Detecting changes in essential ecosystem and biodiversity properties - towards a Biosphere Atmosphere Change Index: BACI

Deliverable 4.4: Global products of carbon and energy fluxes derived from upscaling FLUXNET observations that resolve the diurnal cycle

Project title:	Detecting changes in essential ecosystem and biodiversity properties - towards a Biosphere Atmosphere Change Index
Project acronym:	BACI
Grant agreement number:	640176
Main pillar:	Industrial Leadership
Topic:	EO-1-2014: New ideas for Earth-relevant space applications
Start date of the project:	1st April 2015
Duration of the project:	48 months
Dissemination level:	Public
Responsible of the deliverable:	Martin Jung Phone: ++49 3641 57 6261 Email: mjung@bgc-jena.mpg.de
Authors and contributors:	Paul Bodesheim, Martin Jung, Miguel Mahecha, Markus Reichstein
Date of submission:	29th September 2017

Preface

One of the major objectives of BACI is to derive new downstream data products from Earth observations. In particular, we seek to derive new “essential ecosystem variables” such as land-atmosphere fluxes. understanding interactions between the biosphere and the atmosphere at the subdaily scale is one of the major current frontiers of research. However, so far there are no data products of biosphere-atmosphere fluxes that resolve the diurnal cycle. Currently available global products only provide information in terms of daily average values for the carbon and energy fluxes, but we are interested in data of high temporal resolution and therefore focused on the half-hourly time scale. Hence, within BACI we upscaled diurnal cycles characterized by half-hourly values of four major fluxes: gross primary production (GPP), net ecosystem exchange (NEE), latent heat flux (LE), and sensible heat flux (H). We made use of half-hourly observations at FLUXNET towers and developed machine learning based regression approaches for estimating half-hourly fluxes globally. In the end, we were able to compute global products at half-hourly temporal resolution and half degree spatial resolution for the four aforementioned fluxes (GPP, NEE, LE, H) covering the years 2001 to 2014. We plan to submit our results and findings to a journal within the next weeks and of course, making the computed data products publically available for free goes along with this submission. Since the paper is very close to its final version and best describes our work that has been carried out to compute the global data products in a detailed manner, we have decided to directly use it for this deliverable of the project as well. The following pages therefore contain the preliminary version of the paper including related work and research context, description of the developed upscaling approaches based on machine learning regression techniques, results of cross-validation experiments, and presentation of the global half-hourly data products with some derived statistics and properties.

Upscaled diurnal cycles of land-atmosphere fluxes: a new global half-hourly data product

Paul Bodesheim[1], Martin Jung[1], Miguel D. Mahecha[1,2,3], and Markus Reichstein[1,2,3]

[1]Max Planck Institute for Biogeochemistry, Jena, Germany
[2]German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany
[3]Michael Stifel Center Jena (MSCJ) for Data-Driven & Simulation Science, Jena, Germany

Correspondence to: Paul Bodesheim (pbodes@bgc-jena.mpg.de)

Abstract. Interactions between the biosphere and the atmosphere can be well characterized by fluxes between the two. In particular, carbon and energy fluxes play a major role for understanding biogeochemical processes on ecosystem level or global scale. However, the fluxes can only be measured at individual sites by eddy covariance towers and an upscaling of these local observations is required to analyze global patterns. Previous work focused on upscaling monthly, 8-daily, or daily average values and global maps for each flux have been provided accordingly. In this paper, we raise the upscaling of carbon and energy fluxes between land and atmosphere to the next level by increasing the temporal resolution to subdaily scales. We provide continuous half-hourly fluxes for the period from 2001 to 2014 at half degree spatial resolution, which allows for analyzing diurnal cycles globally. The dataset contains four fluxes: gross primary production (GPP), net ecosystem exchange (NEE), latent heat (LE), and sensible heat (H). We propose two prediction approaches for the diurnal cycles based on large-scale regression models and compare them in extensive cross-validation experiments using different sets of predictor variables. We analyze the results for a set of FLUXNET tower sites showing the suitability of our approaches for this upscaling task. Finally, we have selected one approach to calculate the global half-hourly data products based on predictor variables from remote sensing and meteorology at daily resolution and half-hourly potential radiation. In addition, we provide a derived product that only contains monthly average diurnal cycles, which is a lightweight version in terms of data storage that still enables to study the important characteristics of diurnal courses globally. We recommend to primarily use these monthly average diurnal cycles, because they are less affected by the impacts of day-to-day variation, observation noise, and short-term fluctuations on subdaily scales compared to the plain half-hourly flux products.

1 Introduction

Understanding the coupling of the atmosphere and the biosphere is key to understand Earth system dynamics and ultimately to predict future trajectories based on dynamic and fully coupled Earth system models (Bonan, 2008). Eddy co-variance measurements of energy and carbon fluxes have revealed major insights into land-atmosphere interactions (see the overview by Balddocchi, 2014), but the underlying observations are local by nature and it remains difficult to derive global inferences. To overcome this limitation, continental to global scale products of biosphere-atmosphere fluxes have been produced using machine learning techniques that combine flux tower measurements, observations from remote sensing, and climate data (Jung et al., 2009; Papale et al., 2015). These products proofed to be useful for example in terms of assessing large scale patterns of biosphere-atmosphere fluxes with climate (Jung et al., 2010) or to provide cross-consistency checks for process-model simulations (Bonan et al., 2011).

The general principle of this upscaling approach has been to exploit relationships between climate or satellite-based driver variables like temperature or leaf area index, and the targeted biosphere-atmosphere flux (Xiao et al., 2012). In the first ("training") step, a machine learning model of the flux data is established based on the driver variables across a regional or global network of towers. In the second ("production"[1]) step, the model is being applied to large spatial domains where only gridded estimates of the drivers are available. Machine learning techniques are very effective here since they are fully data-adaptive, do not require making assumptions on functional relationships, and can cope with nonlinear dependencies.

One of the first papers by Jung et al. (2009) deals with empirical upscaling of monthly average values of Gross Primary Production (GPP) obtained from a biosphere model. They propose using a model tree ensemble approach to perform the predictions and introduce both a new model tree induction algorithm and a specific ensemble approach. Later, Beer et al. (2010) estimated GPP for different biomes, focussing on global median annual GPP derived using different prediction approaches. Covering a larger number of variables, Jung et al. (2011) produced global flux products at half degree spatial resolution for monthly average values of GPP, terrestrial ecosystem respiration (TER), net ecosystem exchange (NEE), latent heat (LE), and sensible heat (H). Their findings were confirmed by a comprehensive cross-validation analysis using FLUXNET towers. In the latest study of Jung et al. (2017), they investigate the dependencies of changes in temperature and water availability on the interannual variability of carbon fluxes both locally and globally using their upscaled data products and process-based global land models. There exist further upscaling approaches in the literature based on support vector regression models (Yang et al., 2007; Ueyama et al., 2013; Ichii et al., 2017) that estimate carbon fluxes on regional to continental scale. The work of Xiao et al. (2008, 2010) deals with estimating carbon fluxes for the United States using data from MODIS and AmeriFlux. Only recently, a systematic comparison of different regression algorithms for predicting carbon and energy fluxes has been carried out by Tramontana et al. (2016). The authors of this paper were interested in the best prediction performances for estimating GPP, TER, NEE, LE, H, as well as net radiation at either eight-day or daily temporal resolution. In their cross-validation analysis, they found that prediction performance varies only slightly among different regression algorithms from machine learning. However, they could show that accuracies clearly differ between the individual fluxes, meaning that some fluxes are harder to estimate than others, which is probably due to a lack of information in the set of explanatory variables.

Today, global flux products feature at best a daily temporal resolution as presented by Tramontana et al. (2016). This is partly due to rapidly growing computational issues in the training and production step scaling quadratically with spatial resolution. In addition, consistent global long-term products of driver data with hourly or higher temporal resolutions are lacking or not readily available. Upscaling half-hourly carbon and energy fluxes rises previous upscaling approaches to the next level by increasing the temporal resolution to subdaily scales.

[1] Note that an alternative notion would be to use the term "prediction" here. However, in the climate community "prediction" is typically used for future scenarios, while in machine learning the application domain could be also at ungauged spatial locations etc.

However, there is a need for a global data product of half-hourly fluxes. Such a data product would allow for characterizing sub-daily variations in the diurnal cycles at places where no towers are currently installed. Characterizing typical subdaily flux patterns is critically needed for certain satellite remote sensing applications. For example, the interpretation of satellite retrievals of sun-induced fluorescence as proxy for photosynthesis (Guanter et al., 2014; Sun et al., 2015) or integrated atmospheric column carbon dioxide (XCO2) at certain overpass times (usually around mid-day) requires consideration of strong diurnal variations of biosphere-atmosphere carbon fluxes. Another area where half-hourly data products would be a crucial piece of information are land-atmosphere feedback modeling studies. The derived products could allow to check the cross-consistency, since many processes governing land-atmosphere interactions, e.g., related to the formation of heavy rainfall or heat waves, in fact operate at subdaily time scales (Dirmeyer et al., 2012).

In view of the need for global high-frequency flux data, here we aim at increasing the temporal resolution of data-driven carbon and energy flux products to sub-daily scales by estimating half-hourly values at global scale. We tackle the problem of predicting diurnal cycles of carbon and energy fluxes between biosphere and atmosphere with half-hourly values globally by treating the upscaling task as a large-scale regression problem. From the machine learning perspective, the random forest regression framework serves as a basis for our computations due to its good performance and suitable scaling properties with respect to large data sets. We test two approaches for estimating half-hourly GPP with random forest models and evaluate both of them using a leave-one-site-out cross-validation strategy for the FLUXNET towers of the LaThuile dataset[2]. We produce derived global products with half degree spatial and half-hourly temporal resolution for GPP and NEE as well as for LE and sensible heat H covering the years 2001 to 2014. For the sake of clarity, some figures in this paper only show the results obtained for GPP although similar plots can easily be produced for the other three fluxes that have been considered. Thus, GPP serves as the running example throughout this paper.

The follwing sections are organized as follows. First, we introduce the used data basis by describing both site-level and global forcing data (Sect. 2) that are used in our study. Then, we explain the methodological background (Sect. 3) and the algorithmic concept (Sect. 4) of the proposed upscaling approaches in detail. In Sect. 5, extensive evaluations and comparisons of the different upscaling strategies are presented based on leave-one-site-out cross-validation, which validate the proposed approach and the derived global products. Afterwards, we present the empirical results at global scale in Sect. 6 and highlight intrinsic features of the new data sets. Finally, we discuss both our findings and possible improvements for future applications (Sect. 7). The global data sets presented in this paper will be freely available to any interested user (see Sect. 8).

2 Data sources

In this section, we shortly describe the two data sources we are using in our studies. For learning the relationships between predictor variables and the target fluxes as well as for the cross-validation experiments, we make use of site-level data extracted at FLUXNET sites that are equipped with eddy covariance towers (Sect. 2.1). To perform global upscaling of diurnal cycles, we require gridded data products of the predictor variables on a global scale. The latter are described in Sect. 2.2.

[2]http://fluxnet.fluxdata.org/data/la-thuile-dataset/

2.1 Site-level data

We aim at predicting fluxes on a half-hourly resolution which can currently only be measured using the eddy covariance method at globally distributed towers. The eddy covariance method (Baldocchi et al., 1988; Aubinet et al., 2012) has revolutionized the study into land-atmosphere interactions by offering a means of continuously observing net land-atmosphere fluxes of CO_2, latent heat, and sensible heat (Balddocchi, 2014). Today, the towers are running for sufficient time to enable studies about the interannual variability of land-surface dynamics, yet the temporal representativity is highly uneven (Chu et al., 2017). In our studies, we rely on 222 FLUXNET sites of the LaThuile dataset. All towers are typically equipped with a suite of comparable micrometeorological devices so that a local training of machine learning methods is possible. Gross carbon fluxes can be derived using different flux partitioning methods as described, e.g., by Reichstein et al. (2005) or Lasslop et al. (2010), and here we rely on the former method. In all our experiments, we only rely on measured fluxes, i.e., no gapfilling has been applied and gaps in the half-hourly flux data have simply been ignored. As predictor variables, we use the ones selected by Tramontana et al. (2016, Table 2) in the RS+METEO setup that they use for estimating fluxes at daily resolution. Besides the plant functional type (PFT), these are variables containing remote sensing data from MODIS satellites and meteorological data either in situ measured at the flux tower locations or from long-term time series of the ERA-Interim data set at daily resolution. For detailed descriptions, we refer to the corresponding sections in the paper of Tramontana et al. (2016, Sect. 2.1.3 and 2.1.4).

2.2 Global forcing data

In order to compute the global flux products at half-hourly resolution via upscaling, we require the predictor variables mentioned in the previous section at global scale, i.e., the variables of the RS+METEO setup from Tramontana et al. (2016, Table 2). Concerning the remote sensing variables, MODIS observations are used to compute mean values for each PFT and each day aggregated to half degree spatial resolution. The distributions of each PFT stem from the MODIS collection 5 global land cover product of Friedl et al. (2010). Climatic data for the meteorological variables have been obtained from CRUNCEPv6[3], which denotes a merged data product of monthly observation-based climate variables at half degree spatial resolution from the Climate Research Unit (CRU) and 6-hourly reanalysis data from the National Centers for Environmental Prediction (NCEP).

3 Methodological background: random forest regression

Ensemble methods are powerful machine learning tools that combine the outputs of many individual prediction models to obtain more accurate estimations for a target variable. The random forest approach (Breiman, 2001) denotes a typical example, which consists of a set of randomized decision trees. Decision trees in general can be built for classification or regression purposes and they are therefore also called classification trees or regression trees. Multiple decision trees form a decision forest and learning their decision rules typically involves some randomization, which leads to the name randomized decision forest or short random forest. In the following, the concepts of learning and testing randomized decision trees for regression tasks

[3] http://esgf.extra.cea.fr/thredds/catalog/store/p529viov/cruncep/V6_1901_2014/catalog.html

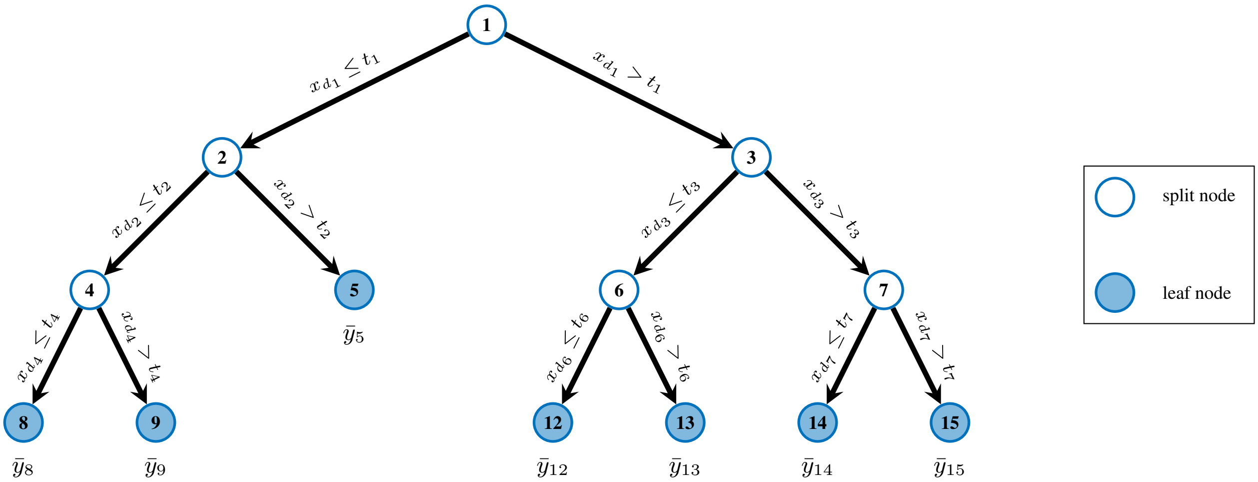

Figure 1. General structure of a decision tree for regression: binary splits with thresholds for individual predictor variables will be used to navigate a sample $\boldsymbol{x}$ to a leaf node that stores a continuous estimate for the target variable.

are briefly summarized, because they denote the essential parts of random forest regression. The reader who is familar with the technical details of random forest regression can skip this section and may directly continue with the proposed upscaling approaches in Sect. 4.

3.1 Randomized decision tree

Given a training set $\mathcal{X} = \left\{\boldsymbol{x}^{(i)} \in \mathbb{R}^D : i = 1,2,\ldots,N\right\}$ of N samples with each sample $\boldsymbol{x}$ being a vector consisting of D predictor variables $x_1, x_2, \ldots, x_D$ and a corresponding real-valued target variable $y \in \mathbb{R}$ with observations $y_1, y_2, \ldots, y_N \in \mathbb{R}$ for the N training samples, the goal is to find a set of rules that allow for predicting y based on $\boldsymbol{x}$. In case of a decision tree, these rules are binary tests for individual predictor variables with simple thresholds. A hierarchical tree structure is built as shown in Fig. 1 by selecting at each node i a predictor variable $d_i \in \{1,2,\ldots,D\}$ and a threshold $t_i \in \mathbb{R}$. The estimate of a node i is the average value $\bar{y}_i$ of the observations computed from training samples that reach this node. The first node of a decision tree called root node contains all training samples and hence, the overall mean value $\bar{y}_1 = \frac{1}{N}\sum_{n=1}^{N} y_n$ of observations y_n from all N training samples is an extremely coarse approximation that needs to be refined depending on the constellation of the input variables $\boldsymbol{x}$.

Starting at node 1 in Fig. 1, the set of training samples is partitioned into two subsets, represented by nodes 2 and 3, based on the result of the binary test $x_{d_1} \le t_1$. Both nodes, node 2 and node 3, have associated predicted outputs $\bar{y}_2$ and $\bar{y}_3$ that are computed as the average observation of samples that reach the corresponding node. The split parameters d_1 and t_1 are optimized such that the mean squared error for the training samples is minimized given the respective predictions from node 2 or node 3. Such splits are then computed for nodes 2 and 3 as well as for further derived nodes until a stopping criterion is fulfilled. Typical stopping criteria are: (i) a split would create nodes with less than $N_{\text{min}} = 5$ samples, (ii) the variance of the observations from samples in a node is smaller than some threhold $\sigma^2_{\text{min}} = 0.05$, or (iii) a maximum depth d_{max} of the tree is reached. The depth of a tree is defined as the largest distance of a node to the root of the tree. Values of the parameters N_{min}, σ^2_{min}, and d_{max} can be changed to obtain either smaller or larger trees, which allows for controlling the runtime of the algorithm.

It is usually the case that multiple stopping criteria are tested and if one of them is fulfilled, the current node is not split but becomes a leaf node that stores a final output prediction. Learning a decision tree therefore consists of computing split parameters until only leaf nodes remain that are not split any further (Fig. 1). Hence, one distinguishes between split nodes as the inner nodes of a tree and leaf nodes, each of which contains the estimated output for any sample that reaches this node.

To reduce overfitting to the training set, the learning process is carried out in a stochastic manner by introducing several types of randomization. Whenever split parameters need to be identified, only a random subset of the D predictor variables is taken into account. Furthermore, only a fixed amount of randomly chosen thresholds is tested. Both randomization techniques also lead to reduced computation costs compared to exhaustive search. To predict the output y^* of a test sample $\boldsymbol{x}^*$, it is passed through the tree according to the evaluation of the split functions at the inner nodes starting at the root node. This is done until a leaf node ℓ is reached, whose precomputed output $\bar{y}_\ell$ is assigned to $\boldsymbol{x}$. However, more accurate predictions can be achieved by considering an ensemble of randomized decision trees.

3.2 Random forest as an ensemble of randomized decision trees

In his work about random forests, Breiman (2001) makes use of a technique called bagging that he has introduced before (Breiman, 1996). Bagging is an acronym for bootstrap aggregating (Breiman, 1996) and stands for aggregating predictions of individual models that have been learned based on different sample sets built from the original training dataset. More precisely, individual sample sets are constructed by random sampling with replacement from the original training set, which is commonly referred to as bootstrapping. If the training set contains N samples, it is possible that each of the sampled sets either contains also N samples (which produces different sets with individual instances occuring several times due to random sampling with replacement) or only a fraction ν of the N samples. In both cases, the random subset selections introduced by bagging additionally prevent overfitting to the training set. For bagging, predictions from an ensemble of individual models are utilized and an ensemble of randomized decision trees is called a random forest or RDF. Each tree in the ensemble is learned separately and independent from the other trees. Due to the involved randomization techniques during learning of a single tree described before, different trees contain different binary tests and provide different estimates for a single input sample $\boldsymbol{x}$. The individual predictions of each tree are then aggregated to obtain a final result, which is typically carried out by simple averaging as shown in Fig. 2. However, the number of trees N_{tree} is a hyperparameter whose value needs to be chosen in advance but good assignments depend on various aspects. Since Breiman (1996) pointed out that bagging leads to predictions which are more stable compared to a single model, especially if the decision function of the single model is highly instable with respect to the training set, a larger number of trees is in favor of higher stability. On the other hand, more trees are causing higher computational costs during both learning and testing. In addition, a saturation effect for the prediction accuracy can typically be observed for an increasing number of trees. Hence, accuracies obtained by cross-validation for different numbers of trees can help to identify this saturation and a proper value for N_{tree}.

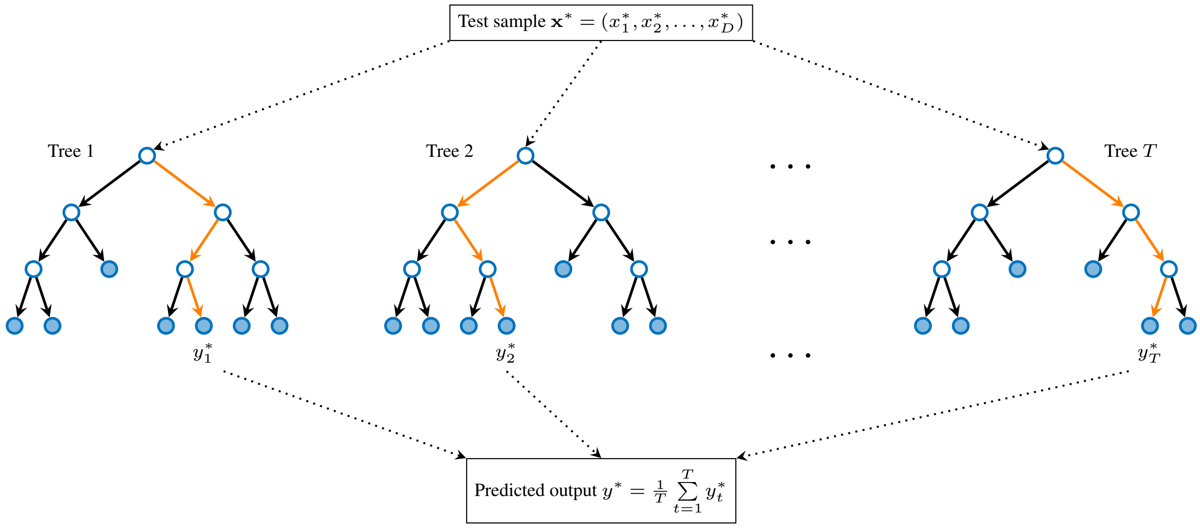

Figure 2. Predicting the output y^* of a sample $\boldsymbol{x}^*$ with a randomized decision forest is carried out by averaging the individual predictions obtained from the T decision trees in the ensemble.

4 Methods for upscaling diurnal cycles

The problem of upscaling diurnal cycles of carbon and energy fluxes can be formulated as a large-scale regression task, i.e., estimating half-hourly fluxes for every grid cell of the globe based on a set of predictor variables. These predictor variables typically encode climate conditions or Earth observations obtained from remote sensing at the corresponding spatial positions. However, the temporal resolutions of variables can be different, not only between the target flux (half-hourly) and a predictor variable (e.g., daily), but also among different predictor variables (e.g., daily and half-hourly). Therefore, two prediction approaches for upscaling diurnal cycles are presented in Sect. 4.1 and Sect. 4.2, respectively, which account for this mismatch of temporal resolutions. Although both approaches can be equipped with any regression algorithm, we have decided to use random decision forests (RDF) as a nonlinear method, which has been summarized in the previous section. The main reasons for this choice are the fast learning and testing algorithms, because the upscaling tasks involve a huge number of samples such that learning nonlinear kernel methods for regression like Gaussian processes (Rasmussen and Williams, 2006) are impractical due to both memory demand and computation time. Furthermore, Tramontana et al. (2016) have shown in their cross-validation experiments that the accuracies for estimating fluxes vary only slightly among different machine learning methods.

4.1 First prediction approach: an individual regression model for every half hour of the day

Recall from the beginning of Sect. 4 that the two main challenges for upscaling diurnal cycles to global scale are the huge amount of data which needs to be handled as well as the mismatch of temporal resolutions between predictor variables and the target fluxes. The first approach for predicting diurnal cycles has the advantage that it allows for using only predictors of daily temporal resolution. This is very important, because daily (average) values are often more accurate with respect to measurement

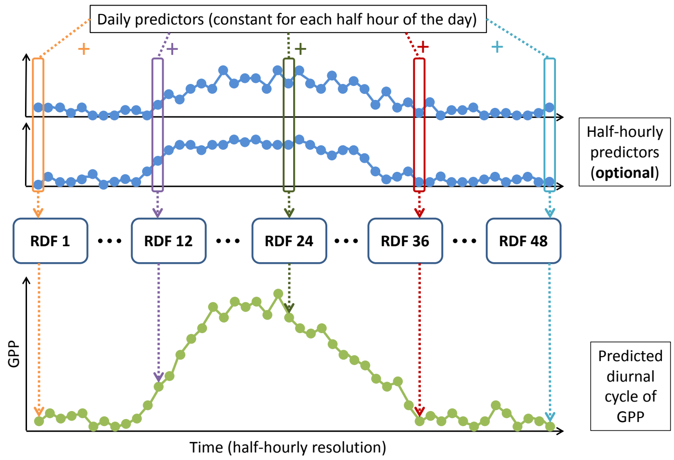

Figure 3. Visualization of the first prediction approach: an individual RDF regression model is learned for each half hour of a day, which allows for predicting diurnal cycles only based on predictors with daily resolution. However, predictors with half-hourly resolution can also be incorporated.

noise and the availability of daily values is much higher compared to half-hourly values, especially when considering global products with values for every grid cell. Furthermore, variables derived from remote sensing are often limited to daily temporal resolution. Therefore, the first prediction approach involves learning an individual regression model for each half hour of the day and as indicated at the beginning of Sect. 4, RDF regression models are utilized for handling large-scale data. A schematic overview for a single day and a diurnal cycle of GPP is shown in Fig. 3.

Even if one uses only predictor variables of daily temporal resolution which can be treated as constant for the whole day, different values of the target flux for different half hours of the day can be estimated. The reason is that the 48 different RDF models are learned with different values for the target output variable y, although the same values for the predictor variables $\boldsymbol{x}$ are used. For example, an RDF model that is learned for a half hour during night only covers the rather small range of observations y that can be observed at this time, while the range of observations around noon is typically much larger, especially during the growing season. Hence, the 48 RDF models and their estimated outputs differ only because of different observations y that are provided during learning together with the same set of samples $\mathcal{X}$. Of course, it is also possible to incorporate predictor variables at half-hourly temporal resolution, which would directly fit to the resolution of the target flux. Such predictor variables could further enhance the distinction of individual half hours of a day and could lead to more accurate estimations. However, they are optional and not required for this prediction approach as indicated in Fig. 3.

4.2 Second prediction approach: a single regression model suitable for all half hours of the day

In contrast to the first prediction approach, the second approach only uses a single regression model that is able to estimate different values for different half hours of the same day. It is then necessary that the distinction between these half hours is somehow encoded in the predictor variables, which is not the case if only predictors of daily resolution are incorporated.

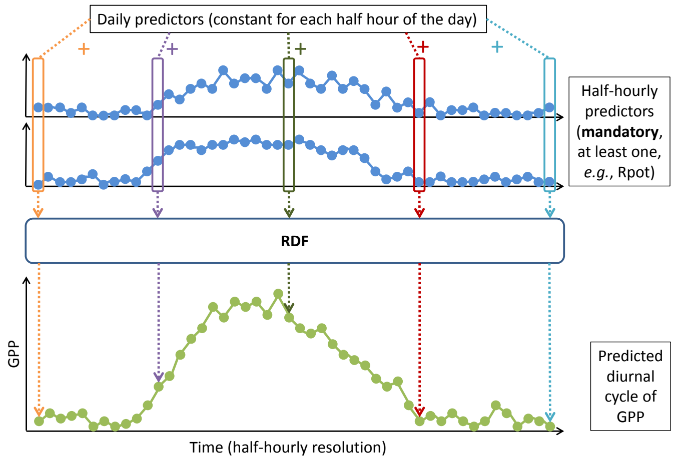

Figure 4. Visualization of the second prediction approach: a single RDF regression model is able to predict the flux at every half hour of the day if at least one predictor variable has a half-hourly temporal resolution (such as the potential radiation Rpot).

Therefore, this approach requires at least one predictor variable at half-hourly temporal resolution. Fortunately, the potential radiation Rpot can be calculated globally at half-hourly resolution, because it only depends on the time as well as the solar angle that is defined given the spatial position via latitude and longitude. Thus, the second approach with a single model, as visualized in Fig. 4, is therefore also applicable for upscaling diurnal cycles to global scale. Again, we make use of an RDF regression model due to its large-scale capabilities.

In addition, besides the potential radiation also its first-order temporal derivative can be incorporated as an additional half-hourly predictor, which helps for distinguishing between morning and afternoon. For all our computations, we have always included this derivative in case we also used Rpot. The nice property of this approach is that information about the physical relationships between the predictor variables and the fluxes can be shared among different half hours during learning of the single regression model, which is not the case for individual models as mentioned in the previous section. This second prediction approach therefore seems to be more plausible from a physical perspective, because the distinction between different half hours of the day is made based on the data, e.g., (potential) radiation, and not enforced by learning independent regression models for each half hour.

Although meteorological variables such as air temperature or vapor pressure deficit (VPD) as well as incoming radiation are also potential candidates for predictors that encode subdaily variations of the fluxes, they are currently only available with a half-hourly resolution at individual sites, e.g., also measured at eddy covariance towers. Due to the missing half-hourly meteorological data products at global scale, it is not possible to use these information for the global upscaling. However, since we are interested in whether such data products could be beneficial for upscaling diurnal cycles, we use the corresponding site-level data in our cross-validation analysis to get further insights. Hence, meteorological variables measured at the eddy covariance towers of FLUXNET can still be used for validating the upscaling approaches and evaluations of cross-validation experiments are presented in the next section.

5 Assessing different upscaling strategies with leave-one-site-out cross-validation

The global products presented in this paper cover diurnal cycles of four fluxes: GPP, NEE, LE, and H. For each of these fluxes, we have consistently performed cross-validation experiments but the results presented in the following only consider GPP as a running example. We have decided to apply random decision forests (RDF) for regression due to its efficient training and testing algorithms even in case of large-scale data as well as its good performance for upscaling daily mean values of GPP (Tramontana et al., 2016). Each RDF was trained with 100 randomized decision trees, because we observed a saturation effect for the prediction performance in preliminary experiments when increasing the number of decision trees. Further parameters have been set to its default values in Matlabs TreeBagger function, e.g., a minimum leaf size of five samples, since we hardly observed any changes in the overall performances when varying the parameter settings. Performances are measured using the Nash-Sutcliffe modeling efficiency (Nash and Sutcliffe, 1970) based on a leave-one-site-out cross-validation scheme.

The motivation for the leave-one-site-out evaluation as a special case of cross-validation is twofold. First, we want to evaluate regression models that have been learned from as many observations as possible and based on training sets that are most similar to the training set that will be used to compute the global products, which will incorporate all the available data from all FLUXNET sites. Second, we intend to mimic a realistic scenario most similar to the upscaling task by predicting fluxes at locations where no training data has been taken from. As a consequence, we predict fluxes at one FLUXNET site using a regression model learned with all observations from all the remaining FLUXNET sites. After doing this for each individual site, we concatenate all site-specific predictions to form a long vector of predictions that can be compared to the corresponding observations measured at the corresponding sites. This allows for a general evaluation of the prediction approaches in a site-independent manner.

5.1 Overview of experiments

We start with a short overview of the experiments that have been conducted in order to make clear our ideas and motivations behind them. In Sect. 5.2, we compare the two different prediction approaches for upscaling diurnal cycles that have been introduced in Sect. 4. Furthermore, we focus on comparing different sets of predictor variables, e.g., the effect of meteorological variables at half-hourly resolution on the prediction performances. Evaluations of the prediction performance for monthly average diurnal cycles derived from the half-hourly values are shown in Sect. 5.3. These average diurnal cycles per month nicely summarize the fluxes over a longer time period (one month) by still keeping a half-hourly pattern that allows for monitoring subdaily variations. In addition, averaging diurnal cycles for a specific month removes noise in the individual half-hourly measurements and reduces the effects of day-to-day variability, e.g., caused by cloud coverage, which allows for comparing the main characteristics of the observations and the predictions for the selected month. The evaluations of monthly average diurnal cycles play an important role for our provided data products, since we also prepare derived products that contain these monthly average patterns only. With two additional experiments presented in Sect. 5.4, we want to demonstrate that the quality of our achieved predictions is not inherently limited by the presented upscaling approaches but rather by missing site-specific information and latent driving forces that are not encoded in the set of predictor variables that has been used. This is

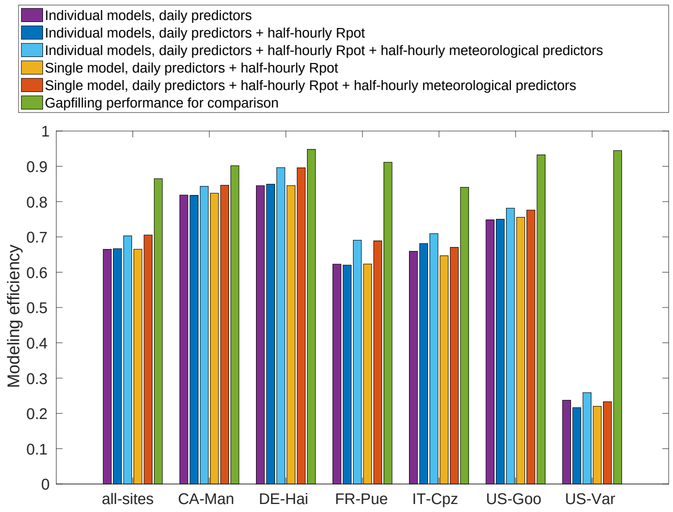

Figure 5. Prediction performances for half-hourly values of GPP depending on different sets of predictor variables are shown. We compare our two proposed prediction approaches (individual RDF models and single RDF model) and also include the results of a gapfilling algorithm for comparison. Looking at the results of all sites as well as at site-specific performances, we observe that meteorological predictor variables at half-hourly resolution clearly improve the accuracies of the estimations.

not a specific problem of upscaling diurnal cycles of fluxes at half-hourly resolution, but a general challenge for all upscaling approaches that deal with carbon and energy fluxes, also at coarser time scales.

5.2 Improved predictions by using half-hourly meteorological data

In the following, we compare the results of our presented prediction approaches for half-hourly GPP depending on different sets of predictor variables, which have been obtained by using the leave-one-site-out strategy explained in the beginning of Sect. 5. As the core for all sets of predictors, we include those variables that have been used for upscaling daily mean GPP values by Tramontana et al. (2016). In fact, we use exactly the set of predictors that corresponds to the RS+METEO setup that has been defined by Tramontana et al. (2016, Table 2) and we refer to this table for details about the variables. Given the explanations of the first prediction approach in Sect. 4.1 with individual regression models for each half hour, we can directly use these predictor variables for estimating half-hourly GPP values. However, we also added the potential incoming radiation (Rpot) at half-hourly resolution to encode subdaily variations in the predictors as well as its first temporal derivative to distinguish between morning and afternoon. Furthermore, we have tested a third set of predictors by additionally incorporating meteorological variables at half-hourly resolution, which are measured at FLUXNET towers, namely air temperature, vapor pressure deficit, and incoming global radiation. The second and third set are also used in the experiments for the second prediction approach that includes only a single regression model, because half-hourly information is encoded in some of the predictor variables.

In Fig. 5, we have visualized the results for all sites as well as for selected FLUXNET towers. Only for comparison to the leave-one-site-out experiments, we also included the modeling efficiency of a gapfilling algorithm (Reichstein et al., 2005) as a potential upper bound for our predictions. In fact, each measured value is also estimated by a gapfilling algorithm that makes use of measurements at the same site under similar climate conditions and hence provides only a theoretical baseline, because it can not be applied for predicting fluxes at locations without any observations. First, we focus on the leftmost group of bars in Fig. 5, which shows the modeling efficiencies for all sites. Looking at the results for the first prediction approach with individual models for each half hour, including half-hourly Rpot only slightly improves the average performance (0.67 compared to 0.66), which is probably also caused by the stochastic nature of the RDF learning algorithm. However, including the meteorological predictors at subdaily temporal resolution leads to an increase of the performance to 0.70 modeling efficiency. A similar improvement can be observed for the second prediction approach with a single model for every half hour of the day, for which the modeling efficiency increases from 0.67 to 0.71 when including half-hourly meteorological data. This highlights that varying subdaily meteorological conditions have a clear impact on the prediction of diurnal cycles of GPP fluxes. On the one side, half-hourly Rpot has almost no influence on the accuracy of the predictions but on the other side, it allows for applying the second prediction approach with only a single regression model. Hence, it may seem more natural from a physical perspective to distinguish individual half hours of the day by the provided half-hourly Rpot rather than enforcing the distinction by separately learned regression models. Comparing both prediction strategies, they achieve similar performances when using the same set of predictor variables. Since it is more convenient from a technical perspective to only handle a single regression model instead of 48 different models, the evaluations in the following sections will focus on the second prediction approach with a single RDF model that is suitable for predicting values at every half hour of the day. It is interesting to note that relative performance differences between the two prediction approaches and among the different sets of predictor variables look very similar when considering single sites only. In all our cross-validation experiments, the best prediction accuracies are always achieved by including half-hourly meteorological variables in the set of predictors. However, absolute performance values vary among sites. As shown in Fig. 5, the accuracies at the sites CA-Man and DE-Hai are between 0.80 and 0.90 modeling efficiency, whereas lower performances (between 0.60 and 0.80) have been achieved for predicting GPP at FR-Pue, IT-Cpz, and US-Goo. Moreover, the fluxes at US-Var seem to be very difficult to estimate, since only modeling efficiencies between 0.20 and 0.30 were obtained.

To further highlight the difference in the predictions when half-hourly meteorology is encoded in the driver variables, we visualize all half-hourly estimations over one year at a specific site using fingerprint plots. A fingerprint in this context is a plot with 365 rows corresponding to 365 days of a year and 48 columns corresponding to 48 half hours of each day such that one fingerprint contains all half-hourly values of a whole year and shows characteristic patterns for the selected site, e.g., length of the growing season. In Fig. 6, the estimations of half-hourly GPP with and without half-hourly meteorological predictors are shown in two individual fingerprint plots and their difference is indicated in a third plot. As expected, the predictions based on half-hourly meteorology contain much more short-term fluctuations during single days, whereas smoother estimations are obtained when only half-hourly Rpot is used as a driver. This can also be observed from the difference of the two fingerprints.

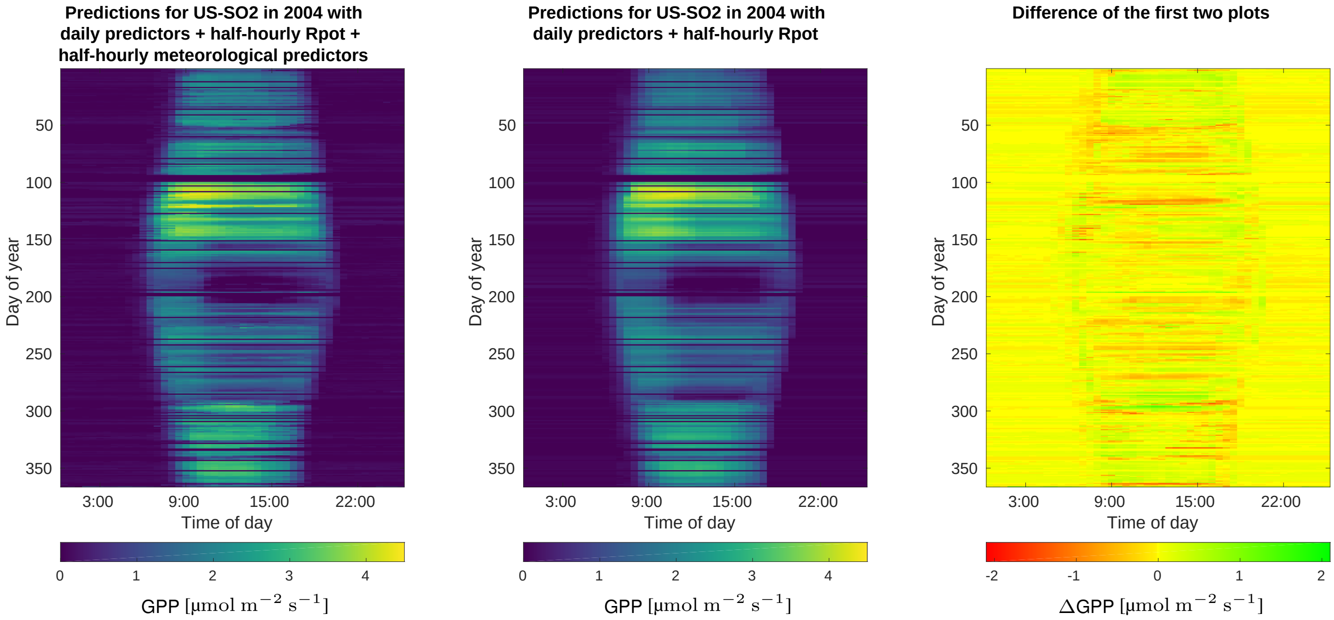

Figure 6. Fingerprint plots of half-hourly GPP fluxes estimated for US-SO2 in 2004 with leave-one-site-out cross-validation show that short-term fluctuations on subdaily scales are captured better when also half-hourly meteorological predictors have been included (left) compared to only using half-hourly Rpot (center). The difference of the first two plots (right) also emphasizes this observation.

Hence, half-hourly meteorological predictor variables are required to better capture high-frequency changes of the fluxes on subdaily scales. In the following, we take a closer look at average diurnal cycles per month.

5.3 Analyzing average diurnal cycles per month

For visual inspection purposes, it is useful to look at average diurnal cycles for individual months at specific sites. Example plots are shown in Fig. 7. They show that our predictions are able to produce the typical shapes of diurnal courses which are in line with corresponding observations. For the depicted predictions, only data from other sites has been used to learn the regression models. It is important to note that averaging diurnal cycles within a month reduces noise in the observations as well as in the predictions of a single day, but also smoothens high frequency short-term fluctuations, e.g., due to (partial) cloud coverage, and yet decreases the influence of day-to-day variations. Hence, these mean diurnal courses are more stable and an evaluation of averaged predictions with respect to averaged observations for all sites leads to larger modeling efficiencies compared to those reported in the previous section. An overview of modeling efficiencies when comparing all half-hourly values versus only looking at the average diurnal cycles is given in Table 1.

In fact, modeling efficiencies for monthly average diurnal cycles increase on average across all sites to a range between 0.78 and 0.80 depending on the set of predictor variables with the best results being accomplished again by incorporating half-hourly meteorological data. This can also be observed from Fig. 8, which is organized in the same way as Fig. 5 but contains the

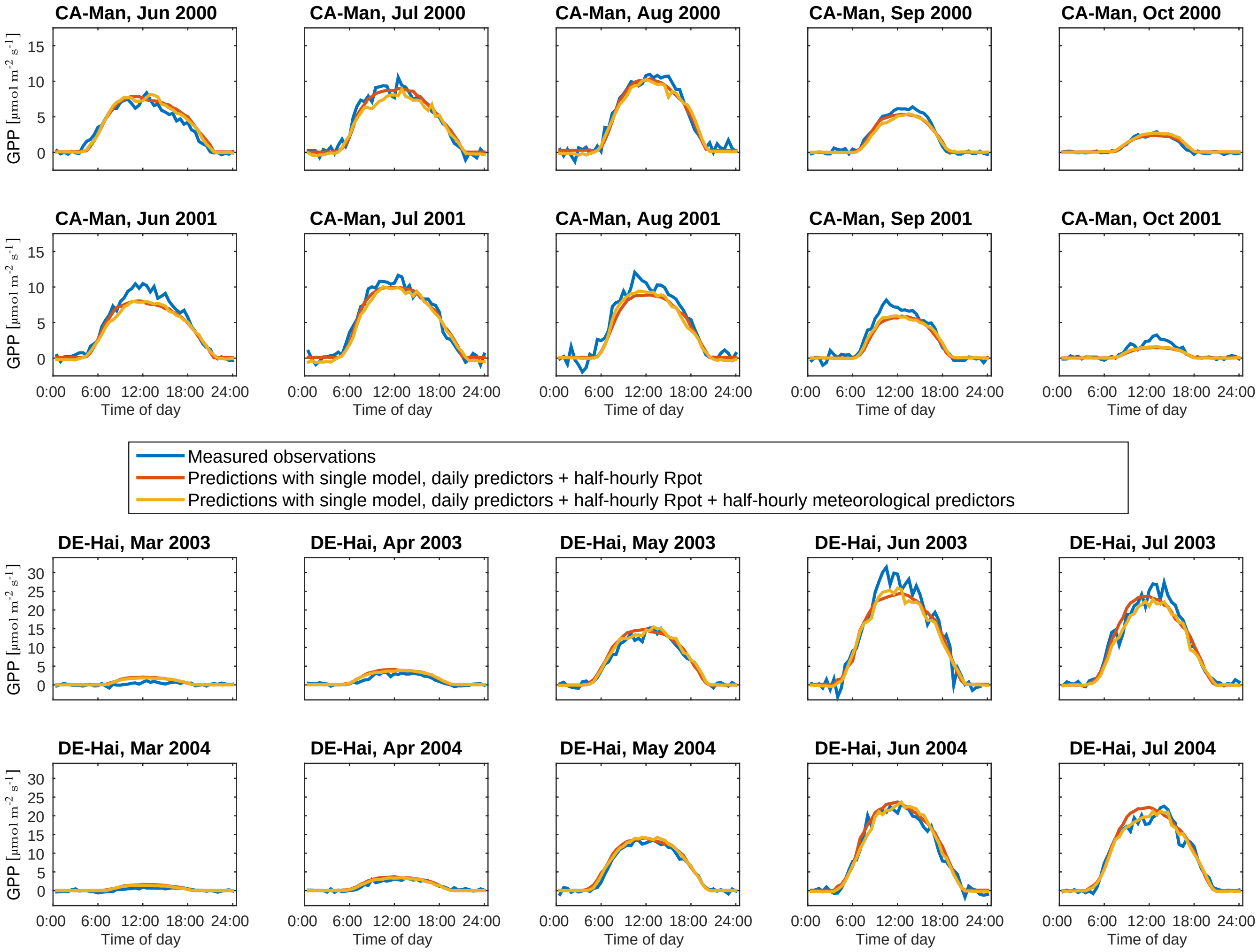

Figure 7. Some example sites with average diurnal cycles for different months comparing two prediction approaches with the observations.

achieved modeling efficiencies for comparing monthly average diurnal cycles of observations and predictions. For the monthly mean diurnal courses, the difference between only using half-hourly Rpot or also including half-hourly meteorology is not so large anymore compared to the evaluations for all half-hourly values. This holds for both the overall accuracies for all sites as well as for single selected sites. As previously mentioned, the reason is that averaging fluxes within a month reduces the effect of short-term fluctuations on subdaily scales. Therefore, if one is only interested in monthly average diurnal cycles, the results obtained by using daily predictors and half-hourly Rpot are only slightly worse compared to including half-hourly meteorology and for some sites, the prediction performances are even on the same level of accuracy. This is important to know, since we also provide a derived product from our global half-hourly fluxes that contains the monthly average diurnal cycles globally at the same spatial resolution (Sect. 6).

Table 1. Modeling efficiencies for the predictions of all sites obtained from the leave-one-site-out experiments are summarized and here we differentiate between comparing all individual half-hourly values with the observations and only looking at monthly average diurnal cycles.

APPROACH	COMPARING ALL HALF-HOURLY VALUES	COMPARING MONTHLY AVERAGE DIURNAL CYCLES
Individual models		
daily predictors	0.66	0.78
daily predictors + half-hourly Rpot	0.67	0.78
daily predictors + half-hourly Rpot + half-hourly meteorological predictors	0.70	0.80
Single model		
daily predictors + half-hourly Rpot	0.67	0.78
daily predictors + half-hourly Rpot + half-hourly meteorological predictors	0.71	0.80
Gapfilling	0.87	0.93

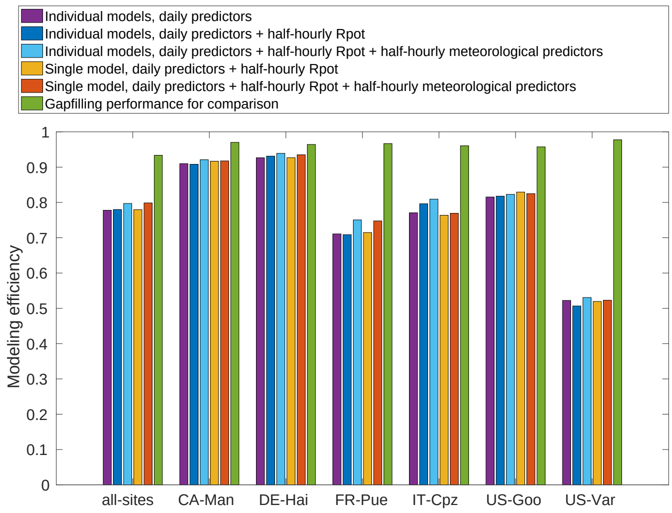

Figure 8. Prediction performances for monthly average diurnal cycles of GPP are shown in the same way as the accuracies for all half-hourly values in Fig. 5.

However, the average diurnal cycles can also be used to identify potential problems of the predictions. In Fig. 9, mean diurnal courses of several months at the sites FR-Pue and IT-Cpz are shown. It can be observed for both sites that the averaged observations are lower in the summer months compared to the corresponding predictions. In other words, the regression models overestimate GPP during these months. We believe that this is caused by the fact that our current prediction models are not able to cope with seasonal droughts, which is not a specific problem of the diurnal upscaling but a challenge that every upscaling

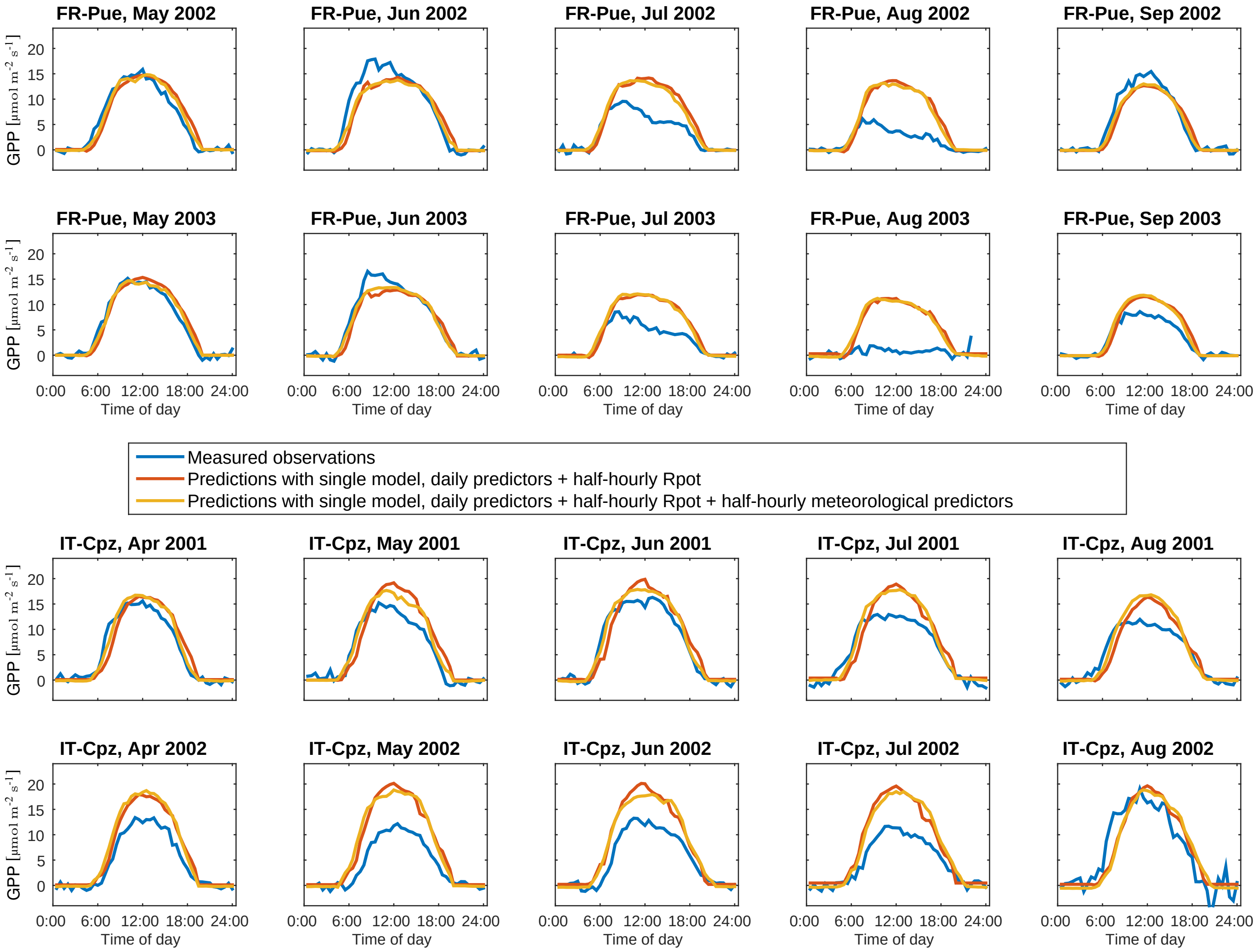

Figure 9. Average diurnal cycles of two sites showing the problems with seasonal droughts. The error in the prediction of the half-hourly fluxes increases during hot and dry summers for both sites, FR-Pue and IT-Cpz.

approach for carbon and energy fluxes needs to tackle. Although the observations show decreased productivity due to drought stress in summer, the regression models still estimate large amplitudes of the diurnal cycles, i.e., a larger productivity. One reason for this behavior could be the insufficient characterization of water availability that is present in the set of predictor variables. Currently, we plan to investigate this issue in further research. In the following, we show that our current sets of predictor variables are lacking some site-specific information, probably not only with respect to water storage capacities.

Table 2. Comparing modeling efficiencies of the two auxiliary experiments (leave-one-month-out setup and including the daily GPP as an additional predictor in the leave-one-site-out setup) to the best performances obtained with the leave-one-site-out experiments by using daily predictor variables, half-hourly potential radiation, and half-hourly meteorological variables.

Approach	Comparing all half-hourly values	Comparing monthly average diurnal cycles
Individual models		
Leave-one-site-out	0.70	0.80
Leave-one-month-out	0.78	0.88
Leave-one-site-out + daily GPP	0.86	0.93
Single model		
Leave-one-site-out	0.70	0.80
Leave-one-month-out	0.79	0.89
Leave-one-site-out + daily GPP	0.87	0.94
Gapfilling	0.87	0.93

5.4 Are we missing (site-specific) information in the predictors?

In order to gain any insights whether site-specific information is currently not well represented in the predictors, we have conducted two auxiliary experiments. During the first experiment, we additionally estimate GPP fluxes at each site in a leave-one-month-out setup and compare the resulting predictions with those of the leave-one-site-out setup. For the leave-one-month-out estimations, we learn and test regression models for each month at each site separately. Furthermore, each regression model for each month is only learned with data from the same site but measured in different months (and years). Hence, the regression models are highly site-specific, since only correspondences between predictor variables and GPP fluxes at a single site are used and predictions are made at the same site but in a different time period. As a result, we have observed improved flux estimations, which is shown exemplarily in Fig. 10 for IT-Cpz. It can be clearly seen that the gaps between averaged observations and averaged predictions are getting smaller and mostly almost disappear, i.e., the predictions match the observations much better in the leave-one-month-out setup. In terms of modeling efficiency, the performances increase to a range between 0.75 and 0.79 when comparing all individual half-hourly predictions from the leave-one-month-out setup at all sites with the observations (best performance with leave-one-site-out is 0.70). Regarding the comparison of averaged predictions with averaged observations as presented in the previous section, the leave-one-month-out setup leads to modeling efficiencies between 0.87 and 0.89. This is clearly larger than the results of the leave-one-site-out-experiment (best performance: 0.80). Table 2 allows for a direct comparison of the results from the leave-one-site-out and the leave-one-month-out experiments using both prediction approaches with the best set of predictor variables, i.e., daily predictor variables, half-hourly Rpot, and half-hourly meteorological variables.

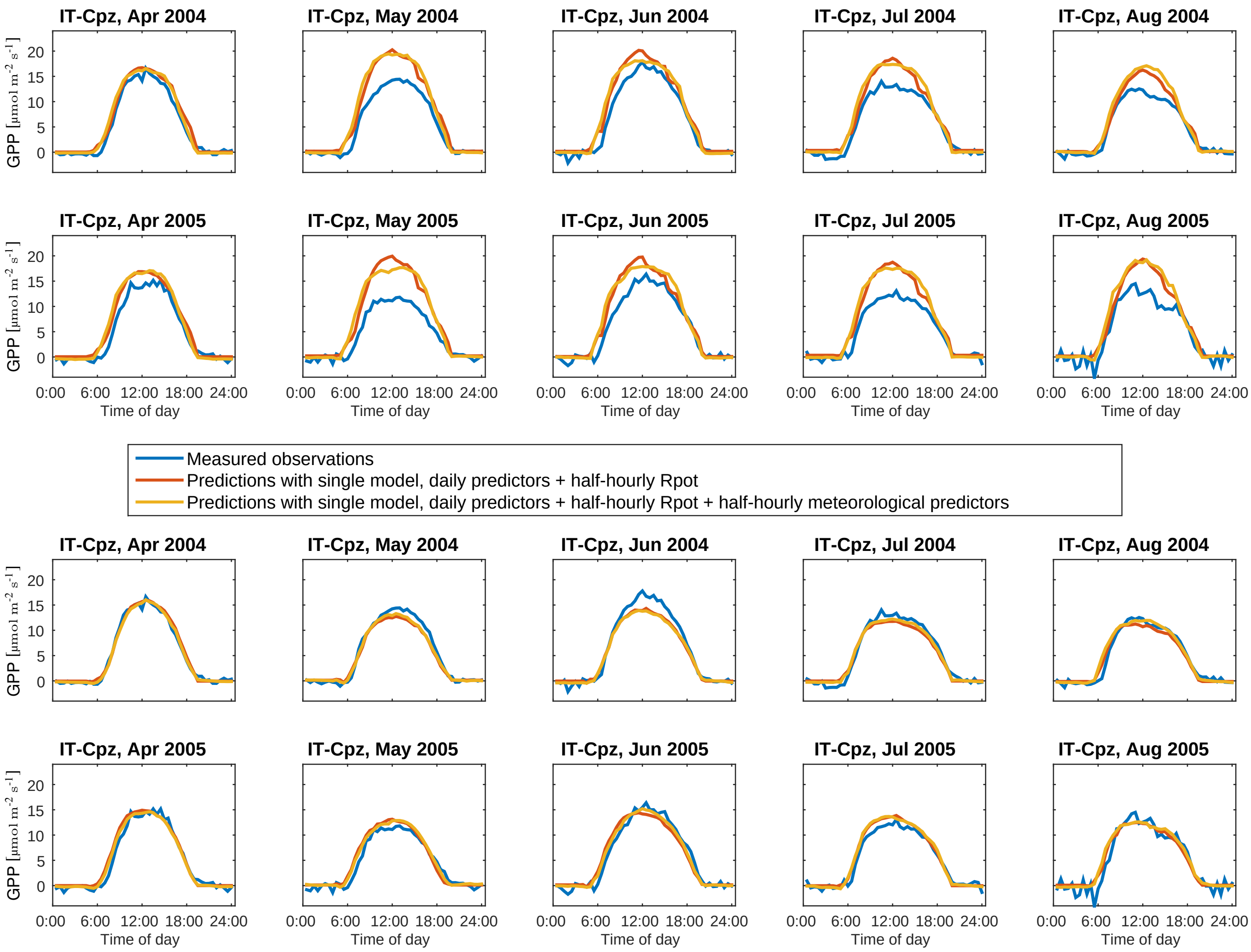

Figure 10. Comparison between leave-one-site-out (top rows) and leave-one-month-out (bottom rows) at IT-Cpz. It can be observed that site-specific training in the leave-one-month-out setup reduces the prediction errors during seasonal droughts. Thus, the drought effects only lead to problems when training across sites and predicting fluxes in the leave-one-site-out setup or for the upscaling when fluxes are estimated at locations where no towers exist.

This table also contains the prediction performances obtained from a second experiment, in which we have used the daily GPP as an additional daily predictor for our regression models in the leave-one-site-out setup. Of course, this is only possible in the cross-validation analysis where we actually have the daily averages of GPP, but the following evaluation reveals interesting insights. Using the daily average GPP basically incorporates information about the amplitudes of the diurnal cycles, hence drought effects of reduced productivity can directly be observed in this additional predictor variable. First of all, it can be seen in Fig. 11 that using the daily GPP as an additional predictor clearly improves the predictions at FR-Pue during summer months. Especially the decreased productivity in July 2005 and August 2005 can nicely be predicted by the regression models.

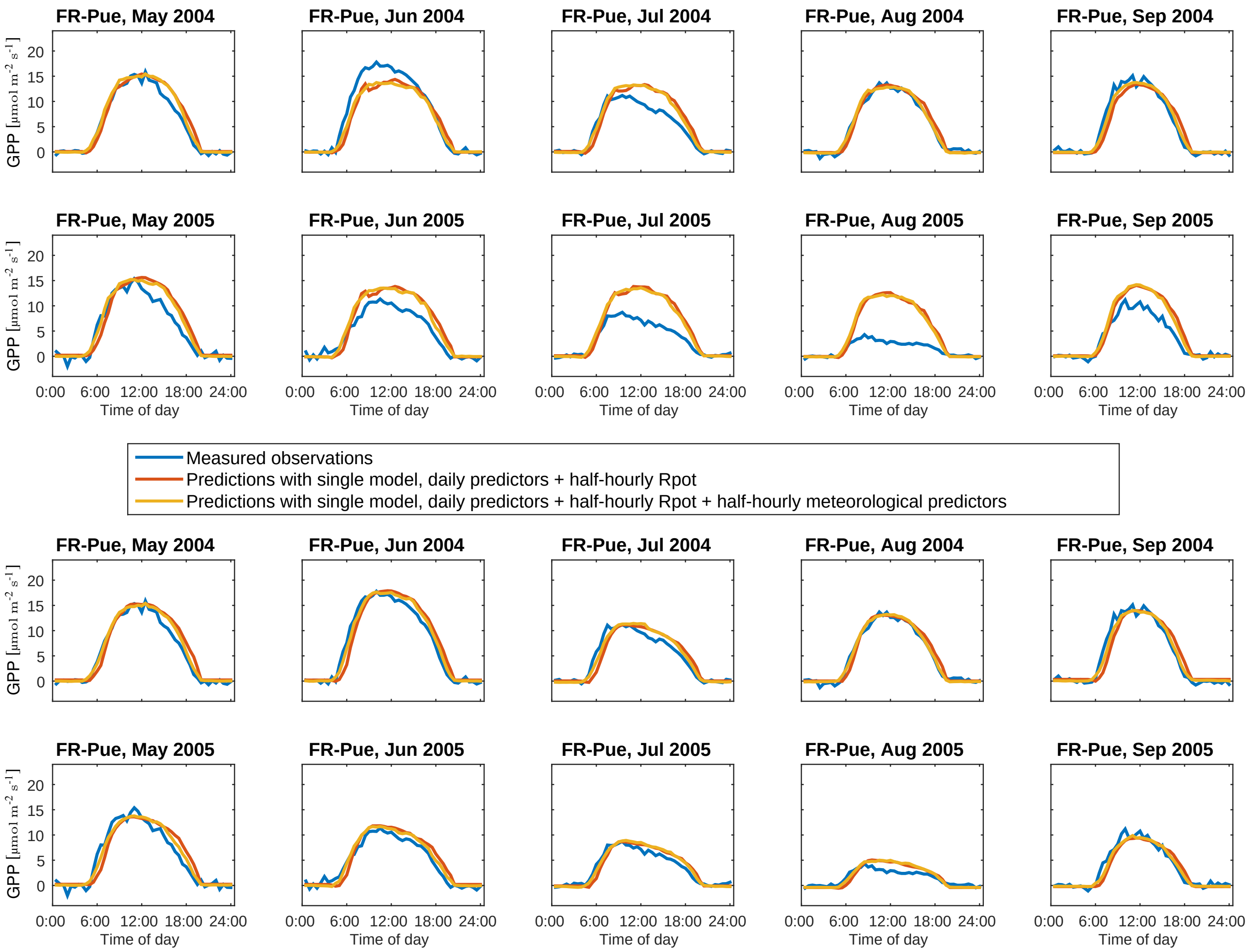

Figure 11. Improvements of the initial estimations (top rows) at FR-Pue can be observed when using daily GPP as an additional daily predictor if it would be available a priori (bottom rows). Hence, the problems with seasonal drought effects would be greatly reduced in the leave-one-site-out setup for every half hour of the diurnal cycle in case an estimate of the daily average value is given.

Since the daily GPP as an additional predictor constrains the size of the peak in a diurnal cycle, the predictions become much more powerful and the characteristic shapes of the diurnal cycles can be produced. The modeling efficiencies are even larger than those obtained with the leave-one-month-out setup. They are in the range of 0.83 to 0.87 for all half-hourly values at all sites, which is comparable with the performance of the gapfilling algorithm that has been included as an additional reference in Fig. 5 as well as in Table 2. The gapfilling also achieves 0.87 modeling efficiency, i.e., the upper performance limit shown as a green bar in the leftmost group in Fig. 5 can be obtained by including the daily GPP as an additional predictor. Regarding monthly averaged diurnal cycles, a modeling efficiency of up to 0.94 is obtained by the regression models that use daily GPP as an additional predictor, while gapfilling reaches 0.93. This is also summarized in Table 2.

From this experiment, we can conclude that the problems for predicting diurnal cycles of GPP are mainly caused by the lack of estimating the daily mean GPP properly. If the daily mean is given, predictions of half-hourly values are much more accurate. Hence, the main problems for the upscaling of half-hourly fluxes are not related to producing the right shapes of the diurnal courses, but turn out to be problems of estimating the correct amplitudes. These are then the same problems as for upscaling daily averge values (or fluxes at coarser time scales) and are not introduced by the step of going to a larger temporal resolution in terms of half hours.

5.5 Key insights from the cross-validation experiments

In this section, we want to shortly summarize the main findings from our cross-validation experiments. First, we have seen that it does not really matter which of the two proposed prediction approaches we are using, since prediction performances hardly differed between the single model and the individual model approach. We prefer to use the single model approach, because it seems to be more plausible from a physical perspective to make distinctions between half hours of a day by the information encoded in the predictor variables and half-hourly Rpot can always be used for this purpose. Second, including half-hourly meteorological information in the predictors clearly helps to improve the prediction performances for fluxes on the half-hourly scale. However, for monthly average diurnal cycles the differences are not so prominent anymore and estimations based on half-hourly Rpot only may be sufficient for analyzing the monthly patterns. Third, we have shown that the main problem for upscaling half-hourly fluxes is not the fact that we increase the temporal resolution, since we are able to reproduce the characteristic subdaily patterns. Moreover, we are lacking additional information in the predictors that encode site-specific characteristics as well as certain special conditions like seasonal droughts. This currently prevents us from getting the day-to-day variability and in the end also the interannual variability right. However, these are also problems that need to be tackled when an upscaling of carbon and energy fluxes at coarser time scales is considered (Tramontana et al., 2016). In the following section, we summarize the results from our cross-validation experiments for all the four fluxes (GPP, NEE, LE, H) with the setup that has been used to compute the global half-hourly products.

5.6 The selected approach for computing the global products

While the previous sections validate the presented prediction approaches and point to potential problems in the estimation of half-hourly fluxes, we also decided to produce first global products of half-hourly GPP and NEE as well as LE and H that will be described in the next section. So far, the analyses have shown that best predictions are obtained by incorporating meteorological variables at half-hourly resolution, but such data products are not available at a global scale. Therefore, we have computed the global products only based on the daily predictors of the RS+METEO setup from Tramontana et al. (2016, Table 2) as well as half-hourly values of Rpot and its first temporal derivative. Note that the used data sources have been described in Sect. 2 and the set of daily predictors varies between carbon and energy fluxes as indicated in the paper of Tramontana et al. (2016, Table 2).

Table 3. Prediction performances in terms of modeling efficiency are estimated from the leave-one-site-out cross-validation experiments with the setup that has been used to compute the global half-hourly products for the four fluxes.

	GPP	NEE	LE	H
Modeling efficiencies related to all individual half-hourly values	0.67	0.61	0.72	0.77
Modeling efficiencies related to monthly mean diurnal cycles	0.78	0.76	0.83	0.86

Furthermore, we have decided to use the second prediction approach (Sect. 4.2) by learning one single regression model that is suitable for all half hours of the day. For us, it seems more natural from a physical perspective to distinguish between different half hours of a day by (potential) radiation as an additional variable rather than enforcing the distinction with individual models for each half hour as it is done in the first prediction approach (Sect. 4.1).

In Table 3, we report the corresponding prediction performances from the leave-one-site-out cross-validation experiments for this setup, i.e., for the selected set of predictor variables and the single regression model approach. The modeling efficiencies for both individual half-hourly values and monthly mean diurnal cycles are stated. Comparing these values, we observe that the accuracies for predicting energy fluxes are higher compared to those for the carbon fluxes. Half-hourly values of the sensible heat flux can be estimated best by achieving a modeling efficiency of 0.77 across all sites. On the other hand, net ecosytem exchange has only been predicted with a modeling efficiency of 0.61. This performance is lower compared to the one for gross primary production (0.67), probably due to missing information in the predictor variables for the respiration component of NEE. For all the four fluxes, the modeling efficiencies are larger when comparing monthly average diurnal cycles of observations and predictions. The main reasons, as also mentioned in Sect. 5.3, are the reduction of noise as well as the smoothing of short-term fluctuations at subdaily scales due to the averaging. In the following section, we present the global half-hourly flux products that have been calculated with the upscaling approach and the setup of this section.

6 Global flux products with half-hourly resolution

For each of the four fluxes (GPP, NEE, LE, H), we have learned a single regression model for all half-hours based on all available half-hourly values of the corresponding flux at the 222 FLUXNET sites, i.e., one regression model for GPP, one for NEE, one for LE, and one for H. These models are then used to compute half-hourly fluxes globally with half degree spatial resolution and continuously from 2001 (January 1st) to 2014 (December 31st) using global forcing data described in Sect. 2.2. As mentioned in the previous section, we have used the daily predictors of the RS+METEO setup from Tramontana et al. (2016, Table 2) as well as half-hourly values of Rpot and its first temporal derivative. Furthermore, it should be noted that the global products have been initially calculated such that they are tiled by PFT. The final flux for each point in space and time has then been determined as a weighted sum depending on the percentage of each PFT to be present in the corresponding grid cell. When looking at annual sums of the half-hourly data products, we observe that these sums are pretty constant among the different years for the individual fluxes. On average, we get 125.94 Pg C for GPP and -21.42 Pg C for NEE as well as 182.22 ZJ for LE and 144.79 ZJ for H.

In addition to the provided half-hourly data, we also offer derived products containing the monthly average diurnal cycles of the four fluxes for the 14 years that are covered by the half-hourly product. For the potential user of the data, it will be much more convenient to directly obtain the monthly average diurnal cycles compared to downloading the much larger half-hourly data product and computing the monthly averages afterwards. Furthermore, the monthly average diurnal cycles are more robust, which has also been shown by larger modeling efficiencies in the experimental evaluations, e.g., as listed in Table 1. Since only daily predictor variables and half-hourly Rpot are used to estimate the half-hourly fluxes, short-term fluctuations on subdaily scales due to cloud cover and other effects can not be captured by the current version of the product. Therefore, also day-to-day variations may not be represented accordingly. However, the averaging to create monthly mean diurnal cycles reduces the impact of these factors and additionally smoothens errors due to observation noise. As a consequence, we recommend to primarily use the monthly average diurnal cycles because of larger robustness and stability. In the following, we show some characteristics of the computed global flux products at half-hourly resolution, which can only be calculated due to the subdaily time scale.

6.1 Global maps and fingerprints

Cutouts of the global products are visualized in Fig. 12. Global maps of GPP and NEE for a specific half hour on June 28th in 2014 are shown in the top row of Fig. 12 and one can nicely distinguish daytime from nighttime for individual regions around the world. Furthermore, selected locations are highlighted and all half-hourly values of the year 2014 for these grid cells are summarized in fingerprint plots, which allow for identifying different characteristics at the corresponding places due to the different patterns in these plots. The fingerprints provide a nice overview of the half-hourly fluxes over a whole year and different lengths of the growing season as well as varying lengths of the day (time between sunrise and sunset) can directly be observed. Corresponding maps for LE and H with fingerprint plots for the same locations are shown in the bottom row of Fig. 12. Larger values of LE compared to H can be observed in Western, Central, and Eastern Europe as well as in the tropical regions of Africa, whereas it is vice versa at the Iberian Peninsula as well as in the northern and southern regions of Africa.

6.2 Maximum diurnal amplitudes within a month

Besides the fingerprint plots summarizing a whole year of half-hourly values for a specific location, it is also possible to compute diurnal amplitudes for each grid cell from the global products. We again picked GPP acting as an example for all the fluxes and determined maximum diurnal amplitudes within each month. In Fig. 13, the maximum diurnal amplitudes of GPP are visualized for June 2014 and December 2014 indicating differences between summer and winter. The biosphere at the northern hemisphere is quite active in June showing large amplitudes, whereas maximum amplitudes are equal or close to zero at almost every grid cell of the northern hemisphere in December. In the tropics, amplitudes of GPP do not vary much between June and December with values around 30 $\mu\mathrm{mol\ m^{-2}\ s^{-1}}$. As expected, the behavior at the southern hemisphere is opposite

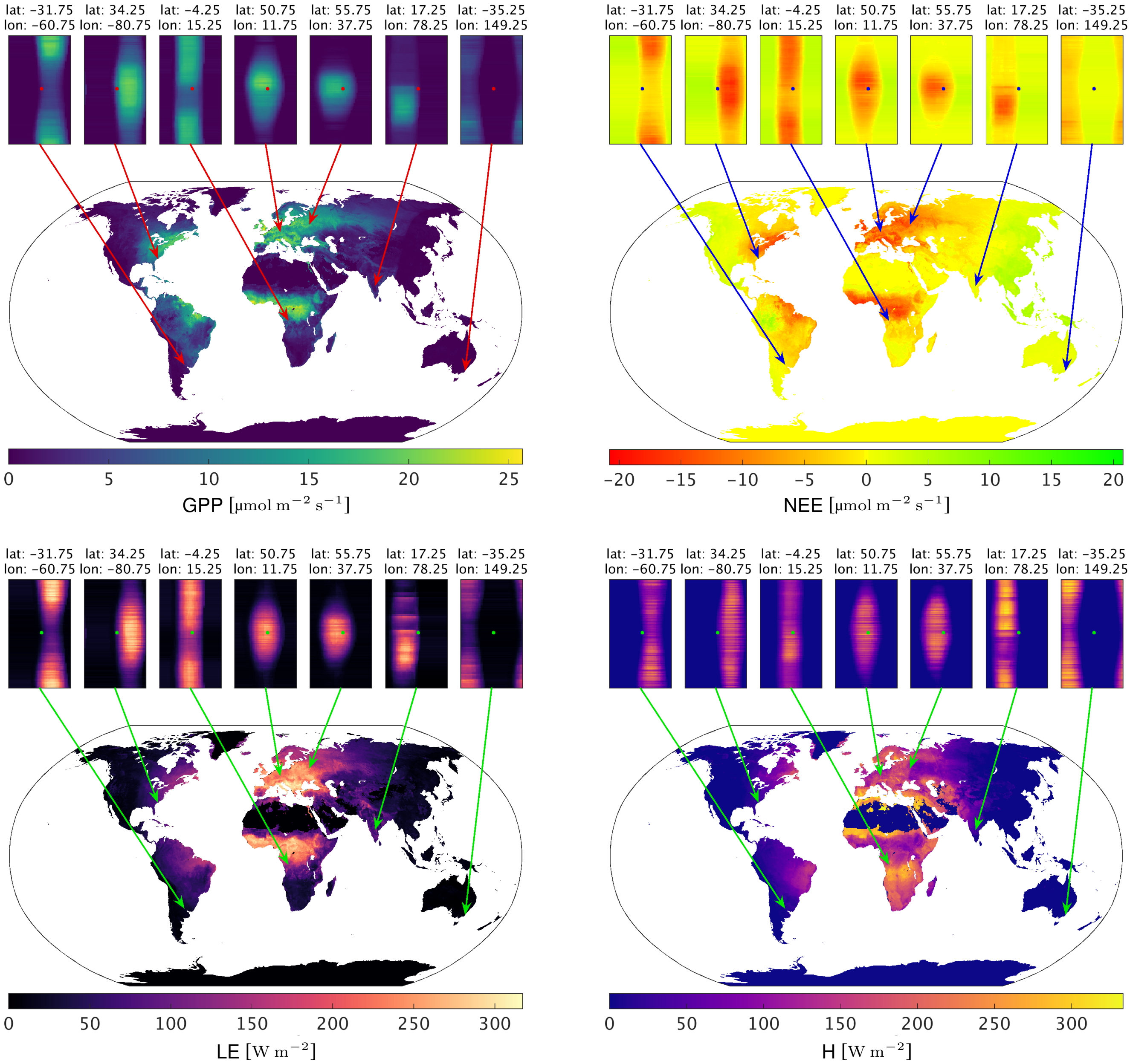

Figure 12. The global maps show estimated values for half-hourly GPP (top left), NEE (top right), LE (bottom left), and H (bottom right) on June 28th, 2014 at 1 pm UTC. In addition, fingerprints for selected grid cells are used to visualize half-hourly values for each day of the year. The dot in each fingerprint marks the value that is shown in the global map. Note that the fingerprints display different extensions of the growing season in different regions and the global maps allow for distinguishing between daytime (e.g., in Europe and Africa) and nighttime (e.g., in East Asia, Australia, and in the western parts of North America).

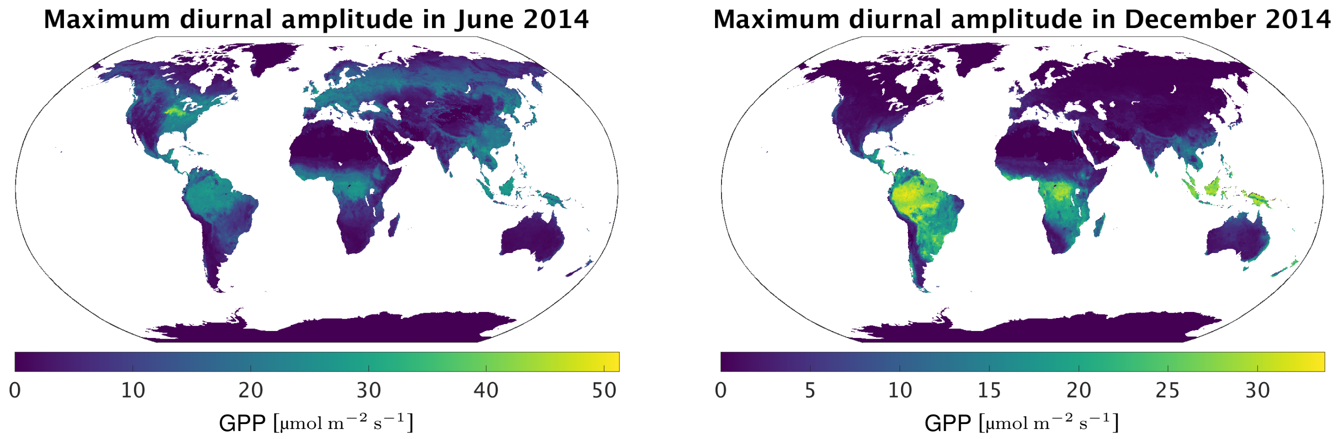

Figure 13. Maximum diurnal amplitudes of GPP within a month are shown for June 2014 (left) and December 2014 (right). Differences between summer and winter for both the northern and the southern hemisphere as well as (almost) constant productivity in tropical regions can be observed from both maps.

to the northern hemisphere, which can be observed at those regions that actually contain vegetation, e.g., large parts of South America and South Africa as well as coastal regions of Australia.

6.3 Maximum half-hourly fluxes

Finally, we have been interested in the maximum flux at each spatial position. These statistics have been calculated among all the years 2001 to 2014 to produce a single map of maximum half-hourly values for each flux. The results are shown in Fig. 14. Those maximum values denote the capabilities of each flux at each grid cell. For GPP, the hot spots with maximum capacities are in the corn belt of the USA, in Eastern China, as well as in the tropical regions. Largest values of NEE are obtained in the tropics as well, especially in the Amazon. Regarding the energy fluxes, it is not so easy to identify few single hot spot regions since large values of LE or H are widespread. However, distinct spatial patterns can be observed in all maps of the maximum fluxes.

7 Conclusions and future work

In this paper, we have shown how to perform an upscaling of half-hourly carbon and energy fluxes from local in-situ measurements to global scale. We have introduced two general prediction approaches to estimate half-hourly values mainly from predictor variables at coarser temporal resolution. Since the problem has been formulated as a large-scale regression task, we have been working with random forest regression although other regression algorithms could be applied as well. Our prediction approaches have been validated by a set of cross-validation experiments employing a leave-one-site-out strategy for the FLUXNET towers that provide the observations. As a result of our analyses, we have presented global flux products at

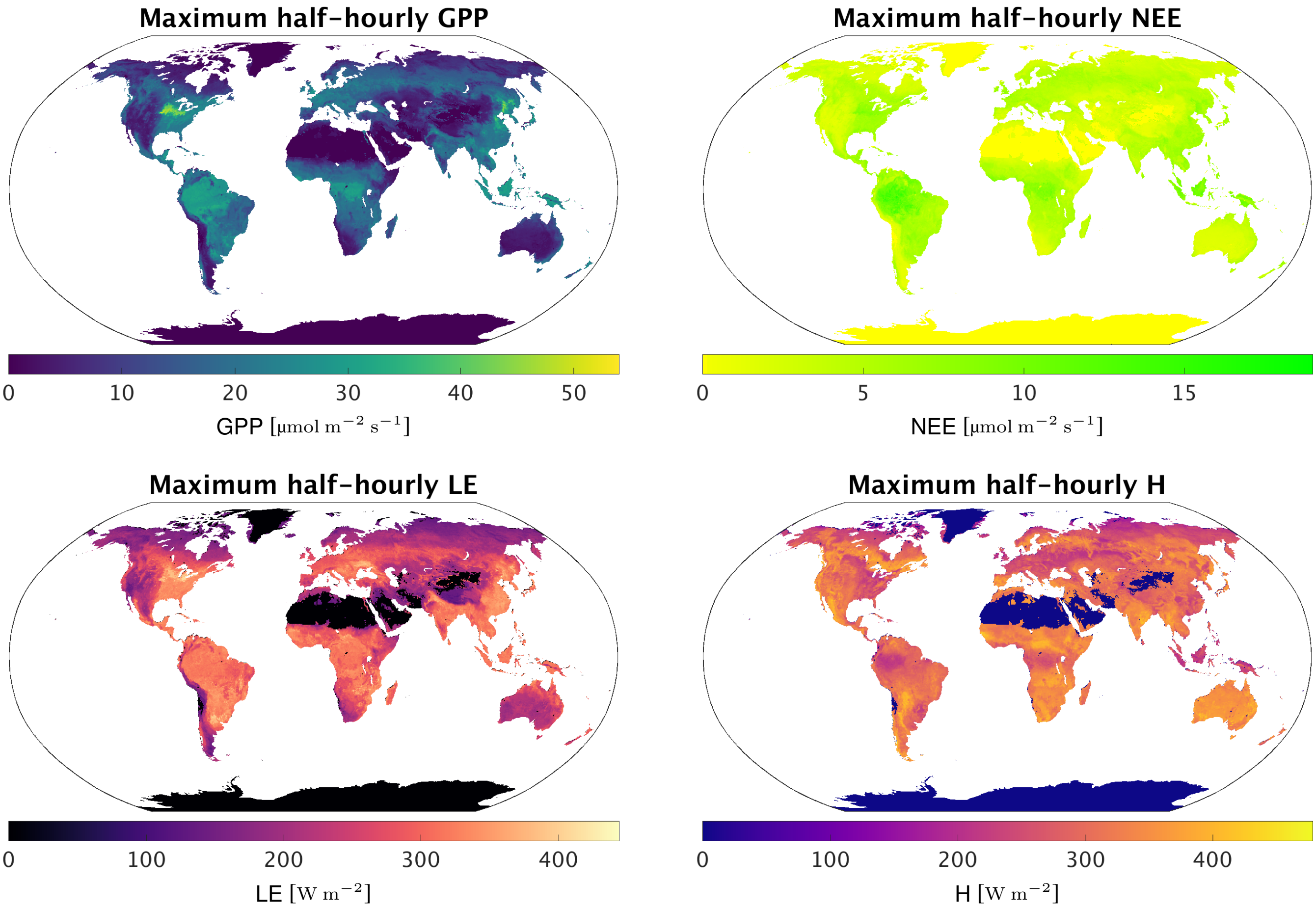

Figure 14. Maximum half-hourly values of GPP (top left), NEE (top right), LE (bottom left) and H (bottom right) during the years 2001 to 2014 are shown for each grid cell.

half-hourly temporal resolution for the years 2001 to 2014 covering four important variables: gross primary production, net ecosystem exchange, latent heat, and sensible heat. Detailed descriptions of the experimental setup for the cross-validation as well as for the computations that have led to the global products were given as well. For the global products, we have also shown derived statistics like maximum diurnal amplitudes of a month as well as maximum half-hourly fluxes at each spatial position. These properties can only be computed from data products with subdaily temporal resolution showing the benefits of our contributions.

In future work, we aim at improving the prediction performance of half-hourly fluxes in various ways. First, we plan to add additional sources of information to the drivers by extending the set of predictor variables to cover further relevant aspects for the individual fluxes like water availability or soil properties. This would allow for tackling difficult scenarios like seasonal droughts, where the current approaches have shown larger errors in the prediction. Second, we also want to incorporate the history of the predictor variables in order to account for lagged effects. So far, samples are treated independently in the prediction but their temporal context due to the time series characteristics may provide additional knowledge that can be exploited for the estimation of fluxes. Third, subdaily meteorology could be included in the calculations of the global products by incorporating new generation meteorological reanalysis data of ERA5 at hourly time scale that will be released in the near future or by exploiting observations from geostationary satellites. Of course, the global products will be updated if these additional ideas lead to better prediction performances. Another import aspect of future work is providing uncertainties for the flux estimations, which could be done by quantile regression approaches (Meinshausen, 2006).

8 Data availability

The calculated global flux products at half-hourly temporal resolution will be made publically available for free upon acceptance of the paper under a creative commons license (CC-BY). More precisely, gridded products at half degree spatial resolution and half-hourly temporal resolution will be provided covering GPP, NEE, LE, and H for the years 2001 to 2014. In addition, a derived product of monthly average diurnal cycles globally for these four fluxes and the given range of years at the same spatial resolution will be prepared for download as well. It will be much more convenient for the potential user to just download the lightweight data of the monthly averages compared to getting the half-hourly data of much larger file size and then performing the averaging on the local machine. As mentioned in the paper, the monthly average diurnal cycles are primarily recommended for usage, since this derived product turned out to be more robust. A permanent link, a DOI, and further information for downloading will be given in this section.

Acknowledgements. The work presented in this paper is part of the project: *Detecting changes in essential ecosystem and biodiversity properties – towards a Biosphere Atmosphere Change Index: BACI*. This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No. 640176.

References

Aubinet, M., Vesala, T., and Papale, D., eds.: Eddy Covariance: A Practical Guide to Measurement and Data Analysis, Springer Atmospheric Sciences, Springer, 1 edn., doi:10.1007/978-94-007-2351-1, 2012.

Balddocchi, D.: Measuring fluxes of trace gases and energy between ecosystems and the atmosphere – the state and future of the eddy covariance method, Global Change Biology, 20, 3600–3609, 2014.

Baldocchi, D. D., Hincks, B. B., and Meyers, T. P.: Measuring Biosphere-Atmosphere Exchanges of Biologically Related Gases with Micrometeorological Methods, Ecology, 69, 1331–1340, doi:10.2307/1941631, 1988.

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

Bonan, G. B.: Forests and climate change: forcings, feedbacks, and the climate benefits of forests, Science, 320, 1444–1449, doi:10.1126/science.1155121, 2008.

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, Journal of Geophyical Research: Biogeosciences, 116, G02 014, doi:10.1029/2010JG001593, 2011.

Breiman, L.: Bagging Predictors, Machine Learning, 24, 123–140, doi:10.1007/BF00058655, 1996.

Breiman, L.: Random Forests, Machine Learning, 45, 5–32, doi:10.1023/A:1010933404324, 2001.

Chu, H., Baldocchi, D. D., John, R., Wolf, S., and Reichstein, M.: Fluxes all of the time? A primer on the temporal representativeness of FLUXNET, Journal of Geophysical Research: Biogeosciences, 122, 289–307, doi:10.1002/2016JG003576, http://dx.doi.org/10.1002/2016JG003576, 2016JG003576, 2017.

Dirmeyer, P. A., Cash, B. A., Kinter III, J. L., Stan, C., Jung, T., Marx, L., Towers, P., Wedi, N., Adams, J. M., Altshuler, E. L., Huang, B., Jin, E. K., and Manganello, J.: Evidence for Enhanced Land-Atmosphere Feedback in a Warming Climate, Journal of Hydrometeorology, 13, 981–995, doi:10.1175/JHM-D-11-0104.1, 2012.

Friedl, M. A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., and Huang, X.: MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets, Remote Sensing of Environment, 114, 168–182, doi:10.1016/j.rse.2009.08.016, 2010.

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, Proceedings of the National Academy of Sciences, 111, E1327–E1333, doi:10.1073/pnas.1320008111, http://www.pnas.org/content/111/14/E1327.abstract, 2014.

Ichii, K., Ueyama, M., Kondo, M., Saigusa, N., Kim, J., Alberto, M. C., Ardö, J., Euskirchen, E. S., Kang, M., Hirano, T., Joiner, J., Kobayashi, H., Marchesini, L. B., Merbold, L., Miyata, A., Saitoh, T. M., Takagi, K., Varlagin, A., Bret-Harte, M. S., Kitamura, K., Kosugi, Y., Kotani, A., Kumar, K., Li, S. G., Machimura, T., Matsuura, Y., Mizoguchi, Y., Ohta, T., Mukherjee, S., Yanagi, Y., Yasuda, Y., Zhang, Y., and Zhao, F.: New data-driven estimation of terrestrial CO2 fluxes in Asia using a standardized database of eddy covariance measurements, remote sensing data, and support vector regression, Journal of Geophysical Research: Biogeosciences, 122, 767–795, doi:10.1002/2016JG003640, 2017.

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., Niovy, N., Weber, U., Williams, C., Wood, E., Zaehle, S., and Zhang, K.: Recent decline in the global land evapotranspiration trend due to limited moisture supply, Nature, 467, 951–954, 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 Geophysical Research: Biogeosciences, 116, G00J07, doi:10.1029/2010JG001566, 2011.

Jung, M., Reichstein, M., Schwalm, C. R., Huntingford, C., Sitch, S., Ahlström , A., Arneth, A., Camps-Valls, G., Ciais, P., Friedlingstein, P., Gans, F., Ichii, K., Jain, A. K., Kato, E., Papale, D., Poulter, B., Raduly, B., Rödenbeck, C., Tramontana, G., Viovy, N., Wang, Y.-P., Weber, U., Zaehle, S., and Zeng, N.: Compensatory water effects link yearly global land CO2 sink changes to temperature, Nature, 541, 516–520, doi:10.1038/nature20780, 2017.

Lasslop, G., Reichstein, M., Papale, D., Richardson, A., 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 Biology, 16, 187–208, doi:10.1111/j.1365-2486.2009.02041.x, 2010.

Meinshausen, N.: Quantile regression forests, Journal of Machine Learning Research, 7, 983–999, 2006.

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

Papale, D., Black, T. A., Carvalhais, N., Cescatti, A., Chen, J., Jung, M., Kiely, G., Lasslop, G., Mahecha, M. D., 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 Geophysical Research: Biogeosciences, 120, 1941–1957, doi:10.1002/2015JG002997, http://dx.doi.org/10.1002/2015JG002997, 2015JG002997, 2015.

Rasmussen, C. E. and Williams, C. K. I.: Gaussian Processes for Machine Learning, The MIT Press, 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., Seufert, G., Tenhunen, J., 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 Biology, 11, 1424–1439, doi:10.1111/j.1365-2486.2005.001002.x, 2005.

Sun, Y., Fu, R., Dickinson, R., Joiner, J., Frankenberg, C., Gu, L., Xia, Y., and Fernando, N.: Drought onset mechanisms revealed by satellite solar-induced chlorophyll fluorescence: Insights from two contrasting extreme events, Journal of Geophysical Research: Biogeosciences, 120, 2427–2440, doi:10.1002/2015JG003150, http://dx.doi.org/10.1002/2015JG003150, 2015JG003150, 2015.

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

Ueyama, M., Ichii, K., Iwata, H., Euskirchen, E. S., Zona, D., Rocha, A. V., Harazono, Y., Iwama, C., Nakai, T., and Oechel, W. C.: Upscaling terrestrial carbon dioxide fluxes in Alaska with satellite remote sensing and support vector regression, Journal of Geophyiscal Research: Biogeosciences, 118, 1266–1281, doi:10.1002/jgrg.20095, 2013.

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., 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., and Torn, M. S.: Estimation of net ecosystem carbon exchange for the conterminous United States by combining MODIS and AmeriFlux data, Agricultural and Forest Meteorology, 148, 1827–1847, 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 U, 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 Sensing of Environment, 114, 576–591, doi:10.1016/j.rse.2009.10.013, 2010.

Xiao, J., Chen, J., Davis, K. J., and Reichstein, M.: Advances in upscaling of eddy covariance measurements of carbon and water fluxes, Journal of Geophysical Research: Biogeosciences, 117, G00J01, doi:10.1029/2011JG001889, 2012.

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 Sensing of Environment, 110, 109–122, doi:10.1016/j.rse.2007.02.016, 2007.