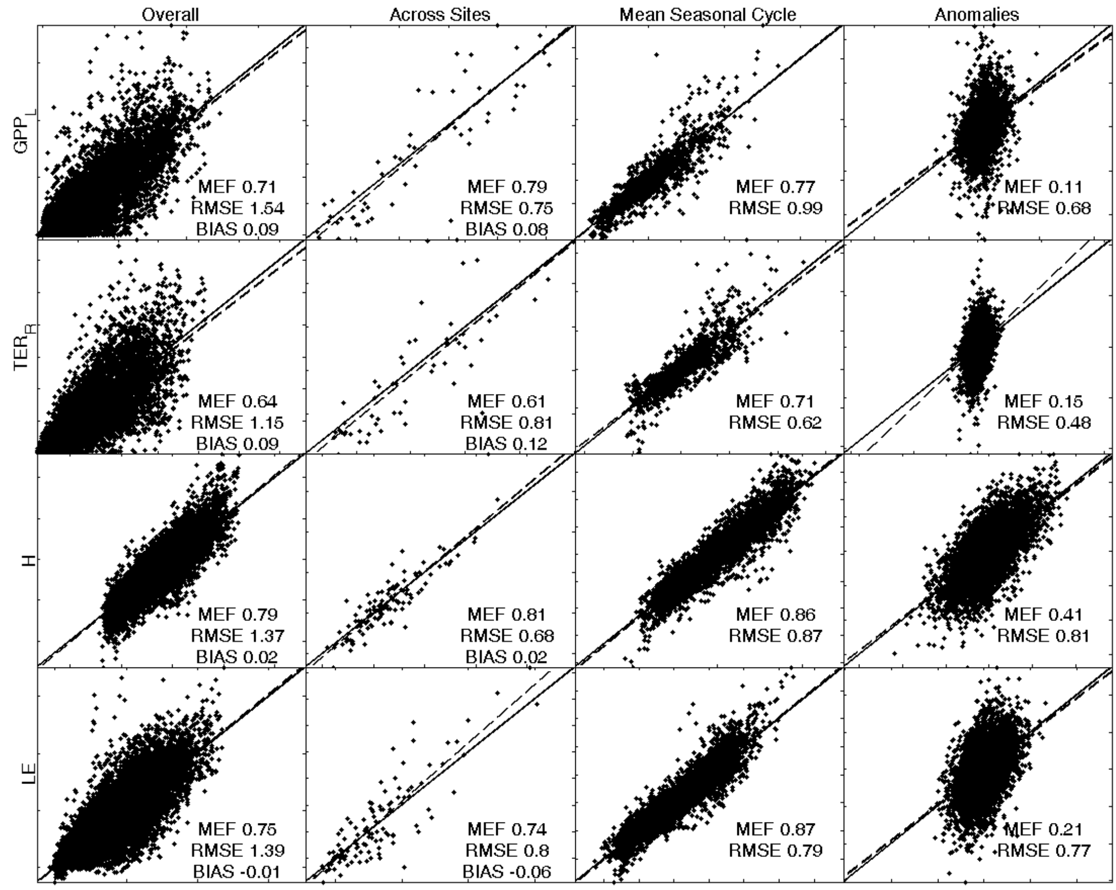

Fig. B2. As Fig. B1 but the predictions (x-axis) were obtained by the RS+METEO setup.