Biogeosciences, 6, 2001–2013, 2009
www.biogeosciences.net/6/2001/2009/

Biogeosciences

Towards global empirical upscaling of FLUXNET eddy covariance observations: validation of a model tree ensemble approach using a biosphere model

M. Jung[1], **M. Reichstein**[1], **and A. Bondeau**[2]

[1]Max Planck Institute for Biogeochemistry, Jena, Germany
[2]Potsdam Institute for Climate Impact Research (PIK), Potsdam, Germany

Received: 7 April 2009 – Published in Biogeosciences Discuss.: 26 May 2009
Revised: 9 September 2009 – Accepted: 14 September 2009 – Published: 6 October 2009

Abstract. Global, spatially and temporally explicit estimates of carbon and water fluxes derived from empirical up-scaling eddy covariance measurements would constitute a new and possibly powerful data stream to study the variability of the global terrestrial carbon and water cycle. This paper introduces and validates a machine learning approach dedicated to the upscaling of observations from the current global network of eddy covariance towers (FLUXNET). We present a new model TRee Induction ALgorithm (TRIAL) that performs hierarchical stratification of the data set into units where particular multiple regressions for a target variable hold. We propose an ensemble approach (Evolving tRees with RandOm gRowth, ERROR) where the base learning algorithm is perturbed in order to gain a diverse sequence of different model trees which evolves over time.

We evaluate the efficiency of the model tree ensemble (MTE) approach using an artificial data set derived from the Lund-Potsdam-Jena managed Land (LPJmL) biosphere model. We aim at reproducing global monthly gross primary production as simulated by LPJmL from 1998–2005 using only locations and months where high quality FLUXNET data exist for the training of the model trees. The model trees are trained with the LPJmL land cover and meteorological input data, climate data, and the fraction of absorbed photosynthetic active radiation simulated by LPJmL. Given that we know the "true result" in the form of global LPJmL simulations we can effectively study the performance of the MTE upscaling and associated problems of extrapolation capacity.

We show that MTE is able to explain 92% of the variability of the global LPJmL GPP simulations. The mean spatial pattern and the seasonal variability of GPP that constitute the largest sources of variance are very well reproduced (96% and 94% of variance explained respectively) while the monthly interannual anomalies which occupy much less variance are less well matched (41% of variance explained). We demonstrate the substantially improved accuracy of MTE over individual model trees in particular for the monthly anomalies and for situations of extrapolation. We estimate that roughly one fifth of the domain is subject to extrapolation while MTE is still able to reproduce 73% of the LPJmL GPP variability here.

This paper presents for the first time a benchmark for a global FLUXNET upscaling approach that will be employed in future studies. Although the real world FLUXNET upscaling is more complicated than for a noise free and reduced complexity biosphere model as presented here, our results show that an empirical upscaling from the current FLUXNET network with MTE is feasible and able to extract global patterns of carbon flux variability.

Correspondence to: M. Jung
(mjung@bgc-jena.mpg.de)

1 Introduction

The establishment of a global database of eddy covariance measurements of CO_2, H_2O and energy, the FLUXNET database (www.fluxdata.org), offers unprecedented opportunities to study the variability of the terrestrial carbon and water cycles. However, this compilation does not provide a complete picture; it has still the character of acupuncture and is heavily biased to regions in the mid-latitudes of the northern hemisphere. Therefore, one objective of the FLUXNET initiative is to derive coherent, spatially and temporally explicit maps of biosphere-atmosphere fluxes from the irregular distributed data points. Here, we call this process of generating spatial fields from point data upscaling.

Upscaling exercises of eddy covariance based carbon fluxes to large regions has been conducted for the US (Xiao

et al., 2008, Yang et al., 2007) and Europe (Jung et al., 2008; Papale and Valentini, 2003; Vetter et al., 2008), which are both characterized by a comparatively dense network of towers. The upscaling principle generally employs the training of a machine learning algorithm to predict carbon flux estimates based on measured meteorological data, remotely sensed vegetation properties, and vegetation type. The trained model can then be applied spatially using grids of the respective input data. Upscaling generally involves both, interpolation and extrapolation. We refer to interpolation when fluxes are predicted at locations whose environmental characteristics are captured by the training data set. Extrapolation occurs if fluxes are predicted for environments, which are not present in the training data set. It is important to note that it is not necessarily the geographical space which determines if inter- or extrapolation takes place but the environmental space. In our sense, an example for interpolation would be where ecosystems from the northern hemisphere may be used to predict carbon fluxes of structurally similar ecosystems in the southern hemisphere. Extrapolation may happen, if for example data from temperate coniferous forests are used to predict the response of temperate grasslands which are geographically nearby but structurally different. In practise, the distinction between inter- and extrapolation can be more fuzzy if it is not exactly known what determines structural similarity or if important characteristics are not known. For example, let us assume that data points from forests on shallow and acidic soils are used to estimate the behaviour of forests on deep and fertile soils which is different. If the soil information is present and if we know that soil is important then we would call it extrapolation, if not we would think of interpolation. This is an example of "hidden extrapolation", i.e. where predictions are made for conditions that are not sampled by the training data (here different soils) although the measured characteristics are captured by the training data (e.g. same climate and vegetation type).

A comparison of different diagnostic approaches to upscale gross primary production (GPP) from eddy covariance towers to Europe has suggested that (1) the method being used for upscaling has a strong effect on the final result, (2) that interannual anomaly patterns are comparatively poorly matching between the upscaled fields (Jung et al., 2008). No actual benchmarking has been carried out for upscaling algorithms and the issue of extrapolation has not been studied yet, which is crucial for large parts of the world with little or no flux towers such as large regions in tundra, boreal and tropical regions. In this paper we propose such a benchmarking by using a biosphere model as surrogate truth. The advantage of this approach is that we know the true result and that we do not confound uncertainties other than the method of upscaling and the distribution of the samples that are available for the training.

So called model trees are one example of a machine learning algorithm that can be trained to predict the fluxes and have been employed for the US to predict NEE (Xiao et al., 2008). Model trees are tree shaped structures that partition the data space into units where a specific model (usually a regression) is valid. This unsupervised stratification approach thus identifies "response units" where particular controlling factors and respective sensitivities govern the fluxes. Therefore, an advantage of model trees is that they partly resolve the problem of representativeness of the training data, by partitioning the data space into units of similar behaviour of the target variable with respect to the explanatory variables. A number of theoretical and empirical studies have shown that ensemble methods where several diverse models are constructed and jointly applied have substantial larger predictive capacity and have become common practise is many forecasting applications (Bates and Granger, 1969; Chandra et al., 2009; Hansen and Salamon, 1990; Kocev et al., 2009; Makridakis et al., 1982). However, we are not aware of any study that developed ensemble model trees or used ensemble methods for the upscaling of biogeochemical flux data.

We propose a new model tree algorithm with some innovations called TRIAL (Tree Induction ALgorithm), and introduce a new method to create model tree ensembles (MTEs), called ERROR (Evolving tRees with RandOm gRowth). Subsequently, we evaluate the efficiency of the proposed algorithms to upscale carbon fluxes from FLUXNET locations. A thorough testing is made possible by using simulations for gross primary production (GPP) of the Lund-Potsdam-Jena managed Land (LPJmL, Bondeau et al., 2007; Sitch et al., 2003) biosphere model as truth that is aimed to be reproduced by the model trees which are trained only at FLUXNET locations. We focus specifically on how well different components of the variability are reproduced such as the mean spatial pattern of GPP, the seasonal cycles, and the monthly anomaly patterns. We dedicate particular emphasis on investigating the extrapolation capacity of the proposed approach, and demonstrate the superiority of the ensemble method over single individual model trees.

2 Materials and methods

2.1 Tree Induction Algorithm (TRIAL)

Model trees (Fig. 1) have been developed from regression trees. Regression trees perform recursive stratification by minimizing the variance within data subsets and the model in the leaf nodes is a constant (the mean). Model trees contain nontrivial models in the leaf nodes, usually a multiple regression and their superiority over regression trees had been demonstrated (e.g. Vens and Blockeel, 2006). Algorithms that learn to generate a model tree are heuristic machine learning approaches, and data mining techniques for knowledge discovery and generally referred to as Top Down Induction of Model Trees (TDIMT). Several model tree induction heuristics have been proposed in the literature that

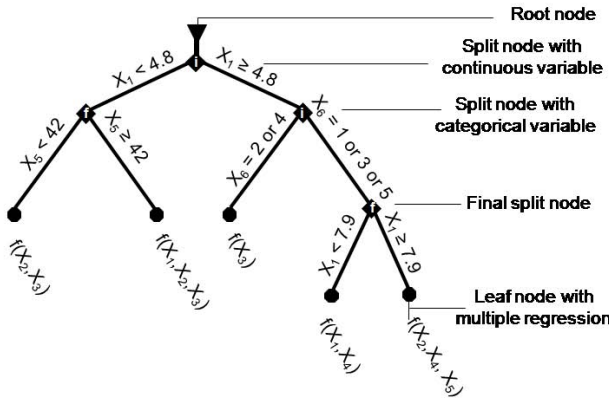

Fig. 1. Conceptual diagram of a model tree structure from TRIAL. X variables denote explanatory variables. Letters "f" and "i" within the split nodes indicate if the split node is a final split node (two leaf children only) or an interior split node (> two leaf children). The split along the categorical variable (X_6) is specific for TRIAL which allows moving several categories into left and right children (see supplementary material for details).

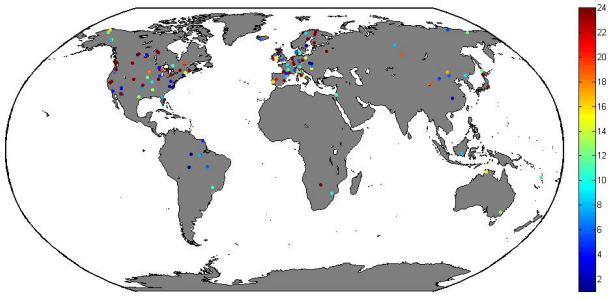

Fig. 2. Map of FLUXNET stations with the number of site-months that passed the quality control (n=3530, 178 sites). The colour gives the number of site months; the colour scale is truncated at 24 months.

share a common strategy (see Vens and Blockeel, 2006 for a review): First, an overly large tree is grown based on recursive partitioning, then the tree is pruned back. Differences among TDIMT algorithms are mainly related to (1) the cost function that is used to find the best split location for a variable X_i, (2) the search algorithm to find the best split along a split variable X_i, and (3) the model in the leaves. Since model tree induction methods are computationally expensive attention is given to keep computation time reasonable. The next sections provide a brief outline of the functioning of TRIAL. An illustration with pseudo-code is given as supplementary material, http://www.biogeosciences.net/6/2001/2009/bg-6-2001-2009-supplement.pdf.

2.1.1 General principle

In contrast to other model tree algorithms TRIAL allows to specify whether the explanatory variables X are (1) only split variables (X_{split}), (2) only regression variables (X_{reg}), or (3) both. The model in the leaf nodes are multiple linear regressions. The central cost function of TRIAL that is minimized is the Schwarz criterion (Schwarz, 1978), also known as Bayesian Information Criterion (BIC):

$$BIC = \log(MSE) \times n + \log(n) \times p \quad (1)$$

where MSE is the mean squared error based on 10-fold cross-validations, n the number of samples, and p the number of parameters (in our case including intercepts). The cross-validation operates in the leaves of the tree and thus provides an assessment of the robustness of the multiple linear regressions with selected regression variables (see below). The MSE of the tree is calculated by adding up the sum of squared errors (SSE) from the cross-validation of all leaves and then dividing by the total number of data points.

BIC contains a strong penalty for the complexity of the model which ensures parsimony. In combination with the MSE estimate from cross-validations, TRIAL is featured by strong overfitting avoidance. Although there is debate in the literature if the less penalizing Akaike's Information Criterion (AIC, Akaike, 1974) or BIC should be used for model selection (see Burnham and Anderson, 2004 and references therein) we favour simplicity of the model and chose BIC. The BIC criterion is used to stop the growth of a tree, to identify the node that should be split, as well as to select the predictor variables of the multiple regressions in the leaf nodes. Instead of a pruning phase after tree growth TRIAL employs pre-pruning by controlling which current leaf node is further partitioned to yield the largest information gain for the entire model tree and stops if further splitting results in an increase of BIC of the tree. Thus TRIAL is not based on truly recursive partitioning but evaluates each time which leaf node should be split. In practise this is facilitated by calculating BIC of the full new model tree for each possible leaf node that could be split and choosing the leaf where BIC of the new model tree is smallest (see also pseudo code in the supplementary material).

The key question of model tree algorithms is how the best split is found for a given node. Since an exhaustive test of all possible splits is often computationally impractical, a smart subset of possible splits is evaluated by computing multiple linear regressions for the left and right child for each tested split location. Subsequently, the location where the joint error of the model in the right and left child is minimal is chosen as the best split of a node. The search for the best spilt of a node is necessarily different for continuous and categorical variables and the next two sections describe the individual strategies in more detail. We refer the interested reader also to the pseudo code in the supplementary material, http://www.biogeosciences.net/6/2001/2009/bg-6-2001-2009-supplement.pdf.

2.1.2 Splits along continuous variables

Splits along continuous variables are determined by finding the split location l_i of each continuous explanatory variable X where the joint sum of squared errors of the left ($X_j < l_i$) and right child ($X_j >= l_i$) is minimal (cf. Karalic, 1992): min (SSE_{left} +SSE_{right}). Instead of evaluating each single possibility, the number of split locations being searched is restricted to a predefined number (default 100) to be tractable for larger datasets (cf. Potts and Sammut, 2005, Vogel et al., 2007). Once the best split variable and corresponding value is found a stepwise forward selection chooses the predictor variables of the multiple regressions based on BIC. Variable selection is a critical point to reduce the complexity of the model and to avoid unwanted effects of colinearity resulting in poorly constrained regression coefficients (Malerba et al., 2004). After the identification of the predictor variables a 10-fold cross-validation is used to estimate an unbiased estimate of the error. Because 10-fold cross-validation may be sensitive to the distribution of training and validation data points due to the random initialization, several repetitions (default: 5) of the 10-fold cross-validation are performed and the mean of the mean squared error over all cross-validations is stored.

2.1.3 Splits along categorical variables

In contrast to classic model and regression tree algorithms, TRIAL does not use so called binary splits for categorical variables where only one single category is separated from a group (Breiman et al., 1984). Our principle is based on iteratively joining two categories into a new aggregated one, which is repeated until only two categories are left which consist of several original categories. Starting from the initial variable with ncat categories, there are $0.5 \times ncat^2 - 0.5 \times ncat$ possibilities which two groups can be joined. For each of these possibilities SSE is computed if the two classes would be joined. Subsequently, the two groups where SSE of the joint multiple regression is minimal are aggregated into a new category, i.e. where two different groups can be best described using one regression. After each step of joining two categories, the new classification contains one category less. This procedure is repeated until only two major categories are left while each consists of several original categories. This approach is several orders of magnitude less computationally expensive than testing each possible way of splitting the categorical variable X_j into two subgroups.

2.1.4 Model tree ensembles (MTE): evolving tRees with RandOm gRowth (ERROR)

Ensemble methods (e.g. Breiman, 1996, Ho, 1998, Freund and Schapire, 1996, Breiman, 2001) where a set of different tree structures are built and jointly applied have been developed for decision and regression trees, and have been shown to outperform single trees (e.g. Dietterich, 2000) including the reduction of extrapolation errors (Loh et al., 2007). The effectiveness of ensembles relies on the accuracy and diversity of the individual members which constitutes a trade-off (Hansen and Salamon, 1990). Surprisingly, ensemble methods for model trees have not attracted attention so far.

The approach we propose follows the idea of Liu et al., 2008 which uses both deterministic splits (by finding the locally best split) and truly random splits without searching for the best split. Random splits are justified because the base search algorithm for the best split operates locally (at one node) only which has little meaning globally, i.e. for the performance of the entire tree (Geurts et al., 2006). Random splits allow also exploiting a substantially larger space of possible tree structures with positive effects on the diversity – accuracy trade-off of ensemble members. However, random splits may result in poor performance of the tree in particular if they occur at final split nodes. Liu et al., 2008 has shown that a combination of random and deterministic splits outcompetes classical ensemble methods based on resampling (e.g. bagging (Breiman, 1996)) and those based on random splits only (Geurts et al., 2006; Liu et al., 2005).

Common practice is to grow a large number of trees starting from the root. We use an evolutionary motivated approach where an existing tree is chosen, a branch is pruned and a new branch is grown with partly random and partly deterministic splits. The tree being selected for modification is partly random but the selection probability scales with the square root of the rank of its performance: for each tree a uniformly distributed random number is generated and multiplied by the square root of the rank of the BIC (best tree has rank 1, worst tree has rank "number of trees") and the tree associated with the minimum product is selected. This successive modification of existing trees allows that already "good" trees can be more easily improved further than growing independent trees from the root where the chances are small to achieve comparable good results again. If a large number of trees are evolved using this approach (e.g. 1000) there will be a sequence of trees that exhibit good performance and are finally independent of each other, i.e. they do not share any part of their structure. A certain fraction (e.g. 25 trees) of these "best-independent" trees is selected for the model tree ensemble.

The starting point of the "evolution" is the deterministic tree that is grown using TRIAL. Subsequently, the tree is pruned at a randomly chosen interior node and truly random splits are used to develop the tree further starting from this node until stopping criteria terminate the tree growth, most likely because inappropriate random splits were tried. Thus we can now use deterministic splits to continue the growth of the tree from the new leaf nodes until it stops again. While for interior split nodes the deterministic split is likely not the "best", the deterministic split is always the best split for final split nodes. Therefore, we impose that all final split nodes must be deterministic and only interior split nodes are allowed to be random.

2.2 Experimental design

The principle idea is to mimic the challenge of upscaling GPP from eddy-covariance sites to the globe by using a process model as "truth". This allows a thorough assessment of the efficiency of the proposed upscaling algorithms (TRIAL and TRIAL+ERROR) given the actual availability of relevant FLUXNET data at site level for training. We use simulations of GPP from the LPJmL biosphere model on monthly time scale from 1998–2005 and with a spatial resolution of 0.5°. We train the model trees to predict the simulated GPP at the respective locations and months where FLUXNET data of sufficient quality are available. We run three realizations to evaluate the relative performances of (1) the deterministic model tree using TRIAL, (2) the best model tree from the TRIAL+ERROR model tree ensemble consisting of 1000 trees, (3) a model tree ensemble consisting of the 25 "best independent" model trees from the 1000 model trees (MTE).

2.2.1 LPJmL simulations

LPJ is a dynamic global vegetation model (DGVM) and originates from the BIOME model family (Haxeltine and Prentice, 1996; Prentice et al., 1992). It simulates the distribution of plant functional types, and cycling of water and carbon on a quasi-daily time-step. LPJ has been used in numerous studies on responses and feedbacks of the biosphere in the Earth System (e.g. Brovkin et al., 2004; Lucht et al., 2002; Schaphoff et al., 2006; Sitch et al., 2005), and is probably the most extensively evaluated biosphere model to date. The version of LPJ used here has been adapted to account for a realistic treatment of croplands and grasslands using a crop functional type (CFT) approach (LPJmL, Bondeau et al., 2007).

The model runs at a spatial resolution of 0.5°, using global data sets of climate, soil type, and land use. The enhanced CRU TS2.1 climate database (CRU-PIK, Österle et al., 2003) provides the historical monthly climatology for the period 1901–2005. The anthropogenic land use information consists of annual cover fractions of 12 CFTs and one managed grassland, while all these are further distinguished according to rainfed and irrigated following (Fader et al., in review). The distribution of natural plant functional types (PFTs) is simulated by the model. A spinup run of 1000 years is first performed by recycling the first 30 years of the climate data in order to generate equilibrium of carbon pools and distribution of the natural plant functional types (PFTs).The model is then run dynamically for the period 1901–2005, responding to CO_2, climate, and land use change.

2.2.2 Explanatory variables for model tree training and upscaling

The explanatory variables being chosen for training and upscaling are those that are also available for the real FLUXNET upscaling endeavour, i.e. climatic/meteorological variables, biophysical state of the vegetation (FAPAR), and vegetation type (Table 1). We include variables that were used to drive LPJmL except for soil properties and atmospheric CO_2, and also include variables not used directly in LPJmL, since we do not always have relevant soil data and also do not know exactly which variables are needed for predicting carbon fluxes in the real FLUXNET upscaling initiative. Climate variables from the Climatic Research Unit (CRU, New et al., 2002) that provide mean annual characteristics and landuse data are only used as split variables for partitioning; they are not predictor variables that could appear in the regression equations. The FAPAR simulated by LPJmL is used as an input for the model tree training because remotely sensed FAPAR constitutes one of the most important information when upscaling carbon fluxes from eddy covariance sites (Jung et al., 2008; Sims et al., 2006). The 9 different natural, and 13 different crop functional types were aggregated into 9 classes (evergreen broadleaf trees, evergreen needleleaf trees, deciduous trees, C3 grass, C4 grass, C3 crop, C4 crop, C3 pasture, C4 pasture) which is compatible with the vegetation classification used in FLUXNET. Consistently with the real FLUXNET data availability, the dominant vegetation type within the gridcell and year was used as categorical explanatory variable for model tree training although LPJmL uses a fractional representation of vegetation types.

In total 21 explanatory variables are provided for model tree training of which 16 operate only as potential split variables. Please note that not all variables are necessarily included in the final model trees since some may not be selected.

2.2.3 Data selection for training at FLUXNET sites

In order to be consistent with the analogue FLUXNET upscaling exercise we extract explanatory variables and LPJmL simulations only exactly for the respective locations, years and months of FLUXNET data, which pass various quality controls. Uncertainty estimates of eddy covariance data is crucial since machine learning algorithms fit the data including their possible biases. We filter the eddy covariance data according to the degree of gap filling and using the data from the latest studies on systematic uncertainties of FLUXNET. This procedure yields a realistic number and distribution of data points that should also be used for the actual upscaling. There is clearly a trade-off between the strictness of the quality control and number of data points available for using in the upscaling. For the eddy covariance measurements and meteorological data (air temperature, global radiation, vapour pressure deficit, precipitation) measured at the sites we allow a maximum of 20% of gap filling within a calendar month. We estimated the u* associated uncertainty for all FLUXNET data using a bootstrapping approach as in Reichstein et al., 2005 and Papale et al., 2006. We reject all data where the 95% confidence interval of this uncertainty

Table 1. List of explanatory variables used for model tree training. Mean climatic variables from CRU are 1961-1990 means. Monthly meteorological data from CRU-PIK are from 1998–2005. Land cover is on annual time step.

Variable	Type	Source/reference	Is LPJmL driver?
Mean annual temperature	Split (continuous)	CRU	No
Mean Annual precipitation sum	Split (continuous)	CRU	No
Mean annual climatic water balance	Split (continuous)	Hargreaves and Samani 1985, Droogers and Allen 2002, CRU	No
Mean annual Potential evaporation	Split (continuous)	Hargreaves and Samani 1985, Droogers and Allen 2002, CRU	No
Mean annual sunshine hours	Split (continuous)	CRU	No
Mean annual number of wet days	Split (continuous)	CRU	No
Mean annual relative humidity	Split (continuous)	CRU	No
Mean monthly temperature	Split (continuous)	CRU	No
Mean monthly precipitation sum	Split (continuous)	CRU	No
Mean monthly climatic water balance	Split (continuous)	Hargreaves and Samani 1985, Droogers and Allen 2002, CRU	No
Mean monthly Potential evaporation	Split (continuous)	CRU	No
Mean monthly sunshine hours	Split (continuous)	CRU	No
Mean monthly number of wet days	Split (continuous)	CRU	No
Mean monthly relative humidity	Split (continuous)	CRU	No
Potential radiation	Split (continuous)	–	(Yes)
Temperature	Split & regression (continuous)	CRU-PIK	Yes
Cloudiness	Split & regression (continuous)	CRU-PIK	Yes
Precipitation	Split & regression (continuous)	CRU-PIK	Yes
fraction of absorbed photosynthetic active radiation (FAPAR)	Split & regression (continuous)	LPJmL	–
Potentially absorbed photosynthetic active radiation (Potential Radiation x FAPAR)	Split & regression (continuous)	(LPJmL)	–
Land Cover	Split (categorical)	PFTs: LPJmL	–
		CFTs: Fader et al. in review	Yes

for GPP exceeds 1 gC/m^2/day on average per month. Recently, Lasslop et al., in review applied an extended light response curve method for separating measured NEE into GPP and TER using primarily day time NEE data, which complements the standard FLUXNET GPP data from the Reichstein et al., 2005 algorithm that is based on estimating TER using night-time NEE data. We exclude (monthly) data points where the absolute difference of GPP from the two independent algorithms exceeds 1 gC/m^2/day. Moreover, we exclude entire sites if the absolute mean difference ("bias") between the two GPP estimates is larger than 120 gC/m^2/year. Such a systematic difference between the daytime and night-time based flux separation methods indicates possible problems with low turbulence and advection losses.

2.2.4 Model tree application to the domain

The model trees are applied to the spatial domain using grids of the explanatory variables. The computation is carried out separately for each vegetation type (i.e. assuming the entire grid would be covered by the same vegetation) and subsequently aggregated based on the fractional land use representation by calculating the weighted mean. Although the land use may change annually in LPJmL we use the mean fraction over the eight years (1998–2005) in order to be consistent with real FLUXNET upscaling.

The model tree ensemble is given as the median of the 25 independent values from each ensemble member. We calculate the uncertainty of the model tree ensemble using a robust

estimate of the standard deviation over the 25 ensemble trees, which is given as the median absolute deviation (MAD) × 1.4826. Multiplying the MAD with 1.4826 yields the standard deviation of a normal distribution.

2.2.5 Extrapolation detection

Detecting extrapolation in the upscaling is relevant for further analysis because the results may not be trustable, and is in any case interesting in terms of representativeness and future measurement network design of FLUXNET. Intuitively, extrapolation occurs when conditions are present that are not captured by the training dataset and this may be quantified by distance measures of environmental conditions. However, the shortcoming with this approach is that not all variables are equally important and that the importance of these variables (controlling factors) change in space and time, which is not known a priori. This problem can be circumvented by using our proposed ensemble method and we propose a new and simple way of detecting extrapolation which is also computationally inexpensive. Each model tree of the ensemble has learned a different way to predict the target variable from the same training dataset with roughly similar performance. We exploit this equifinality feature and argue that the different model tree estimates are similar if these conditions are known to them due to the training and that the estimates diverge if unknown conditions occur. We use a simple heuristics to flag extrapolation by testing if the uncertainty of the ensemble estimate is larger than the 99th percentile of the uncertainty of the ensemble from the training sequence. This binary flag can be further converted into an index of extrapolation that can be mapped spatially by computing the relative frequency of the extrapolation flag in the time domain for each pixel.

2.2.6 Statistical analysis

We report two standard statistical measures to assess the quality of the upscaling: the root mean squared error (RMSE), and the coefficient of determination. For the latter we do not compute the squared Pearson correlation coefficient but follow the definition that is also known as modelling efficiency to measure the deviation from the 1:1 line.

$$\mathrm{RMSE}=(\mathrm{SSE}/\mathrm{n})^{0.5} \quad (2)$$

$$R^2=1-\mathrm{SSE}/\mathrm{SS} \quad (3)$$

where SSE denotes the sum of squared errors, n the number of data points, and SS the sum of squares of the target variable (GPP from LPJmL). Please note that with this definition of R^2 negative values are possible if SSE exceeds SS. Whenever we refer to the "true" RMSE or R^2 we calculate the measures between the model tree based GPP and LPJmL GPP (truth) over the full modelling domain. We refer to training RMSE or R^2 when both measures are derived from the training data points only using 10 fold crossvalidations.

We decompose the spatio-temporal data of GPP into three components that help to understand uncertainties of different aspects: (1) between site or spatial variability, (2) mean seasonal variation, (3) between year variability of the monthly fluxes. We define (1) as the spatial field of the mean value computed over the time domain. Seasonal variation is given as the mean seasonal cycle minus its mean. Anomalies are calculated by subtracting the mean seasonal cycle. This decomposition allows quantifying global measures of the explained variance of the mean spatial pattern, mean seasonal variation, and monthly anomalies by the model tree results.

In addition to these global measures of model performance we extract the dominant modes of variability of the seasonal and interannual variability using principal component analysis (PCA). PCA is effective in reducing the dimensionality of a data set and to extract dominant patterns of variability. We use PCA to reduce the dimensionality in the time domain which yields spatial patterns of variability. The first principal component is used to map the dominant pattern of the seasonal and interannual variability. For the latter, interannual anomalies are computed by the difference of annual GPP and mean annual GPP for each pixel.

3 Results and discussion

3.1 Performance of individual trees and the model tree ensemble

Performance statistics of the deterministic tree, best tree, and ensemble of 5% of the best trees in Table 2 shows that (1) the model tree(s) are able to accurately reproduce ($R^2>0.96$) the LPJmL GPP in the training data set and (2) that the ERROR algorithm was able to generate a number of hybrid model trees that are superior over the deterministic tree in terms of the fit of the training samples. However, statistics of the true model tree performances (Table 3) which is computed over the full LPJml model domain after upscaling reveals that performance statistics from the training are not directly transferable to the actual upscaling product. For example, the best tree of the training exhibits poorer description of the global spatio-temporal variability of LPJmL than the deterministic tree, essentially because a substantial degree of extrapolation is necessary. The model tree ensemble yields the best performance for the full modelling domain.

The very high accuracy of 92% of explained variance of the global multi-year GPP of LPJmL by the model tree ensemble is surprising given that with the training data set less than 0.1% of the domain was sampled, and even in a geographically clumped way. The superiority of the model tree ensemble over individual trees is further illustrated in Fig. 3. The model tree ensemble shows always better performance than any of the individual trees overall, for the mean spatial pattern, seasonal variation, and anomalies of LPJmL GPP. While the mean spatial pattern and seasonal variation that

Table 2. Performance statistics of model trees from the training procedure (n=3530). MTE refers to the model tree ensemble (mean of 5% of the best hybrid trees (=25) with standard deviations given in brackets). Number of parameters includes intercepts of the regression models. Please note that all measures originate from five repetitions of a 10fold cross-validation within the leaf nodes.

	Deterministic tree	Best tree	MTE
R^2	0.963	0.97	0.966 [0.002]
RMSE	0.638	0.572	0.606 [0.02]
BIC	−2470	−3016	−2560 [205]
# strata	20	29	30.9 [4.9]
# parameters	86	114	119.4 [18]

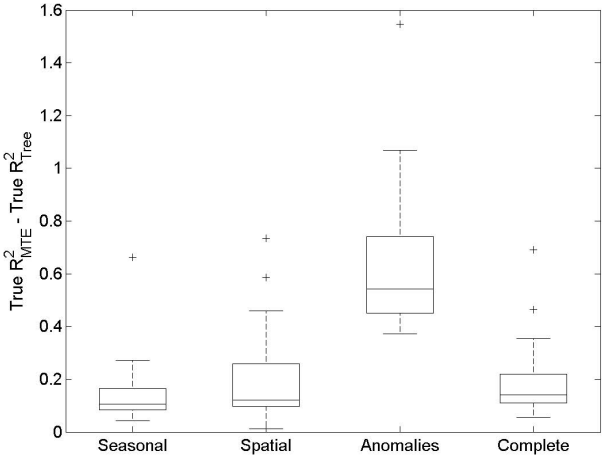

Fig. 3. Gain of true R^2 of the model tree ensemble over the 25 individual trees in the ensemble.

contribute large variance in the LPJmL GPP are already well captured by single trees, the GPP anomalies which constitute by far the lowest variance component is poorly reproduced by all individual trees but substantially improved by the model tree ensemble.

Having illustrated the improved efficiency of the ensembling method we next focus on a more detailed evaluation of the model tree ensemble upscaling. As indicated by the statistical measures above, Fig. 4 shows that the mean spatial pattern, and dominant mode of the seasonal variability are very well reproduced by the model tree ensemble and differences to LPJmL are hardly detectable visually. Although interannual variability is the component with least accuracy, its dominant pattern is consistently extracted by the MTE as seen by the first principal component (Fig. 4). This dominant pattern of interannual variability seems to largely represent the carbon cycle response to the El Nino Southern Oscillation (ENSO) climate phenomenon (Jones et al., 2001; Knorr et al., 2007; Qian et al., 2008). Also when aggregated to latitudinal bands the anomaly time series derived from the MTE

Table 3. Statistics of the true performance of the deterministic tree, best tree from the training, and model tree ensemble (MTE). Statistics were computed over the full modelling domain and against the "truth" (LPJmL). The decomposition into the components mean spatial pattern, seasonal variation, and monthly anomalies is described in Sect. 2.2.6. Please note the small fraction of variance of the anomalies, and the improved performance of the model tree ensemble in particular for the anomalies.

		Total	Spatial	Seasonal	Anomalies
Variance	Det Tree	9.79	3.99	5.49	0.32
[gC/m^2/day]	Best Tree	10.93	4.24	6.09	0.58
	MTE	9.32	3.98	5.1	0.24
	LPJmL	8.98	3.44	5.15	0.39
True R^2	Det Tree	0.86	0.93	0.87	0.09
	Best Tree	0.78	0.89	0.8	−0.4
	MTE	0.92	0.96	0.94	0.41
True RMSE	Det Tree	1.13	0.49	0.82	0.6
[gC/m^2/day]	Best Tree	1.4	0.6	1.02	0.74
	MTE	0.83	0.37	0.57	0.48

compares much more favourably to the original LPJmL dynamics than might be thought from Table 3, where individual pixels were compared (Fig. 5).

There are several reasons why the interannual variability is less well reproduced by MTE. Firstly, the signal is small in comparison to the spatial and seasonal variability as indicated by the variances in table 3, which also implies that comparatively little emphasis is given to that small fraction of variance during model tree training. Secondly, the controlling factors for the spatial GPP gradients and between year variability may differ (Reichstein et al., 2007), which might cause a conflict. Thirdly, we used mean annual fractions of vegetation types for MTE while the land use in LPJmL may show some variations over the year, for instance in consequence of fires. Fourthly, in many regions the GPP interannual variability as simulated by LPJmL is controlled by variations of soil moisture (e.g. Jung et al., 2007; Weber et al., 2009). Soil moisture is a storage term and causes memory effects of the system, which is not taken into account by MTE. Given that all these factors are not considered by MTE the results are still rather encouraging. However, if the interannual variability is of particular interest it may be possible to further improve the performance of MTE for interannual variability by training directly on the anomalies, i.e. using temperature, precipitation, and FAPAR anomalies etc. to predict carbon flux anomalies. This approach would solve the first two problems that the factors controlling interannual variability can differ from those determining the spatial and seasonal gradients, and that only little emphasis is normally given to reproducing the interannual variability due to its small contribution to the total variance. Further improvements may include the addition of variables with

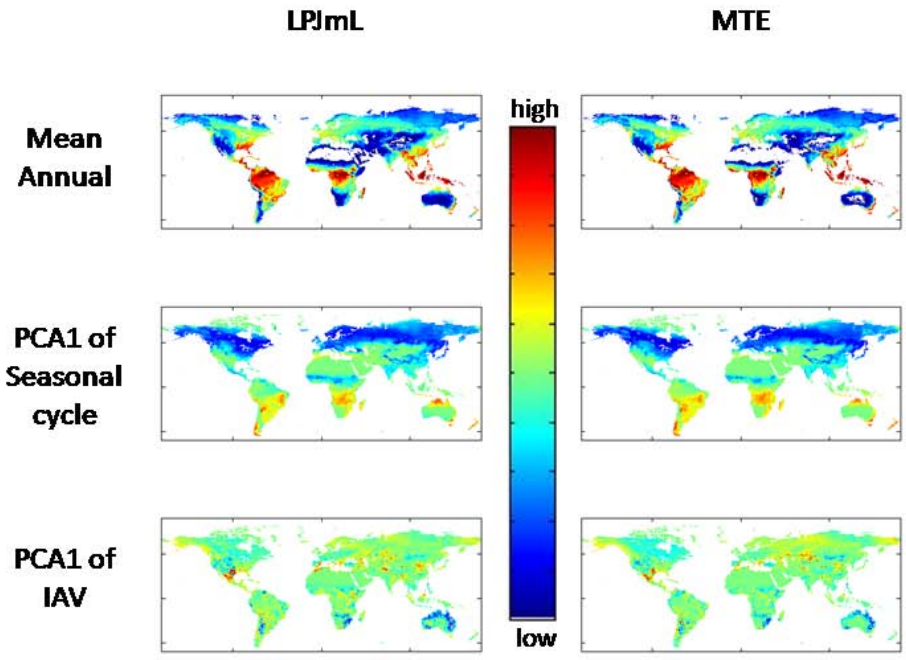

Fig. 4. Comparison of the mean annual, dominant seasonal, and dominant interannual patterns between the ensemble (left) and LPJmL (right). Dominant mode of the seasonal variation is given as the first principal component (PC) of the mean seasonal variation which explains 70% and 76% for LPJmL and MTE respectively. Dominant mode of interannual variability is given as the first PC of interannual anomalies which explains 27% and 29% of the variance for LPJmL and MTE respectively.

lag (e.g. precipitation anomaly of the previous one, two, or three months) or cumulated variables such as temperature sums or cumulative water balance indicators as proxy for soil moisture as additional explanatory variables. Such additions would enable to describe memory effects, i.e. effects of past conditions on the current fluxes.

3.2 Uncertainty estimates and extrapolation capacity of the model tree ensemble

Providing realistic uncertainty estimates of the upscaling products is essential for their scientific use. The uncertainty of the upscaling from ensemble methods described by a robust estimate of the standard deviation is weakly correlated with the true absolute error between the MTE and LPJmL (r=0.37 (Pearson) over the full spatio-temporal domain) indicating that between-tree variability does not necessarily imply a large prediction error also. As described in Sect. 2.2.5 we can flag situations when extrapolation is likely which are characterized by a prediction variability among individual trees that goes beyond those present for the training data. We estimate that about one fifth (21.5%) of the pixel-months are subject to extrapolation. An index of extrapolation defined as the fraction of months per pixel flagged and extrapolated is mapped in Fig. 6. On the one hand it is evident that in particular tropical areas are subject to extrapolation, but that even the few flux towers effectively constrain the MTEs

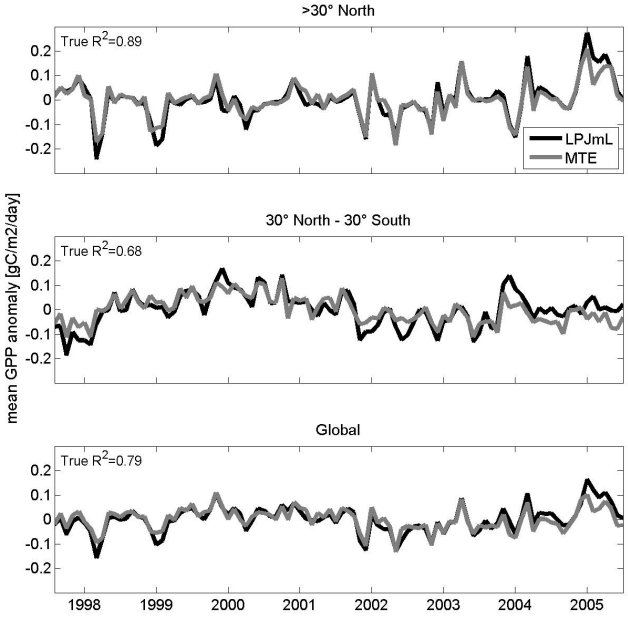

Fig. 5. Comparison of the monthly anomalies between LPJmL and MTE for latitudinal bands and global.

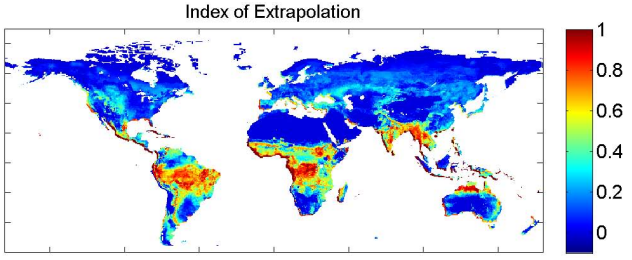

Fig. 6. Index of extrapolation derived from the model tree ensemble. See text for details.

for a considerable part (e.g. northern Amazon forest areas, parts of Indonesia). Moreover, the lack a towers over boreal Siberia seems to introduce less extrapolation problems than expected, since environmental conditions seem to be sampled well by Canadian and European flux towers. Here the strength of the model tree approach is illustrated because data from ecosystems that function similarly are identified via the stratification and can be used to predict similar ecosystems that are geographically far away. However, the good performance in Siberia will be also related to simplifications made in LPJmL to some extent, such as one general parameter set for boreal forests.

By using this extrapolation flag we can demonstrate the substantially improved extrapolation capacity of the ensemble relative to individual trees by computing the true performance separately for non-extrapolation and extrapolation conditions (Fig. 7). While the individual trees show high and only a small spread of performance for interpolation, the individual trees give poor results and diverge in performance for extrapolation. However, the ensemble as a combined estimate over the individual trees gives substantially improved results for conditions of extrapolation (True R^2 =0.74) and also a small gain for non-extrapolation situations (True R^2 =0.95). Thus, even when the estimates of different trees diverge the median value appears to be a robust approximation of the true value in many cases which underlines once more the advantages of ensemble methods.

The (robust) standard deviation of the predictions between the trees of MTE can be interpreted as a measure of prediction uncertainty. Taking advantage of the fact that we know the true values we can evaluate if this measure of uncertainty is sensible. From a theoretical statistical point of view, for example 95% of the true observations should lie within +-2 standard deviations of the mean MTE estimate assuming normal distribution of the error. Hence, Fig. 8 summarizes the percentage of true observations being within a certain multitude of standard deviations. The true value is within one or two standard deviations in 73% or 90% of the cases respectively. Given that the error distribution is not necessarily Gaussian this result indicates that the estimation of uncertainty is reasonable. Interestingly, under extrapolation conditions always a larger percentage of true observations is within

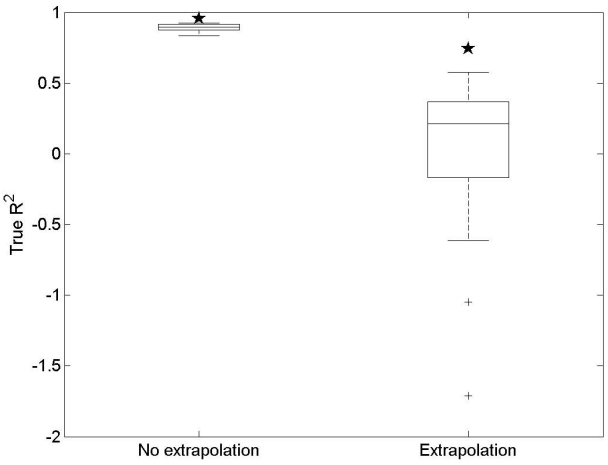

Fig. 7. True R^2 for interpolation and extrapolation conditions of the individual trees of the ensemble (box plot) and the model tree ensemble (star).

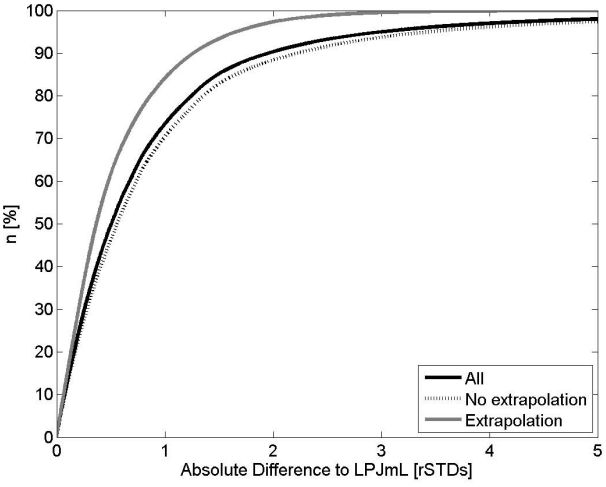

Fig. 8. Cumulative distribution of the number of standard deviations that are needed to capture the true value, stratified for extrapolation and no extrapolation conditions.

the estimated uncertainty range than under non-extrapolation conditions. This illustrates that the uncertainty estimate tends to be conservative and is also valid when extrapolating.

3.3 Remarks regarding FLUXNET upscaling

In this section we discuss briefly the meaning of our synthetic test case for real FLUXNET upscaling projects and propose additional steps that can be taken to further study and improve FLUXNET upscaling. The primary objective of this paper was to introduce the method of model tree ensembles and an evaluation of its efficiency to derive spatial and temporal fields from highly clumped and irregularly spaced data like FLUXNET. The presented test case using a biosphere

model as surrogate truth is a necessary first step to gain confidence in the technique. If MTE would have failed to adequately reproduce the LPJmL simulations from the flux tower locations, it would not be worth applying it using real world flux data. We suggest that our approach of testing the method of empirical upscaling flux tower data to continents and the globe should become a required standard.

The fact that MTE could reproduce the global LPJmL simulations very well does not proof that the real FLUXNET upscaling products using MTE will generate global carbon flux fields of comparable accuracy. Most importantly, the biosphere is more complex than LPJmL, which uses a relatively small number of plant functional types associated with constant sets of parameters to discretize global ecosystems, uses necessary simplifications of physical and physiological processes, and lacks other processes that can be important such as nutrient cycles. The artificial experiment using the biosphere model also represents a "perfect world" without any uncertainties in the data while data uncertainties are present (and an issue) in both, the flux tower data, and driver data such as grids of meteorological fields and satellite based estimates of fAPAR. The current study is also simpler in the sense that we use training data at 0.5° resolution to predict at 0.5° resolution globally. In reality, the training data (meteorology and fluxes) are measured at the towers plus time series of satellite fAPAR products of 1 or 2 km resolution. On top of the uncertainty of all these measurements there is additional uncertainty originating from a mismatch between the footprints of the individual instruments at the tower and of the satellite. This footprint mismatch introduces additional noise.

The role of the data uncertainties could be assessed for example by adding noise and bias to the (simulated) training data that are comparable to those inferred from studying the uncertainties of eddy covariance data (Lasslop et al., 2008; Richardson et al., 2006). This approach would be effective to evaluate how well the machine learning tools are capable of extracting general relationships from the noisy data or tend to overfit. If successful, machine learning tools could be used to assess the information content and signal to noise ratios of real world data. Clearly, additional confidence of FLUXNET derived upscaling products is required by corroboration against independent data.

4 Summary and conclusion

We have presented a new model tree ensemble machine learning algorithm and provided empirical evidence for its efficiency. We performed an upscaling of simulated GPP from LPJmL from the highly clumped distribution of FLUXNET sites to the globe and evaluated this product against the actual LPJmL simulations which here constitutes the truth. The model tree ensemble result explains overall 92% of the variance of the global LPJmL GPP simulations, 96% of the mean spatial pattern, 94% of the seasonal variability, and 41% of the monthly anomalies. The uncertainty estimates of the model tree ensemble, given as the robust standard deviation of the individual tree estimates, was confirmed to be a useful indicator of the true uncertainty. The true value was within one standard deviation for 73% of the cases. We developed an indicator for extrapolation based on the spread of the model tree estimates which yields plausible results showing that overall about one fifth of the global spatio-temporal domain was subject to extrapolation, primarily large parts of the tropics, which, however, does not necessarily imply poor performance of the ensemble estimate. We demonstrate that the ensemble method is particularly powerful in enhancing extrapolation capacity yielding a true R^2 of 73% when extrapolating (95% when interpolating).

This study constitutes a benchmark for the method of upscaling carbon and water fluxes from FLUXNET sites to the globe which is enabled by using a biosphere model as surrogate truth. We can conclude that the proposed method is highly efficient to perform this upscaling and is able to generate good and substantially better results than single trees also in situations of extrapolation. The retrieved performance statistics can certainly not be directly transferred to the real FLUXNET upscaling exercise where a more complex world than LPJmL, noise of explanatory variables, and possible systematic biases in the flux measurements must be expected and taken into account. Nevertheless, we have now improved confidence that future FLUXNET upscaling products using our method will be a new and useful information stream derived from observations that will help to better understand the variability of the global terrestrial carbon cycle.

The service charges for this open access publication have been covered by the Max Planck Society.

Edited by: E. Falge

References

Akaike, H.: A new look at the statistical model identification, IEEE Transactions on Automatic Control, 19(6), 716–723, 1974.

Bates, J. M. and Granger, C. W. J.: The combination of forecasts, Operations Research Quarterly, 20, 451–468, 1969.

Bondeau, A., Smith, P. C., Zaehle, S., et al.: Modelling the role of agriculture for the 20th century global terrestrial carbon balance, Glob.l Change Biol., 13(3), 679–706, 2007.

Breiman, L.: Bagging predictors, Mach. Learn., 24(2), 123–140, 1996.

Breiman, L.: Random forests, Mach. Learn., 45(1), 5–32, 2001.

Breiman, L., Friedman, J., Olshen, R., and Stone J.: Classification and Regression Tree, Wadsworth and Brooks, 1984.

Brovkin, V., Sitch, S., von Bloh, W., Claussen, M., Bauer, E., and Cramer, W.: Role of land cover changes for atmospheric CO2

increase and climate change during the last 150 years, Glob. Change Biol., 10(8), 1253–1266, 2004.

Burnham, K. P. and Anderson, D. R.: Multimodel inference – understanding AIC and BIC in model selection, Sociological Methods and Research, 33(2), 261–304, 2004.

Chandra, D. K., Ravi, V., and Bose I.: Failure prediction of dotcom companies using hybrid intelligent techniques, Expert Syst. Appl., 36, 4830–4837, 2009.

Dietterich, T. G.: An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization, Mach. Learn., 40(2), 139–157, 2000.

Fader, M., Rost, S., and Müller, C.: Virtual water content of temperate cereals and maize: Present and potential future pattern, J. Hydrol., in review.

Freund, Y. and Schapire,R. E.: Experiments with a new boosting algorithm, Proceedings of the 13th International Conference on Machine Learning, 148–156, 1996.

Geurts, P., Ernst, D., and Wehenkel, L.: Extremely randomized trees, Mach. Learn., 63(1), 3–42, 2006.

Hansen, L. and Salamon, P.: Neural network ensembles, IEEE Trans. Pattern Anal. Mach. Intell., 12, 993–1001, 1990.

Haxeltine, A. and Prentice, I. C.: BIOME3: an equilibrium terrestrial biosphere model based on ecophysiological constraints, resource availability and competition among plant functional types, Global Biogeochem. Cy., 10, 693–710, 1996.

Ho, T. K.: The random subspace method for constructing decision forests, Ieee Transactions on Pattern Analysis and Machine Intelligence, 20(8), 832–844, 1998.

Jones, C., Collins, M., Cox, P., and Spall, S. A.: The Carbon Cycle Response to ENSO: A Coupled Climate-Carbon Cycle Model Study, J. Climate, 14, 4113–4129, 2001.

Jung, M., Verstraete, M., Gobron, N., Reichstein, M., Papale, D., Bondeau, A., Robustelli, M., and Pinty, B.: Diagnostic assessment of European gross primary production, Glob. Change Biol., 14(10), 2349–2364, 2008.

Jung, M., Vetter, M., Herold, M., et al.: Uncertainties of modeling gross primary productivity over Europe: A systematic study on the effects of using different drivers and terrestrial biosphere models, Global Biogeochem. Cy., 21, GB4021, doi:10.1029/2006GB002915, 2007.

Karalic, A.: Employing linear regression in regression tree leaves, Proceedings of the 10th European Conference on Artificial Intelligence, 440–441, 1992.

Knorr, W., Gobron, N., Scholze, M., Kaminski, T., Schnur, R., and Pinty, B.: Impact of terrestrial biosphere carbon exchanges on the anomalous CO_2 increase in 2002–2003, Geophys. Res. Lett., 34, L09703, doi:10.1029/2006GL029019, 2007.

Kocev, D., Dzeroski, S., White, M. D., Newell, G., and Griffioen, P.: Using single- and multi-target regression trees and ensembles to model a compound index of vegetation condition, Ecol. Modell., 220, 1159–1168, 2009.

Lasslop, G., Reichstein, M., Kattge, J., and Papale, D.: Influence of observation errors in eddy flux data on inverse model paramter estimation, Biogeosciences, 5, 1311–1324, 2008, http://www.biogeosciences.net/5/1311/2008/.

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, Glob. Change Biol., doi:10.1111/j.1365-2486.2009.02041.x, in press, 2009.

Liu, F. T., Ting, K. M., and Fan, W.: Maximizing Tree Diversity by Building Complete-Random Decision Trees, Advances in Knowledge Discovery and Data Mining, 9th Pacific-Asia Conference, PAKDD 2005, 2005.

Liu, F. T., Ting, K. M., Yu, Y., and Zhou, Z. H.: Spectrum of variable-random trees, J. Artif. Intell. Res., 32, 355–384, 2008.

Loh, W., Chen, C. W., and Zheng, W.: Extrapolation errors in linear model trees, ACM Trans. Knowl. Discov. Data, 1(2), 6, ISSN:1556-4681, 2007.

Lucht, W., Prentice, I. C., Myneni, R. B., Sitch, S., Friedlingstein, P., Cramer, W., Bousquet, P., Buermann, W., and Smith, B.: Climatic control of the high-latitude vegetation greening trend and Pinatubo effect, Science, 296(5573), 1687–1689, 2002.

Makridakis, S., Anderson, A., Carbone, R., Fildes, R., Hibdon, M., and Lewandowski, R.: The accuracy of extrapolation (time series) methods: Results of a forecasting competition, J. Forecast., 1, 111–153, 1982.

Malerba, D., Esposito, F., Ceci, M., and Appice, A.: Top-Down Induction of Model Trees with Regression and Splitting Nodes, IEEE Trans. Pattern Anal. Mach. Intell., 26(5), 612–625, 2004.

New, M., Lister,D. Hulme, M. and Makin I.: A high-resolution data set of surface climate over global land areas, Climate Res., 21, 1–25, 2002.

Österle, H., Gerstengarbe, F.-W., and Werner, P. C.: Homogenisierung und Aktualisierung des Klimadatensdatzes der Climate Research Unit of East Anglia, Norwich, Terra Nostra, 6, 326–329, 2003.

Papale, D. and Valentini, A.: A new assessment of European forests carbon exchanges by eddy fluxes and artificial neural network spatialization, Glob. Change Biol., 9(4), 525–535, 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, 2006, http://www.biogeosciences.net/3/571/2006/.

Potts, D. and Sammut, C.: Incremental learning of linear model trees, Mach. Learn., 61(1–3), 5–48, 2005.

Prentice, I. C., Cramer, W., Harrison, S. P., Leemans, R., Monserud, R. A., and Solomon, A. M.: A Global Biome Model Based on Plant Physiology and Dominance, Soil Properties and Climate, J. Biogeography, 19(2), 117–134, 1992.

Qian, H., Joseph, R., and Zeng, N.: Response of the terrestrial carbon cycle to the El Nino-Southern Oscillation, Tellus B, 60(4), 537–550, 2008.

Reichstein, M., Papale, D., Valentini, R., et al.: Determinants of terrestrial ecosystem carbon balance inferred from European eddy covariance flux sites, Geophys. Res. Lett., 34, L01402, doi:10.1029/2006GL027880, 2007.

Reichstein, M., Falge, E., Baldocchi, D., et al.: On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm, Glob. Change Biol., 11(9), 1424–1439, 2005.

Richardson, A. D., Hollinger, D. Y., Burba, G. G., et al.: A multisite analysis of random error in tower-based measurements of carbon and energy fluxes, Agr. Forest. Meteorol., 136(1–2), 1–

18, 2006.

Schaphoff, S., Lucht, W., Gerten, D., Sitch, S., Cramer, W., and Prentice, I. C.: Terrestrial biosphere carbon storage under alternative climate projections, Climatic Change, 74(1–3), 97–122, 2006.

Schwarz, G.: Estimating the dimension of a model, Ann. Stat., 6(2), 461–464, 1978.

Sims, D. A., Rahman, A. F., Cordova, V. D., et al.: On the use of MODIS EVI to assess gross primary productivity of North American ecosystems, J. Geophys. Res.-Biogeosci., 111(G4), G04015, doi:10.1029/2006JG000162, 2006.

Sitch, S., Brovkin, V., von Bloh, W., van Vuuren, D., Assessment, B., and Ganopolski, A.: Impacts of future land cover changes on atmospheric CO_2 and climate, Global Biogeochem. Cy., 19(2), GB2013, doi:10.1029/2004GB002311, 2005.

Sitch, S., Smith, B., Prentice, I. C., et al.: Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model, Glob. Change Biol., 9(2), 161–185, 2003.

Vens, C. and Blockeel, H.: A simple regression based heuristic for learning model trees, Intelligent Data Analysis, 10, 215–236, 2006.

Vetter, M., Churkina, G., Jung, M., Reichstein, M., Zaehle, S., Bondeau, A., Chen, Y., Ciais, P., Feser, F., Freibauer, A., Geyer, R., Jones, C., Papale, D., Tenhunen, J., Tomelleri, E., Trusilova, K., Viovy, N., and Heimann, M.: Analyzing the causes and spatial pattern of the European 2003 carbon flux anomaly using seven models, Biogeosciences, 5, 561–583, 2008, http://www.biogeosciences.net/5/561/2008/.

Vogel, D. S., Asparouhov, O., and Scheffer, T.: Scalable look-ahead linear regression trees, in: Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining, edited by: ACM, San Jose, California, USA, 2007.

Weber, U., Jung, M., Reichstein, M., Beer, C., Braakhekke, M. C., Lehsten, V., Ghent, D., Kaduk, J., Viovy, N., Ciais, P., Gobron, N., and Rödenbeck, C.: The interannual variability of Africa's ecosystem productivity: a multi-model analysis, Biogeosciences, 6, 285–295, 2009, http://www.biogeosciences.net/6/285/2009/.

Xiao, J. F., Zhuang, Q. L., Baldocchi, D. D., et al.: Estimation of net ecosystem carbon exchange for the conterminous United States by combining MODIS and AmeriFlux data, Agr. Forest Meteorol., 148(11), 1827–1847, 2008.

Yang, L., Ichii ,K., White, M. A., Hashimoto, H., Michaelis, A., Votava, P., Zhu, A., Huete, A., Running, S., and Nemani, 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, 2007.