JGR Biogeosciences

RESEARCH ARTICLE
10.1029/2020JG005814

Upscaling Net Ecosystem Exchange Over Heterogeneous Landscapes With Machine Learning

O. Reitz[1], A. Graf[2], M. Schmidt[2], G. Ketzler[1], and M. Leuchner[1]

[1]Department of Geography, RWTH Aachen University, Aachen, Germany, [2]Agrosphere Institute, Forschungszentrum Jülich, Jülich, Germany

Key Points:

- CO_2 flux upscaling with Random Forest can be improved with a backward feature elimination and strict quality control of flux data
- Vegetated parts of the Rur catchment were predicted to be a CO_2 sink on average, with the highest uptake in late spring and early summer
- The Enhanced Vegetation Index and potential evapotranspiration are useful predictors for the regionalization of CO_2 flux measurements

Correspondence to:
O. Reitz,
oliver.reitz@geo.rwth-aachen.de

Citation:
Reitz, O., Graf, A., Schmidt, M., Ketzler, G., & Leuchner, M. (2021). Upscaling net ecosystem exchange over heterogeneous landscapes with machine learning. *Journal of Geophysical Research: Biogeosciences*, *126*, e2020JG005814. https://doi.org/10.1029/2020JG005814

Received 30 APR 2020
Accepted 29 NOV 2020

Abstract This paper discusses different feature selection methods and CO_2 flux data sets with a varying quality-quantity balance for the application of a Random Forest model to predict daily CO_2 fluxes at 250 m spatial resolution for the Rur catchment area in western Germany between 2010 and 2018. Measurements from eddy covariance stations of different ecosystem types, remotely sensed vegetation data from MODIS, and COSMO-REA6 reanalysis data were used to train the model and predictions were validated by a spatial and temporal validation scheme. Results show the capabilities of a backwards feature elimination to remove irrelevant variables and an importance of high-quality-low-quantity flux data set to improve predictions. However, results also show that spatial prediction is more difficult than temporal prediction by reflecting the mean value accurately though underestimating the variance of CO_2 fluxes. Vegetated parts of the catchment acted as a CO_2 sink during the investigation period, net capturing about 237 g C m^{-2} y^{-1}. Croplands, coniferous forests, deciduous forests and grasslands were all sinks on average. The highest uptake was predicted to occur in late spring and early summer, while the catchment was a CO_2 source in fall and winter. In conclusion, the Random Forest model predicted a narrower distribution of CO_2 fluxes, though our methodological improvements look promising in order to achieve high-resolution net ecosystem exchange data sets at the regional scale.

Plain Language Summary Whether ecosystems absorb or emit CO_2 plays a major role in the global carbon cycle and impacts climate change. This exchange is already measured at scattered stations, but creating spatially resolved data sets remains a challenge. In this paper, we used satellite images of vegetation and meteorological data to predict the CO_2 exchange of the Rur catchment area near the German-Dutch-Belgian border for every day from 2010 to 2018. In order to assess the prediction quality, we compared actual measurements from several stations within the catchment with the predictions at the locations of these stations. Results show that our method could increase prediction quality compared to previous process-based models, though the error remains rather high. Vegetated parts of the catchment including coniferous forests, deciduous forests, grasslands, and croplands were all CO_2 sinks on average. In late spring and early summer, they were the strongest sink, but in fall and winter a CO_2 source.

1. Introduction

Land use changes are important drivers of anthropogenic climate change. For example, deforestation or afforestation can highly affect the carbon uptake and storage capacities of an ecosystem (Schimel et al., 2001). Net ecosystem exchange (NEE), the difference between carbon dioxide (CO_2) uptake through photosynthesis and respiration within an ecosystem (Luyssaert et al., 2007), represents a major feature of the global carbon cycle and, thus, helps to assess ecosystem services and the impact of land use changes on them (negative NEE = CO_2 uptake, positive NEE = CO_2 emission) (Abdalla et al., 2013; Schmitt et al., 2010; Xu et al., 2017).

The eddy covariance (EC) technique is the most direct way of measuring CO_2 fluxes; however, it provides point measurements from a sparse network of stations each representing a footprint with an along-wind extent typically less than 800 m (Chen et al., 2009). Top-down approaches for spatial NEE assessment include global atmospheric inversion models from satellites such as GOSAT and OCO-2 (H. Wang et al., 2019), which are especially useful for areas with limited or no EC coverage (Kondo et al., 2015) but are restricted to a coarse spatial resolution. Thus, bottom-up approaches scaling up EC measurements are expedient to quantify CO_2 fluxes for larger areas (Denman et al., 2007; Xiao et al., 2012), though they are also challenging due to the high spatiotemporal variability of those fluxes (Borchard et al., 2015; Kondo et al., 2017).

Process-based biogeochemical models have been widely applied for this purpose (e.g., Post et al., 2018; Xiao et al., 2011), but subject to assumptions and model parametrizations. Data-driven machine learning techniques such as Random Forest (RF) are another promising approach to predict NEE as they can grasp even highly nonlinear relationships to explanatory variables as is usual in environmental data (Cutler et al., 2007). Previous attempts using statistical modeling include nonspatial predictions of NEE at the EC tower scale (Dou et al., 2018; Safa et al., 2019; Zhou et al., 2019). Other attempts aimed at upscaling of carbon fluxes to the continental or national scale (Papale et al., 2015; Sun et al., 2011; Xiao et al., 2008) or the globe, most notably the FLUXCOM approach (Bodesheim et al., 2018; Jung et al., 2011, 2020). Upscaling to the regional scale at high spatial resolution has rarely been done although NEE estimates of heterogeneous regional and local ecosystems are of high value for assessing ecosystem services in spatial planning (Tammi et al., 2017). Furthermore, products at a finer spatial resolution are less prone to contain mixed pixels with contamination of the main land use class by, for example, roads, settlements, or tree rows. Zhang et al. (2011) developed a regression model for the U.S. Great Plains based on EC towers representing grassland only. Post et al. (2018) already upscaled NEE to the study area of this analysis, though with a process-based model.

Spatial cross validation, that is, excluding whole locations from model training and testing the model on them, is crucial for a realistic assessment of the reliability of spatial predictions beyond the locations of training data. A substantial performance decrease is to be expected in comparison to a random split of data points into training and test sets, which hence overrates model performance due to spatial autocorrelation (Meyer et al., 2018; Roberts et al., 2016). Tramontana et al. (2016) conducted a profound cross validation analysis for spatial predictions of various carbon and energy fluxes with the conclusion that NEE is especially difficult to predict. Feature selection of explanatory variables, on the other hand, can considerably improve data-driven model performance as it reduces overfitting and removes irrelevant or redundant variables (Hall & Smith, 1999). Meyer (2018) proposed a sequential feature selection algorithm based on spatial cross validation to remove spatially autocorrelated predictors. In contrast to this, conventional feature selection as implemented in the caret package (Classification And REgression Training, Kuhn, 2020) is based on internal cross validations within the training data, and hence fails to improve model performance when testing on locations not used for model training (Meyer et al., 2019). Genetic algorithms like the Guided Hybrid Genetic Algorithm (GHGH, Jung & Zscheischler, 2013) are useful for very large numbers of features (>100), though generally they do not rely on spatial cross validation. Quality of EC data is another issue for upscaling attempts, especially when aggregating half-hourly to daily fluxes. While data quality improves when excluding low-confidence values based on quality control, too small data set. also limit the learning capacities of machine learning algorithms (Ließ et al., 2012). A common practice is to indicate daily data as missing if more than 20% of half-hourly values are missing or of low quality (Tramontana et al., 2016; Yuan et al., 2010). However, to our knowledge a sensitivity analysis to different percentages has not been done before.

In conclusion, NEE upscaling with data-driven methods at high spatiotemporal resolutions and incorporating different land uses remains a major task to be handled in order to approach the goal of flux information "everywhere, all of the time" (Baldocchi, 2014). Thus, the objectives of this paper are i) to perform upscaling of daily NEE over heterogeneous landscapes of the Rur catchment in western Germany for the years 2010–2018 with a RF model incorporating EC measurements, remote sensing and reanalysis data and ii) to assess the impact of EC data quality and feature selection on the model performance.

2. Data and Methods

2.1. Study Area

The Eifel/Lower Rhine Valley Observatory covers the Rur catchment located near the German-Belgian-Dutch border and is one of four Terrestrial Environmental Observatories (TERENO) in Germany (Zacharias et al., 2011). These areas were selected for the TERENO network because they are representative of typical landscapes found in Central Europe (Bogena et al., 2016). The catchment covers an area of 2,354 km^2 and can be divided into a northern lowland part with intensive agriculture and a relatively high proportion of built-up areas and a southern low mountain part where pastures and forests prevail, as shown in Figure 1. Based on a simplified land cover classification by Lussem and Herbrecht (2019), the catchment

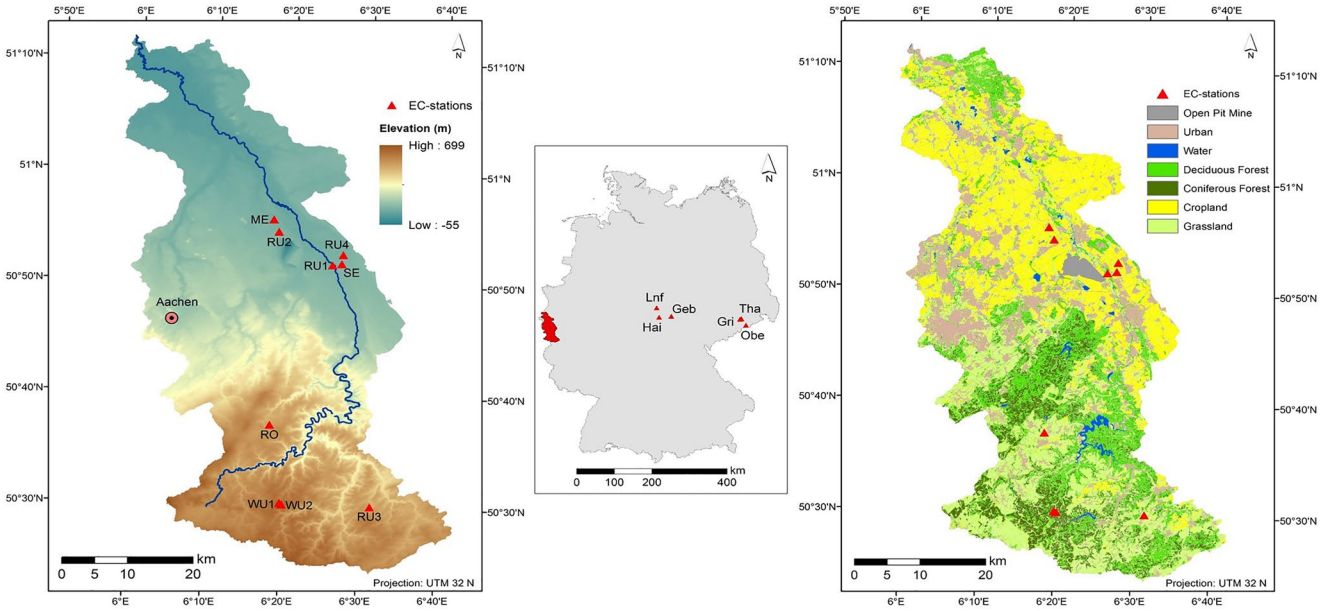

Figure 1. Elevation from SRTM data (Jarvis et al., 2008) and eddy covariance (EC) stations used for training within the Rur catchment area (left), the location of the study area and the FLUXNET stations within Germany (middle) and simplified land cover classes after Lussem and Herbrecht for the Rur catchment (2019) (right).

consists of 27.6% grassland, 25.7% cropland, 17.7% deciduous forest, 8.5% coniferous forest, and 20.4% other land cover types including urban areas, open cast mines and water bodies. Mean annual temperatures range from about 7.5–10.2°C, increasing from south to north. Mean annual precipitation decreases from 1,200 mm in the southern low mountain parts to 700 mm in the north (Baatz et al., 2014).

2.2. Eddy Covariance Data

CO_2 flux measurements from nine EC stations covering different land cover types and elevations within the study area have been used for model training and prediction (see Table 1 for details and abbreviations). The nine stations are all part of the TERENO network (Zacharias et al., 2011). Measurements from these stations were processed with the TK3 software (Mauder & Foken, 2011); 20 Hz frequency data were hereby processed to 30 min fluxes and corrected for storage terms to obtain NEE values. All processing and quality control schemes were carried out according to the standardized strategy presented by Mauder et al. (2013), which also includes a test on developed turbulence after Foken and Wichura (1996). Detailed measurement and processing descriptions can be found in the references listed in Table 1, a short description of each site is given here.

RO is an extensively managed grassland site, which is cut several times per year and mostly consists of ryegrass and smooth meadow grass (*Lolium perenne*, *Poa pratensis*). The EC tower was placed in the middle of two neighboring pastures with slightly different management regimes (Borchard et al., 2015; Korres et al., 2010). RU3 is a grassland site with similar characteristics (Post et al., 2015), while RU1 is a grassland site at significantly lower elevation (Lussem & Herbrecht, 2019). SE, ME, RU2, and RU4 are cropland sites with rotating crops, mostly sugar beet, winter wheat and winter barley (Eder et al., 2015; Lussem & Herbrecht, 2019; Post et al., 2015; Schmidt et al., 2012). WU1 is located above a planted spruce forest (*Picea abies*) of uniform height (Graf et al., 2014), while nearby WU2 is placed in an 8.6 ha area which was deforested in 2013 to allow a natural succession toward a European beech forest (Ney et al., 2019; Wiekenkamp et al., 2016). Currently (2020), spontaneous vegetation of the deforested area consists mostly of grass, shrubs (e.g., *Cytisus scoparius*) and young trees (*Sorbus aucuparius, Betula pubescens*).

EC data were aggregated from half-hourly fluxes to daily data. As only high to moderate quality EC data were used (quality flags 0 and 1), frequent gaps were created. The number (n) of days containing all 48

Table 1
Description of the Eddy Covariance Stations Providing CO_2 Flux Data Used for Model Training

Name	ID	Sensing period	Ecosystem type	Elevation (m)	Latitude	Longitude	Documentation
TERENO							
Merzenhausen	ME	05/2011–12/2018	Cropland	92	50.9297879	6.2969924	Eder et al. (2015)
Rollesbroich	RO	05/2011–12/2018	Grassland	514	50.6219142	6.3041256	Gebler et al. (2015)
Ruraue	RU1	10/2011–08/2017	Grassland	101	50.8636289	6.4274571	Schmidt and Schween (2018)
Engelsdorf	RU2	08/2012–10/2012	Cropland	108	50.9115426	6.3088546	Schmidt (2019)
Kall-Sistig	RU3	08/2012–07/2013	Grassland	499	50.5026827	6.525401	Post et al. (2015)
Niederzier	RU4	04/2013–07/2013	Cropland	101	50.8795149	6.4498871	Schmidt (2014)
Selhausen	SE	04/2011–12/2018	Cropland	101	50.8658521	6.4473198	Schmidt et al. (2012)
Wüstebach forest	WU1	06/2010–12/2018	Coniferous forest	624	50.5049269	6.33096248	Graf et al. (2014)
Wüstebach deforested	WU2	09/2013–12/2018	Deforested	628	50.50305	6.33596	Wiekenkamp et al. (2016)
FLUXNET							
Gebesee	DE-Geb	01/2002–12/2014	Cropland	162	51.09973	10.91463	10.18140/FLX/1440146
Grillenburg	DE-Gri	01/2004–12/2014	Grassland	385	50.95004	13.51259	10.18140/FLX/1440147
Hainich	DE-Hai	01/2002–12/2012	Deciduous Forest	430	51.07921	10.45217	10.18140/FLX/1440148
Leinefelde	DE-Lnf	01/2002–12/2012	Deciduous Forest	451	51.32822	10.3678	10.18140/FLX/1440150
Oberbärenburg	DE-Obe	01/2008–12/2014	Coniferous Forest	734	50.78666	13.72129	10.18140/FLX/1440151
Tharandt	DE-Tha	01/2002–12/2014	Coniferous Forest	385	50.96256	13.56515	10.18140/FLX/1440152

half-hourly intervals (100%) was only 386 for all TERENO stations combined, which is about 3.2% of all possible days and constitutes the first data set. Additional data sets were created with a varying number of missing 30-min intervals allowed: minimum 45/48 (93.75%) intervals of high to moderate quality (n = 1,035; 8.5% of possible days), 42/48 (87.5%) (n = 2,032; 16.6% of possible days), and 36/48 (75%) (n = 3,996; 32.7% of possible days). For the calculation of these daily NEE values, gap-filled data inferred with the REddyProc package (Wutzler et al., 2018) were used. In case gap-filled data were not available, the mean was calculated of all nonfilled values of each respective day. Based on the minimum of reliable half-hourly values included, these data sets are referred to in this paper as *48, 45, 42, and 36,* respectively.

Forest sites were underrepresented in the TERENO data, as only one coniferous forest site and no deciduous forest site were included. To achieve a better representation of each ecosystem type and to broaden the environmental envelope, we added daily NEE data with variable ustar-thresholds of six further stations (two coniferous forest, two deciduous forest, one grassland, one cropland site) from the FLUXNET2015 database (Pastorello et al., 2020), as shown in Table 1. Because quality-flag schemes may have differed among these sites, we filtered the FLUXNET data according to the relative uncertainty instead. In order to create data sets of equal proportions as the *48, 45, 42,* and *36* data sets, we took the X days with the lowest relative uncertainty, with X being 3.2%, 8.5%, 16.6%, and 32.7%, respectively. Finally, these FLUXNET data sets, were added to the *48, 45, 42,* and *36* data sets. The sensitivity of each of these data sets with a varying quality-quantity balance to the RF performance was then further evaluated with the feature selection and cross validation strategies described below.

2.3. Raster Data

Explanatory variables were compiled from various sources and were of different spatial and temporal resolutions, as shown in Table 2. These variables were chosen because they are regarded to potentially affect NEE, and were selected by availability for the whole time period 2010–2018. Vitale et al. (2016) showed that variations of vegetation indices such as Leaf Area Index (LAI) can highly influence carbon fluxes. Ishtiaq and Abdul-Aziz (2015) concluded that meteorological parameters have a strong linkage with CO_2 fluxes, especially "radiation-energy" components. Datetime variables such as Day of Year can also be a useful proxy for estimating CO_2 fluxes (Acosta et al., 2018). We used the following remotely sensed

Table 2
Predictor Variables Used for Model Training

Nr.	ID	Name	Source	Temporal resolution	Spatial resolution	Unit
1	NDVI	Normalized Difference Vegetation Index	MODIS	8 days	250 m	
2	EVI	Enhanced Vegetation Index	MODIS	8 days	250 m	
3	LAI	Leaf Area Index	MODIS	8 days	500 m	m^2/m^2
4	FPAR	Fraction of absorbed Photosynthetic Active Radiation	MODIS	8 days	500 m	%
5	Tree	Percent_Tree_Cover	MODIS	Yearly	250 m	%
6	NonTree	Percent_NonTree_Vegetation	MODIS	Yearly	250 m	%
7	NonVeget	Percent_NonVegetated	MODIS	Yearly	250 m	%
8	SWI	Downward Shortwave Radiation	Heliosat (2010–2017) MODIS (2018)	Daily	0.05° (Heliosat) 5 km (MODIS)	W/m^2
9	SWB	Net Shortwave Radiation	COSMO-REA6	Daily	6 km	W/m^2
10	LWI	Downward Longwave Radiation	COSMO-REA6	Daily	6 km	W/m^2
11	LWU	Upward Longwave Radiation	COSMO-REA6	Daily	6 km	W/m^2
12	LWB	Net Longwave Radiation	COSMO-REA6	Daily	6 km	W/m^2
13	Precip	Precipitation	COSMO-REA6	Daily	6 km	mm
14	Tair	Air Temperature (2 m)	COSMO-REA6	Daily	6 km	K
15	Tsoil	Soil Temperature	COSMO-REA6	Daily	6 km	K
16	rH	Relative Humidity (2 m)	COSMO-REA6	Daily	6 km	%
17	*u*	Zonal Wind Speed (10 m)	COSMO-REA6	Daily	6 km	m/s
18	*v*	Meridional Wind Speed (10 m)	COSMO-REA6	Daily	6 km	m/s
19	ETpot	Potential Evapotranspiration	DWD	Daily	1 km	mm
20	Soilm	Soil Moisture	DWD	Daily	1 km	%
21	Month	Month		Monthly	Static	
22	Season	Season		3 months	Static	
23	DoY	Day of Year		Daily	Static	
24	DEM	Elevation	SRTM	Static	30 m	m

MODIS Version 6 products: MCD18A1 for downward shortwave radiation (2018 only) (D. Wang, 2017), MOD44B for percent tree cover, percent nontree vegetation and percent nonvegetated land cover (Dimiceli et al., 2015), MCD15A2H for LAI and Fraction of absorbed Photosynthetic Active Radiation (FPAR) (Myneni et al., 2015), and MYD13Q1 and MOD13Q1 for Normalized Difference Vegetation Index (NDVI) and Enhanced Vegetation Index (EVI) (Didan, 2015). All of these data sets were quality controlled to exclude contaminated pixels with the quality assurance raster included in the MODIS products. Subsequently, for NDVI, EVI, LAI, and FPAR a Whittaker smoother (Atzberger & Eilers, 2011) was applied to fill gaps and smooth out noise in the data occurring from undetected clouds. Finally, these vegetation data sets were linearly interpolated in time from 8-day to daily data.

Daily gridded data for the meteorological variables air temperature and relative humidity in 2 m, soil temperature, precipitation, zonal, and meridional wind speed in 10 m, long wave upward and downward radiation at the surface and net shortwave and longwave radiation at the surface were obtained from the COSMO-REA6 regional reanalysis data set (Bollmeyer et al., 2015) and were regridded with Climate Data Operators (Schulzweida, 2019). Furthermore, daily downward shortwave radiation from 2010 to 2017 was acquired from the Heliosat (SARAH-2) Surface Solar Radiation Data Set (Pfeifroth et al., 2019), other variables include a digital elevation model from the Shuttle Radar Topography Mission (Jarvis et al., 2008), and soil moisture and potential evapotranspiration from the German Weather Service (DWD, 2019) based on Löpmeier (1994).

Raster data were used and further processed at two different steps in the analysis, to i) extract values at the coordinates of each site for model training and validation, and ii) predict NEE for the entire catchment area. For the latter, all raster sets were homogenized to the same extent and same spatial resolution of 250 m with bilinear interpolation of the raster package in R (Hijmans, 2020).

2.4. RF Model

RF is a machine learning method based on an ensemble of many binary decision trees. The algorithm was introduced by Breiman (2001) and is widely used for classification and regression in ecology (e.g., Aide et al., 2012; Prasad et al., 2006; Tramontana et al., 2016). Each decision tree is grown with a random subsample with replacement of the input data, called bootstrapping (Efron, 1979). At each node in the decision tree, a threshold of a randomly selected explanatory variable is ascertained to split the data into the two most homogeneous subgroups, i.e. with the lowest variance. The leaf nodes at the end of the tree do not further split the data but contain predictions of the target variable. This value is the mean of the target variable of all elements in the corresponding subgroup. For the final prediction, results of all trees (in this case 500) are averaged to overcome weaknesses of single trees. One consequence of this algorithm, however, is that predictions cannot be out of bounds of the training range. In this study, we used the implementation of the RF code in the randomForest package in R by Liaw and Wiener (2002) to predict NEE in a regression approach. In order to identify an ideal number of predictor variables used at each split node ("mtry"), model tuning was conducted with the caret package, which is a wrapper to perform model tuning for various predictive models.

In order to perform feature selection, we first split the data into spatial and temporal folds (described in the next section in detail) with the CAST package (Meyer, 2018). In a next step, we applied the forward feature selection (FFS) procedure of this package with root mean squared error (RMSE) as performance metric to punish high errors in particular. The advantage of CAST FFS is that feature selection results are based on spatiotemporal cross validation rather than on training data only. However, as FFS sometimes selected very few variables with unsatisfactory performance (see Section 3.1), a slightly modified version of this procedure has been devised and compared to FFS. We applied a backward feature elimination (BFE), which starts with all features and iteratively removes the worst feature based on a spatial or temporal cross validation. Conversely, FFS iteratively adds features to the best combination of two features. Since RF relies on randomization, results can significantly differ between model runs. Hence, each iteration within BFE was repeated five times to average out such randomization effects. This leads, however, to a significant increase in computation time. The general algorithm of the BFE procedure is described in Table 3. To illustrate the impact of these two feature selection procedures on model performance, model runs without any feature selection were evaluated too.

2.5. Cross Validation

In order to assess RF performance beyond the scope of training data, NEE predictions have been cross-validated for (a) spatially and (b) temporally independent test sets. The additional FLUXNET data from outside the catchment were only used for training, whereas the TERENO data were used for training and testing. Figure 2 displays the cross validation strategies in a schematic way. Especially spatially independent test sets may be important for the assessment of the upscaled NEE predictions presented in Section 3.2 because they simulate predictions for pixels without any information used for model training. These cross validation strategies have been performed for all different NEE data sets and feature selection methods.

a) Data were split into folds by station ID. Stations were only considered for leave-out if they i) were not the only station of their land cover class to ensure that the class was still contained in the training data when excluding the station and ii) had data spanning over at least three years to ensure representative results. Therefore, five stations (SE, RU1, ME, RO, and WU1) were regarded, though all other stations were included for training.
b) Data were split into folds by year. Each fold containing one year of data was left out once and predicted by a model trained with the other years. 2010 has only been regarded for training and not for testing as only data from one TERENO site was available for 2010.

Table 3
General Algorithm of the Backward Feature Elimination (BFE), the Indications in Square Brackets Refer to the R-Packages Used for the Respective Step

Algorithm 1	
01:	Split data into spatial or temporal folds [CAST]
02:	Do 5 times with all predictor variables:
03:	Train and test model with a leave-one fold-out cross validation [caret]
04:	Calculate RMSE over all folds
05:	Get *previous* ← Averaged RMSE over five repetitions
06:	**For** each remaining predictor variable do:
07:	Exclude variable
08:	Do 5 times:
09:	Train and test model with a leave-one fold-out cross validation [caret]
10:	Calculate RMSE over all folds
11:	Average RMSE over 5 repetitions
12:	Get *bestSubset* ← variable subset with lowest average RMSE
13:	Get *bestRMSE* ← RMSE of *bestSubset*
14:	**If** *bestRMSE* < *previous*: **stop**
15:	**Else:** *previous* ← *bestRMSE*
16:	**repeat** from step 06 onwards with *bestSubset*

Abbreviation: RMSE, root mean squared error.

We used the coefficient of determination (R^2), the mean absolute error (MAE) and RMSE as metrics to evaluate model performance.

The relative importance of selected variables for model building was assessed through RF's internal variable importance metric implemented in the randomForest package. For this, data points of each variable are randomly permutated and the relative increase of the mean squared error (MSE) based on an internal cross validation within in the training data is measured. This error is assumed to increase if the variable is important.

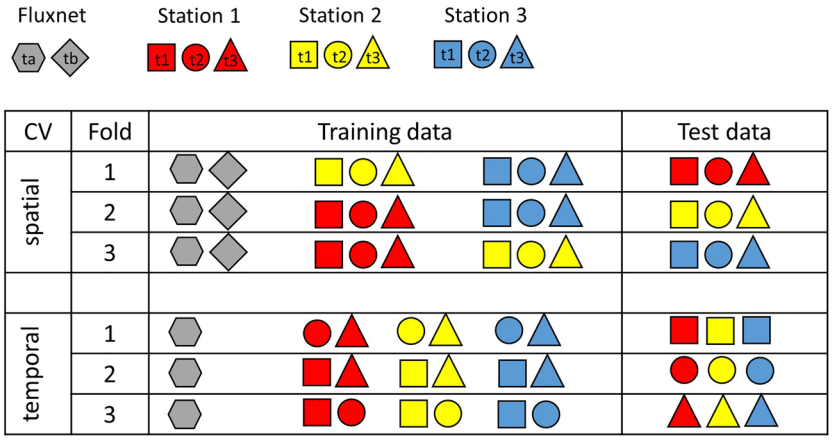

Figure 2. Schematic depiction of the spatial and temporal cross validation (CV) strategies that have been applied for this analysis. t1, t2, and t3 refer to three different time steps (years in our case), while ta refers to the FLUXNET data which are outside the timespan of TERENO data (prior to 2010) and tb refers to the years included in the TERENO data (2010–2014). The figure was recreated and modified after Meyer et al. (2018).

Table 4
Cross Validation Results for Different NEE Data sets and Different Feature Selection Methods

Cross validation	Data set	Feature selection	RMSE	MAE	R^2	*n* Var	Variables selected
Spatial							
	48	FFS	2.96	2.25	0.35	6	9,12,19,22,24
		BFE	2.72	1.9	0.41	15	2,5:7,10:12,14:20,22
		None	3.01	2.19	0.34	24	1:24
	45	FFS	2.88	2.07	0.43	10	2,5,7,8,12,14,19:21,23
		BFE	2.84	2.07	0.42	10	5,7,11,12,14,17,19:22
		None	3.16	2.24	0.34	24	1:24
	42	FFS	3.18	2.23	0.34	9	1,2,5,7,8,12,15,21,22
		BFE	3.2	2.25	0.33	14	1,2,5:8,12,14,:16,19,21,23
		None	3.32	2.37	0.31	24	1:24
	36	FFS	3.4	2.42	0.32	10	1,2,5:9,11,21,22
		BFE	3.4	2.41	0.32	17	1:3,5:9,11,14:18,20:22
		None	3.51	2.54	0.3	24	1:24
Temporal							
	48	FFS	2.54	1.89	0.31	8	5,6,8,9,16,19,22,24
		BFE	2.55	1.78	0.32	11	1,5,8,14,16,19,21:24
		None	2.91	2.02	0.24	24	1:24
	45	FFS	2.67	1.84	0.46	8	1,5,10,17,19,21,23:24
		BFE	2.59	1.79	0.5	15	1,5:8,11,12,15,17,19:24
		None	2.66	1.82	0.46	24	1:24
	42	FFS	3.01	2.08	0.44	12	1,2,6:9,14,17:19,23:24
		BFE	2.95	2.05	0.45	14	2,5:9,11,14,15,17,18,21:22,24
		None	3.06	2.1	0.43	24	1:24
	36	FFS	3.34	2.32	0.44	13	1:3,5:8,11,17:19,21,24
		BFE	3.32	2.28	0.44	16	2,3,5:11,14,17,18,21:22,24
		None	3.41	2.39	0.42	24	1:24

Note. Performance is displayed as root mean squared error (RMSE; in g C $m^{-2}d^{-1}$), mean absolute error (MAE; in g C m $m^{-2}d^{-1}$) and coefficient of determination (R^2). *n* Var gives the number of selected variables, variables selected lists which variables were selected by the Nr. Stated in Table 2. RMSE, MAE, and R^2 values are reported as averages of the respective folds.

Abbreviations: BFE, backward feature elimination; FFS, forward feature selection.

3. Results

3.1. Cross Validation Results

We evaluated NEE predictions with a leave-one-fold-out cross validation by withholding either sites (spatial cross validation) or years (temporal cross validation). Table 4 summarizes the spatial and temporal cross validation results of different NEE data sets and feature selection methods. The results indicate that spatial prediction was generally more difficult than temporal prediction. Models trained with the *48* data set had the lowest RMSE values, though with the *45* data set slightly higher R^2 could be obtained. Especially for temporal cross validation, RMSE/MAE and R^2 did not coincide well with each other. The feature selection methods FFS and BFE were either closely in line with each other or BFE gave considerably better results than FFS. In these cases, FFS always selected less variables than BFE. Furthermore, BFE always performed better than no feature selection, indicating that BFE is more suitable than FFS or no feature selection for this analysis. Few patterns can be observed regarding selected variables, most notably that elevation was selected for every model with temporal cross validation, though very seldom for spatial prediction. The

Table 5
Importance of Each Variable Selected by BFE for the 48 Data set

Variable	%IncMSE
EVI	13.85
ETpot	12.29
Season	3.49
rH	3.02
LWB	2.70
Tree	2.55
LWU	2.26
LWI	2.24
Tsoil	2.18
NonTree	2.11
Tair	2.06
Soilm	1.74
v	1.11
NonVeget	1.01
u	0.60

Note. %IncMSE describes the percental increase of the mean squared error after permutation as implemented in the randomForest package.
Abbreviations: BFE, backward feature elimination; EVI, Enhanced Vegetation Index; LWB, Net Longwave Radiation; LWI, Downward Longwave Radiation; LWU, Upward Longwave Radiation.

variables LAI, FPAR, and Precip, however, were almost always excluded. For the final upscaling we selected the model with the best RMSE by spatial cross validation in order to optimize it for spatial prediction and avoiding high errors. Hence, we chose the model trained with the *48* data set and tuned with BFE.

The relative importance of selected variables was assessed by the importance function of the randomForest package. As shown in Table 5, the most important variables were EVI and ETpot. Figure 3 displays an assessment of the quality of NEE predictions in comparison to observed TERENO validation data. While predictions and observations have almost the same mean values (−2.31 and −2.3 g C $m^{-2}d^{-1}$) and rather similar median values (−1.84 and −1.33 g C $m^{-2}d^{-1}$), and the regression line a slope close to 1 ($y = 0.15 + 1.06x$; Figure 3a), the standard deviation of predictions (2.5 g C $m^{-2}d^{-1}$) is much lower compared to observations (3.73 g C $m^{-2}d^{-1}$). The interquartile range of predictions is also narrower than of observations (−3.96 to −0.27 compared to −4.49 to 0.29 g C $m^{-2}d^{-1}$; Figure 3b). As for predictions, values from about −5 to 1 g C $m^{-2}d^{-1}$ were more frequent, and outside of that range less frequent than in observations, resulting in a narrower distribution of values (Figure 3c). This results in higher absolute errors for high flux magnitudes, especially for positive fluxes (Figure 3d).

3.2. Upscaling Results

We predicted daily NEE data at 250 m spatial resolution for the Rur catchment from 2010 to 2018. According to the results of the previous section, the *48* NEE data set and explanatory variables selected with spatial BFE were used for model training. Table 6 shows the upscaled results aggregated by land cover class and season. To put these results into perspective, such aggregations over actual measurements within the catchment are also included in Table 6. Pixels classified as urban or built-up were excluded from the analysis because anthropogenic CO_2 emissions were not represented in the training data. Results show that vegetated areas of the Rur catchment were on average a CO_2 sink between 2010 and 2018 with about −0.65 g C $m^{-2}d^{-1}$. Grasslands and deciduous forests were the strongest sink (−0.76 g C $m^{-2}d^{-1}$ and −0.72 g C $m^{-2}d^{-1}$, respectively), while croplands captured the least net amount of CO_2 (−0.56 g C $m^{-2}d^{-1}$). During winter (December–February) and fall (September-November), the Rur catchment was a CO_2 source (0.86 g C $m^{-2}d^{-1}$ and 0.75 g C $m^{-2}d^{-1}$, respectively), while in spring (March–May) it was a strong sink (−2.14 g C $m^{-2}d^{-1}$), closely followed by summer (June–August; −2.02 g C $m^{-2}d^{-1}$). Figure 4 shows yearly courses of predicted NEE aggregated by land cover for the investigation period. Additionally, daily NEE raster were aggregated by season and the whole investigation period (Figure 5). These results show that all land cover classes were a CO_2 source in fall and winter and sink in spring and summer, although the CO_2 uptake started decreasing in summer already. Croplands were the earliest to become a sink in spring and also to turn into a source after day of year 200. This NEE sink capacity decrease of croplands from spring to summer is also observable in Figure 5 as croplands prevail in the northern half of the catchment. Forests were a stronger source than croplands and grasslands in fall and winter, though deciduous forests were also the strongest sink with average NEE below −5 g C $m^{-2}d^{-1}$ around day of year 170. However, coniferous forests and especially deciduous forests were a greater sink in summer in actual measurements, and no CO_2 source in fall. In contrast, grasslands were predicted to be a greater sink in summer compared to actual measurements. Differences between land cover classes were in general less pronounced in upscaled predictions than in measurements.

4. Discussion

The results of the study showed that a data-driven upscaling of NEE to the regional scale predicted the average NEE well though underestimated the variance (Figure 3b). Feature selection and the right quality-quantity balance of NEE data, however, can improve model performance. Similar to our results, high errors for

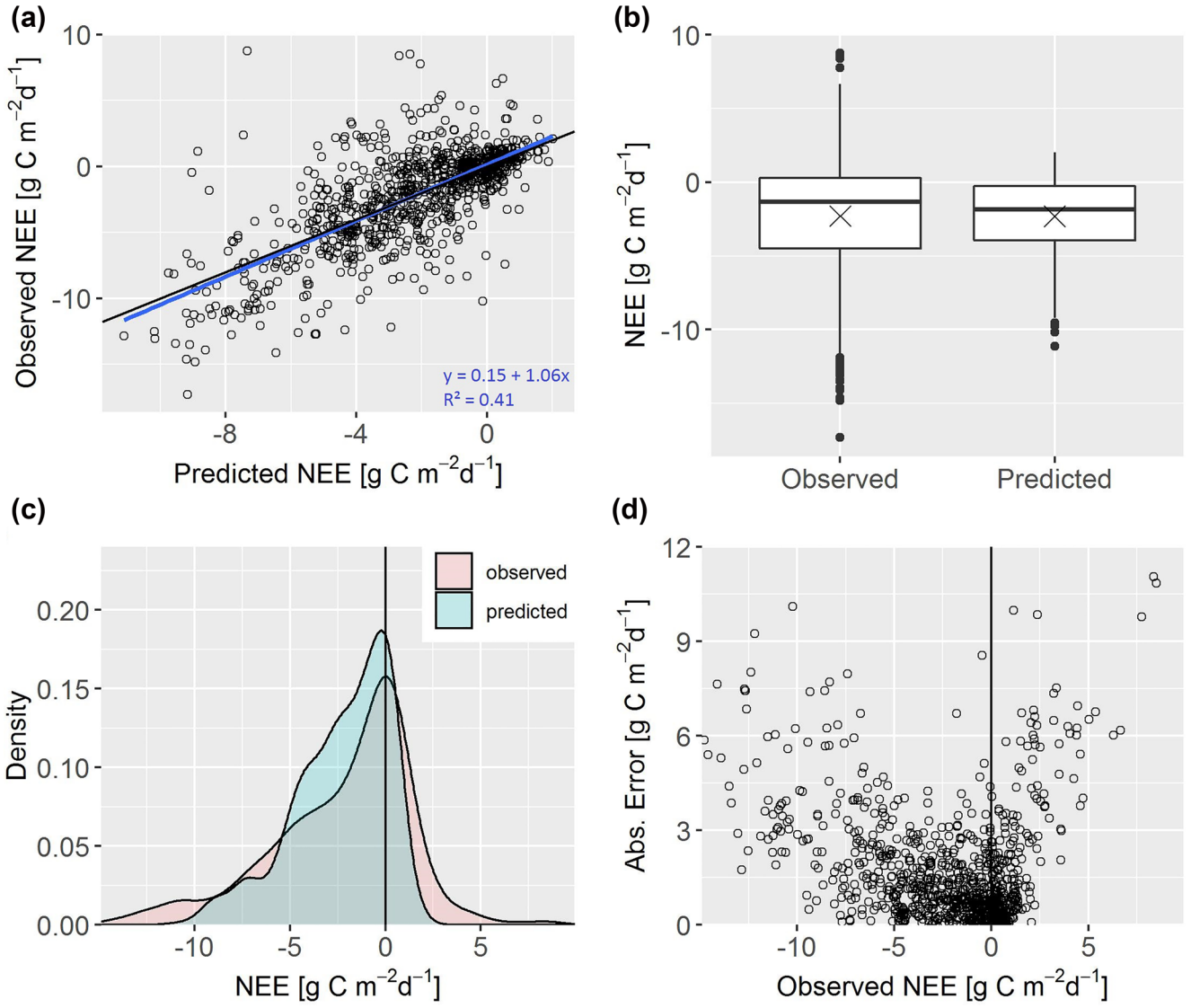

Figure 3. Assessment of prediction quality in comparison to observations from TERENO test data. (a) scatter plot with regression line (blue) and identity line (black); (b) boxplots with mean values displayed as crosses; (c) density plot of predicted (blue) and observed (red) NEE values; (d) absolute error by observed NEE. The vertical line at zero marks the border between CO_2 sinks (left) and CO_2 sources (right).

NEE upscaling have also been observed in other studies (Jung et al., 2011; Tramontana et al., 2016). Xiao et al. (2011) showed that an ecosystem model predicted NEE with an R^2 between 0 and 0.66, depending on the site. Richardson et al. (2012) demonstrated increasing random errors with flux magnitude for half-hourly CO_2 flux measurements. The absolute errors displayed in Figure 3d show a similar pattern, indicating that high flux magnitudes may be difficult to predict and validate because actual measurements in those ranges are already error-prone.

Meyer et al. (2018) demonstrated that random cross validation lead to an overoptimistic view of the model performance compared to spatial cross validation. In our case, RMSE could be improved to 1.85 g C m^{-2}d^{-1} and R^2 up to 0.82 with a random cross validation, indicating that mere data reproduction is much easier than actual spatial prediction. Elevation was named as a typical example of a spatially autocorrelated predictor by Meyer et al. (2019). Hence, it is reasonable that it was removed by FFS and BFE for spatial cross validation, but not for temporal cross validation. Besnard et al. (2019) concluded that integrating memory effects of past disturbances in a recurrent neural network outperforms nondynamic statistical models like RF. So not including memory effects in our study might limit the prediction capacities.

Table 6
NEE in g C $m^{-2}d^{-1}$ Aggregated by Land Cover and Season of Upscaled Data Over Vegetated Parts of the Rur Catchment (above) and of eddy covariance (EC) Measurements Within the Catchment Described in 1 (below)

Upscaling results					
Land cover	Winter	Spring	Summer	Fall	Year
Coniferous F.	0.92	−1.75	−2.54	0.9	−0.63
Deciduous F.	1.07	−1.99	−2.77	0.85	−0.72
Grassland	0.87	−2.14	−2.35	0.6	−0.76
Cropland	0.67	−2.41	−1.17	0.74	−0.56
Study Area	0.86	−2.14	−2.02	0.75	−0.65
EC tower measurements					
Land cover	Winter	Spring	Summer	Fall	Year
Coniferous F.	0.46	−2.36	−3.5	−1.1	−1.63
Deciduous F.*	1.15	−0.95	−6.11	−0.64	−1.64
Grassland	0.56	−1.71	−0.83	0.55	−0.36
Cropland	0.38	−3.08	−1.79	0.76	−0.93

Note. *Values for the class deciduous forest were calculated from the two FLUXNET stations De-Hai and De-Lnf as no measurements of that class were available within the catchment.

One intrinsic feature of RF is to not extrapolate beyond the input data due to the prediction being the average target value of the subgroup within a leaf node. Although we attempted to overcome this limitation by including sites from outside the catchment and hence broadening the environmental envelope, outliers

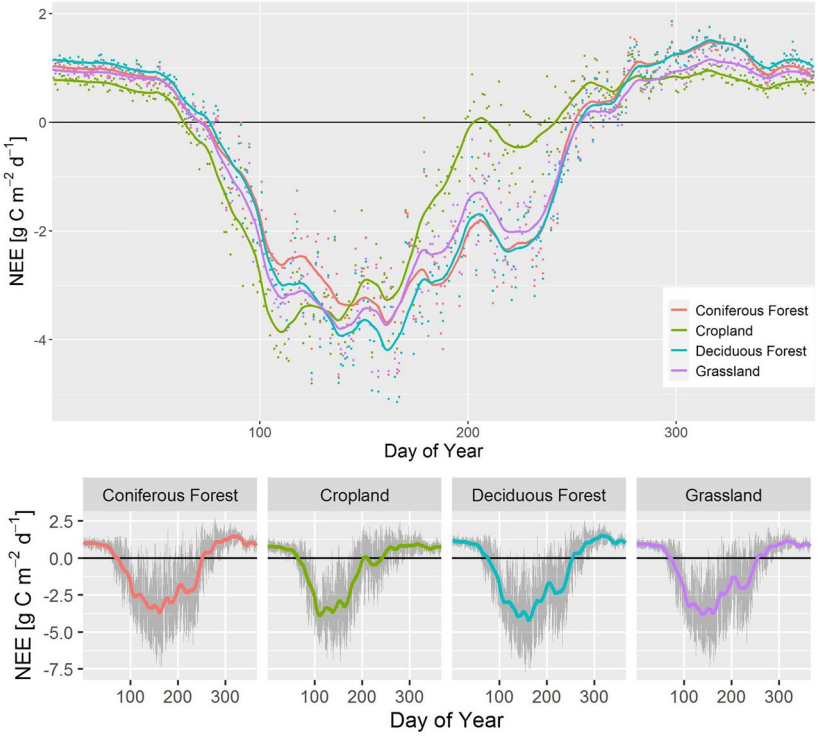

Figure 4. Yearly courses of predicted net ecosystem exchange (NEE) aggregated by each land cover class for the period 2010–2018 (points), and smoothed by a loess function (lines; above). Standard deviation (±) for each of those lines separately (gray bands; below).

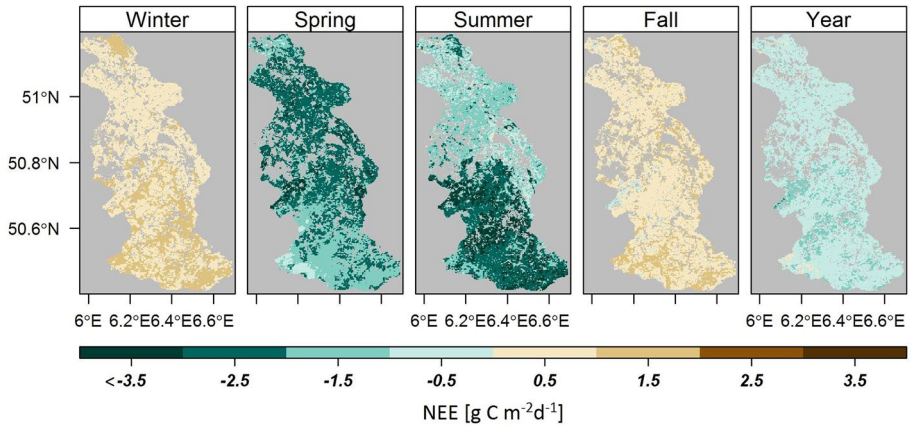

Figure 5. Predicted net ecosystem exchange (NEE) of the Rur catchment aggregated for the whole investigation period 2010–2018 and for each season. Mean values for each map are 0.86 g C $m^{-2}d^{-1}$ (winter), −2.14 (spring), −2.02 (summer), 0.75 (fall), and −0.65 (year).

with high flux magnitudes were still underpredicted. A portion of the prediction error can also be attributed to uncertainties in the raster data sets used for model training and predictions. Some of the most important variables such as ETpot, Tair, and rH were also measured in situ at the EC stations SE, RO, WU1, and WU2. Averaged over these four stations, Tair from Cosmo-REA6 coincided very well with in situ Tair ($R^2 = 0.99$); the same applies to rH ($R^2 = 0.88$), and modeled ETpot ($R^2 = 0.93$). SWI was combined from two different sources without data overlap but both sources also agree well with in situ measurements (Heliosat: R^2 of 0.96; MODIS: R^2 of 0.92). However, we assume that MODIS-based vegetation indices did not capture small-scale vegetation structures well and hence contributed to prediction errors. A possibility to improve vegetation data to inform the RF model would be to use remote sensing data with a higher spatial resolution such as Sentinel-2 (Drusch et al., 2012), which was not used here because it did not cover the whole investigation period. Another limitation of our study comprises the exclusion of 20.4% of the land area from the analysis because anthropogenic fluxes were not measured. This high proportion results from the high population density in the northern part of the catchment and the relatively large (13 km^2) Inden open pit mine. However, only small biospheric net fluxes are to be expected from these areas as they are to a large extent non vegetated and thus may not contribute much to the overall biospheric fluxes of the catchment.

The results indicate that smaller data sets incorporating only few (<6.25%) or no low-quality intervals in the aggregated daily fluxes are more beneficial than larger data sets with more low-quality data. Small data sets can increase overfitting of a predictive model; however, the ensemble characteristic of RF of averaging multiple trees also counteracts overfitting. Thus, it seems reasonable for RF to select for small data sets with higher quality. Although a standardized quality-flag scheme was used on the TERENO-data set, it should be noted that quality-flagging is not fully standardized in the flux-community yet. Thus, our thresholds may not be transferable to other schemes.

As uncertainty is correlated with flux magnitude, filtering the FLUXNET data by small relative uncertainties has the side-effect to favor large NEE values and discriminate small ones, whereas quality flags are not correlated with magnitude. However, the distribution of the TERENO NEE magnitudes shows that, naturally, small fluxes occur much more frequent than large fluxes (Figure 3c). Such imbalanced data is a problem for RF, which requires about equally sized domains in the training data to not overpredict the largest domains (Krawczyk, 2016; Torgo et al., 2013). Therefore, favoring large fluxes in the FLUXNET data improves their representation in the training data sets. The test data sets, however, consisting only of quality-flag filtered data, remained unbiased and are thus regarded suitable for model cross validation. Even so, Figure 3c shows that the maximum around NEE = 0 in the training data was still overpredicted and rare domains underpredicted, indicating that the training data probably was still not balanced enough.

The performance differences between BFE and FFS can partly be explained by local optima of variable subsets, as sequential feature selection algorithms are prone to being trapped in such local optima (Liu &

Motoda, 2008). In these cases, the first local optimum trap for BFE is much closer or even identical to the absolute optimum than the first local optimum trap for FFS. Hence, a BFE is regarded superior in such cases. The relatively high variance between RF model runs increased fluctuations and can thus amplify this effect by creating artificial local optima, leading to a suboptimal variable selection. Averaging five model runs reduced the variance within 100 model runs by about 76%, generally leading to more robust results. However, it should be noted that variance between model runs can be lower for other machine learning algorithms and that repeating and averaging is computationally expensive and therefore not suitable for large numbers of variables to select from. For such cases, a genetic algorithm like GHGA (Jung & Zscheischler, 2013) may be more appropriate.

The relatively high elevations of forests and grasslands in the catchment resulted in lower average annual temperatures in years 2010–2018 (coniferous forests: 8.9°C, deciduous forests: 9.6°C, grasslands: 9.5°C) compared to croplands (10.8°C), and hence a later start of the growing season might be an explanation for croplands being an earlier CO_2 sink. Deciduous trees, on the other hand, first need to build-up the canopy leaf area to utilize suitable conditions for photosynthesis, though having higher photosynthesis capacities when fully leafed. However, differences between land cover types were less pronounced in upscaled results than in actual measurements (Table 6). One explanation for this might be mixed pixels in MODIS EVI (250 m spatial resolution) containing spectral responses from different land cover types.

The catchment was a slightly stronger CO_2 sink in spring than in summer. Lindroth et al. (2008) stated that net CO_2 uptake in Swedish spruce forests is shifted toward the earlier parts of the growing season because respiration was still low while radiation was already high. Managed grasslands on the other hand, usually are cut several times during summer. For example, Rollesbroich was cut three times in the growing season of 2013 (Borchard et al., 2015) and each defoliation had the potential to turn grassland temporally into a CO_2 source (Wohlfahrt et al., 2008). Croplands showed the largest decrease of CO_2 uptake in late summer. Schmidt et al. (2012) analyzed vegetation parameters of a winter wheat field in the catchment over the course of two years. LAI of living/green leaves reached the maximum in early May, plant senescence (LAI of brown leaves) already started in late April and reached its peak in July. Although these patterns can differ for other crops, the results still indicate that specific croplands uptake the most CO_2 in spring. The EC aggregations in Table 6 further confirm a decrease of CO_2 uptake in summer for croplands and grasslands. In comparison, the aggregated EVI of the whole Rur catchment started slowly increasing in late February, peaked in early June, and declined afterward. Graf et al. (2020) showed that the exceptional drought and heat across Central Europe during the 2018 growing season resulted in a reduced net CO_2 uptake for many drought-affected EC stations, including SE, RO and WU1. The whole catchment was predicted to be a significantly weaker CO_2 sink in summer 2018 (-0.89 g C $m^{-2}d^{-1}$) compared to 2010–2017 (-2.16 ± 0.45 g C $m^{-2}d^{-1}$). Whereas in spring 2018 the sink capacity decrease to 2010–2017 was less distinct (-1.92 g C $m^{-2}d^{-1}$ in 2018 compared to -2.16 ± 0.51 g C $m^{-2}d^{-1}$ 2010–2017), indicating that the seasonal averages may be influenced by one exceptional year. In view of these findings, we consider the seasonal variations of upscaled NEE as largely plausible.

5. Conclusion

In this study, we scaled up daily EC NEE data to the regional scale at 250 m spatial resolution with a RF model integrating remote sensing and reanalysis data. Furthermore, we evaluated the impact of feature selection and NEE data quality-quantity balance on the model performance. We conclude that upscaling results can be improved with a BFE to remove unnecessary predictors and by incorporating no or only small (<6.25%) amounts of low-quality intervals in the aggregated daily NEE data. Therewith, we provided a data-driven approach for predicting spatial NEE data sets which can be used for assessing the CO_2 uptake of heterogeneous local and regional ecosystems or calibrating and validating process-based models. However, the spread of NEE observations and differences between land cover types were underestimated.

Vegetated parts of the Rur catchment acted as a CO_2 sink between 2010 and 2018 with about -0.65 g C $m^{-2}d^{-1}$. The catchment was predicted to be a slightly stronger sink in spring than in summer probably partly due to plant senescence increasing in summer in cropland and grassland ecosystems, while it was a CO_2 source during fall and winter. In future work, a model incorporating emissions from urban and built-up

areas should be implemented to produce spatially continuous NEE data sets. Furthermore, remotely sensed vegetation products with a higher spatial resolution are likely to improve model accuracy as they would allow to distinguish small-scale vegetation structures.

Data Availability Statement

Data used in this study are freely available from the TERENO and TR32 portals, the FLUXNET2015 database, the LP DAAC Catalog (MCD18A1, MOD44B, MCD15A2, MYD13Q1, MOD13Q1), the EUMETSAT Navigator, and CGIAR-CSI. Upscaled daily NEE data for the Rur catchment and data from the German Weather Service used in this study are stored at https://doi.org/10.5281/zenodo.3776011.

Acknowledgments
The eddy covariance data set used in this study was contributed by within the projects TERENO and the collaborative research center TR32 funded by the DFG (Deutsche Forschungsgemeinschaft). In particular, we thank the Meteorological (Jan Schween, site RU1), and the Geographical Institute (Karl Schneider, site ME) of the University of Cologne, the Institute of Environmental Meteorology of the University of Trier (Uwe Baltes, Clemens Drüe, Günther Heinemann, site WU1) for hardware and maintenance assistance as well as Daniel Dolfus, Nicole Adels, David Symanzik and Martina Kettler, Agrosphere Institute of the Forschungszentrum Jülich, for maintenance and calibration. We also thank Alexander Knohl (University of Göttingen), Christian Bernhofer (TU Dresden) and Christian Brümmer (Thünen Institute) for providing data to the FLUXNET2015 data set. We thank Ladislav Šigut and one anonymous reviewer for very helpful comments during the review process, which improved the manuscript substantially. Open access funding enabled and organized by Projekt DEAL.

References

Abdalla, M., Saunders, M., Hastings, A., Williams, M., Smith, P., Osborne, B., et al. (2013). Simulating the impacts of land use in Northwest Europe on Net Ecosystem Exchange (NEE): The role of arable ecosystems, grasslands and forest plantations in climate change mitigation. *Science of the Total Environment*, *465*, 325–336. https://doi.org/10.1016/j.scitotenv.2012.12.030

Acosta, M., Darenova, E., Krupkova, L., & Pavelka, M. (2018). Seasonal and inter-annual variability of soil CO2 efflux in a Norway spruce forest over an eight-year study. *Agricultural and Forest Meteorology*, *256–257*, 93–103. https://doi.org/10.1016/j.agrformet.2018.03.005

Aide, T. M., Clark, M. L., Grau, H. R., Lopez-Carr, D., Levy, M. A., Redo, D., et al. (2012). Deforestation and reforestation of Latin America and the Caribbean (2001–2010). *Biotropica*, *45*, 262–271. https://doi.org/10.1111/j.1744-7429.2012.00908

Atzberger, C., & Eilers, P. H. C. (2011). A time series for monitoring vegetation activity and phenology at 10-daily time steps covering large parts of South America. *International Journal of Digital Earth*, *4*(5), 365–386. https://doi.org/10.1080/17538947.2010.505664

Baatz, R., Bogena, H. R., Hendricks Franssen, H. J., Huisman, J. A., Qu, W., Montzka, C., & Vereecken, H. (2014). Calibration of a catchment scale cosmic-ray probe network: A comparison of three parameterization methods. *Journal of Hydrology*, *516*, 231–244. https://doi.org/10.1016/j.jhydrol.2014.02.026

Baldocchi, D. D. (2014). 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. https://doi.org/10.1111/gcb.12649

Besnard, S., Carvalhais, N., Arain, M. A., Black, A., Brede, B., Buchmann, N., et al. (2019). Memory effects of climate and vegetation affecting net ecosystem CO_2 fluxes in global forests. *PloS One*, *14*(2), e0211510. https://doi.org/10.1371/journal.pone.0211510

Bodesheim, P., Jung, M., Gans, F., Mahecha, M. D., & Reichstein, M. (2018). Upscaled diurnal cycles of land–atmosphere fluxes: A new global half-hourly data product. *Earth System Science Data*, *10*, 1327–1365. https://doi.org/10.5194/essd-10-1327-2018

Bogena, H., Borg, E., Brauer, A., Dietrich, P., Hajnsek, I., Heinrich, I., et al. (2016). TERENO: German network of terrestrial environmental observatories. *Journal of Large-Scale Research Facilities*, *2*, A52. http://dx.doi.org/10.17815/jlsrf-2-98

Bollmeyer, C., Keller, J. D., Ohlwein, C., Wahl, S., Crewell, S., Friederichs, P., et al. (2015). Toward a high-resolution regional re-analysis for the European CORDEX domain. *Quarterly Journal of the Royal Meteorological Society*, *141*, 1–15. https://doi.org/10.1002/qj.2486

Borchard, N., Schirrmann, M., von Hebel, C., Schmidt, M., Baatz, R., Firbank, L., et al. (2015). Spatio-temporal drivers of soil and ecosystem carbon fluxes at field scale in an upland grassland in Germany. *Agriculture, Ecosystems & Environment*, *211*, 84–93. https://doi.org/10.1016/j.agee.2015.05.008

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5–32.

Chen, B., Black, T. A., Coops, N. C., Hilker, T., Trofymow, J. A., & Morgenstern, K. (2009). Assessing tower flux footprint Climatology and scaling between remotely sensed and eddy covariance measurements. *Boundary-Layer Meteorology*, *130*, 137–167. https://doi.org/10.1007/s10546-008-9339-1

Cutler, D. R., Edwards, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J. (2007). Random forests for classification in ecology. *Ecology*, *88*, 2783–2792.

Denman, K. L., Brasseur, G., Chidthaisong, A., Ciais, P., Cox, P. M., Dickinson, R. E., et al. (2007). Couplings between changes in the climate system and biogeochemistry. In S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, et al. (Eds.), *Climate change 2007: The physical science basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 499–588). Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press.

Didan, K. (2015). *MOD13Q1 MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid V006 [Data set]. NASA EOSDIS Land Processes DAAC*. https://doi.org/10.5067/MODIS/MOD13Q1.006

Dimiceli, C., Carroll, M., Sohlberg, R., Kim, D. H., Kelly, M., & Townshend, J. R. G. (2015). *MOD44B MODIS/Terra Vegetation Continuous Fields Yearly L3 Global 250m SIN Grid V006 [Data set]. NASA EOSDIS Land Processes DAAC*. https://doi.org/10.5067/MODIS/MOD44B.006

Dou, X., Yang, Y., & Luo, J. (2018). Estimating forest carbon fluxes using machine learning techniques based on eddy covariance measurements. *Sustainability*, *10*(1), 203. https://doi.org/10.3390/su10010203

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., et al. (2012). Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sensing of Environment*, *120*, 25–36. https://doi.org/10.1016/j.rse.2011.11.026

DWD Climate Data Center. (2019). Daily grids of potential evapotranspiration over grass and soil moisture under grass and sandy loam. version 0.x.

Eder, F., Schmidt, M., Damian, T., Träumner, K., & Mauder, M. (2015). Mesoscale Eddies affect near-surface turbulent exchange: Evidence from Lidar and tower measurements. *Journal of Applied Meteorology and Climatology*, *54*, 189–206. https://doi.org/10.1175/JAMC-D-14-0140.1

Efron, B. (1979). Bootstrap methods: Another look at the Jackknife. *Annals of Statistics*, *7*(1), 1–26.

Foken, T., & Wichura, B. (1996). Tools for quality assessment of surface-based flux measurements. *Agricultural and Forest Meteorology*, *78*, 83–105. https://doi.org/10.1016/0168-1923(95)02248-1

Gebler, S., Hendricks Franssen, H.-J., Pütz, T., Post, H., Schmidt, M., & Vereecken, H. (2015). Actual evapotranspiration and precipitation measured by lysimeters: A comparison with eddy covariance and tipping bucket. *Hydrology and Earth System Sciences*, *19*, 2145–2161. https://doi.org/10.5194/hess-19-2145-2015

Graf, A., Bogena, H. R., Drüe, H., Pütz, T., Heinemann, G., & Vereecken, H. (2014). Spatiotemporal relations between water budget components and soil water content in a forested tributary catchment. *Water Resources Research*, *50*, 4837–4857. https://doi.org/10.1002/2013WR014516

Graf, A., Klosterhalfen, A., Arriga, N., Bernhofer, C., Bogena, H., Bornet, F., et al. (2020). Altered energy partitioning across terrestrial ecosystems in the European drought year 2018. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *375*(1810). https://doi.org/10.1098/rstb.2019.0524

Hall, M. A., & Smith, L. A. (1999). *Feature selection for machine learning: Comparing a correlation-based filter approach to the wrapper*. Paper presented at 12th International FLAIRS Conference, 235–239.

Hijmans, G. (2020). *raster: Geographic data analysis and modeling*. R package version 3.0-12. Retrieved from https://CRAN.R-project.org/package=raster

Ishtiaq, K. S., & Abdul-Aziz, O. I. (2015). Relative linkages of canopy-level CO2 fluxes with the climatic and environmental variables for US deciduous forests. *Environmental Management*, *55*, 943–960. https://doi.org/10.1007/s00267-014-0437-1

Jarvis, A., Reuter, H. I., Nelson, A., & Guevara, E. (2008). *Hole-filled seamless SRTM data V4*. International Centre for Tropical Agriculture (CIAT). Retrieved from http://srtm.csi.cgiar.org

Jung, M., Reichstein, M., Margolis, H. A., Cescatti, A., Richardson, A. D., Arain, M. A., et al. (2011). 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*, *116*, G00J07. https://doi.org/10.1029/2010JG001566

Jung, M., Schwalm, C., Migliavacca, M., Walther, S., Camps-Valls, G., Koirala, S., et al. (2020). Scaling carbon fluxes from eddy covariance sites to globe: Synthesis and evaluation of the FLUXCOM approach. *Biogeosciences*, *17*, 1343–1365. https://doi.org/10.5194/bg-17-1343-2020

Jung, M., & Zscheischler, J. (2013). A guided hybrid genetic algorithm for feature selection with expensive cost functions. *Procedia Computer Science*, *18*, 2337–2346. https://doi.org/10.1016/j.procs.2013.05.405

Kondo, M., Ichii, K., Takagi, H., Sasakawa, M. (2015). Comparison of the data-driven top-down and bottom-up global terrestrial CO_2 exchanges: GOSAT CO_2 inversion and empirical eddy ux upscaling. *Journal of Geophysical Research: Biogeosciences*, *120*, 1226–1245. https://doi.org/10.1002/2014JG002866

Kondo, M., Saitoh, T. M., Sato, H., & Ichii, K. (2017). Comprehensive synthesis of spatial variability in carbon flux across monsoon Asian forests. *Agricultural and Forest Meteorology*, *232*, 623–634. http://dx.doi.org/10.1016/j.agrformet.2016.10.020

Korres, W., Koyama, C. N., Fiener, P., & Schneider, K. (2010). Analysis of surface soil moisture patterns in agricultural landscapes using Empirical Orthogonal Functions. *Hydrology and Earth System Sciences*, *14*, 751–764. https://doi.org/10.5194/hess-14-751-2010

Krawczyk, B. (2016). Learning from imbalanced data: Open challenges and future directions. *Progress in Artificial Intelligence*, *5*, 221–232. https://doi.org/10.1007/s13748-016-0094-0

Kuhn, M. (2020). *caret: Classification and regression training*. R package version 6.0-86. Retrieved from https://CRAN.R-project.org/package=caret

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, *2*(3), 18–22.Retrieved from https://www.stat.berkeley.edu/∼breiman/RandomForests/

Ließ, M., Glaser, B., & Huwe, B. (2012). Uncertainty in the spatial prediction of soil texture: Comparison of regression tree and random forest models. *Geoderma*, *170*, 70–79. https://doi.org/10.1016/j.geoderma.2011.10.010

Lindroth, A., Klemedtsson, L., Grelle, A., Weslien, P., & Langvall, O. (2008). Measurement of net ecosystem exchange, productivity and respiration in three spruce forests in Sweden shows unexpectedly large soil carbon losses. *Biogeochemistry*, *89*, 43–60. https://doi.org/10.1007/s10533-007-9137-8

Liu, H., & Motoda, H. (2008). Computational methods of feature selection. *Data Mining and Knowledge Discovery Series*. Boca Raton, FL: Chapman & Hall/CRC.

Löpmeier, F.-J. (1994). Berechnung der Bodenfeuchte und Verdunstung mittels agrarmeteorologischer Modelle. *Zeitschrift für Bewässerungswirtschaft*, *29*, 157–167.

Lussem, U., & Herbrecht, M. (2019). Land use classification of 2018 for the Rur catchment. *TR32DB*. https://doi.org/10.5880/TR32DB.38

Luyssaert, S., Inglima, I., Jung, M., Richardson, A. D., Reichstein, M., Papale, D., et al. (2007). CO_2 balance of boreal, temperate, and tropical forests derived from a global database. *Global Change Biology*, *13*(12), 2509–2537. https://doi.org/10.1111/j.1365-2486.2007.01439

Mauder, M., Cuntz, M., Drüe, C., Graf, A., Rebmann, C., Schmid, H. P., et al. (2013). A strategy for quality and uncertainty assessment of long-term eddy-covariance measurements. *Agricultural and Forest Meteorology*, *169*, 122–135. https://doi.org/10.1016/j.agrformet.2012.09.006

Mauder, M., & Foken, T. (2011). Documentation and instruction manual of the eddy-covariance software package TK3. Documentation and instruction manual of the eddy-covariance software package TK3 (46). Arbeitsergebnisse: Universität Bayreuth, Abteilung Mikrometeorologie. ISSN 1614-8924 60 pp.

Meyer, H. (2018). *CAST: "caret" Applications for spatial-temporal models*. R package version 0.3.1. Retrieved from https://CRAN.R-project.org/package=CAST

Meyer, H., Reudenbach, C., Hengl, T., Katurji, M., & Nauss, T. (2018). Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation. *Environmental Modelling & Software*, *101*, 1–9. https://doi.org/10.1016/j.envsoft.2017.12.001

Meyer, H., Reudenbach, C., Wöllauer, S., & Nauss, T. (2019). Importance of spatial predictor variable selection in machine learning applications – Moving from data reproduction to spatial prediction. *Ecological Modelling*, *411*, 108815. https://doi.org/10.1016/j.ecolmodel.2019.108815

Myneni, R., Knyazikhin, Y., & Park, T. (2015). MCD15A2H MODIS/Terra+Aqua Leaf Area Index/FPAR 8-day L4 Global 500m SIN Grid V006 [Data set]. *NASA EOSDIS Land Processes DAAC*. https://doi.org/10.5067/MODIS/MCD15A2H.006

Ney, P., Graf, A., Bogena, H., Diekkruger, B., Drüe, C., Esser, O., et al. (2019). CO_2 fluxes before and after partial deforestation of a Central European spruce forest. *Agricultural and Forest Meteorology*, *274*, 61–74. https://doi.org/10.1016/j.agrformet.2019.04.009

Papale, D., Black, T. A., Carvalhais, N., Cescatti, A., Chen, J., Jung, M., et al. (2015). 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. https://doi.org/10.1002/2015JG002997

Pastorello, G., Trotta, C., Canfora, E., Chu, H., Christianson, D., Cheah, Y.-W., et al. (2020). The FLUXNET2015 data set and the ONEFlux processing pipeline for eddy covariance data. *Scientific Data*, *7*, 225. https://doi.org/10.1038/s41597-020-0534-3

Pfeifroth, U., Kothe, S., Trentmann, J., Hollmann, R., Fuchs, P., Kaiser, J., & Werscheck, M. (2019). *Surface radiation data set - Heliosat (SARAH) - Edition 2.1. Satellite application facility on climate monitoring*. https://doi.org/10.5676/EUM_SAF_CM/SARAH/V002_01

Post, H., Hendricks Franssen, H. J., Graf, A., Schmidt, M., & Vereecken, H. (2015). Uncertainty analysis of eddy covariance CO_2 flux measurements for different EC tower distances using an extended two-tower approach. *Biogeosciences*, *12*, 1205–1221. https://doi.org/10.5194/bg-12-1205-2015

Post, H., Hendricks Franssen, H.-J., Han, X., Baatz, R., Montzka, C., Schmidt, M., & Vereecken, H. (2018). Evaluation and uncertainty analysis of regional-scale CLM4.5 net carbon flux estimates. *Biogeosciences*, *15*, 187–208. https://doi.org/10.5194/bg-15-187-2018

Prasad, A. M., Iverson, L. R., & Liaw, A. (2006). Newer classification and regression tree techniques: Bagging and random forests for ecological prediction. *Ecosystems*, *9*, 191–199. https://doi.org/10.1007/s10021-005-0054-1

Richardson, A. D., Aubinet, M., Barr, A. G., Hollinger, D. Y., Ibrom, A., & Lasslop, G. (2012). Uncertainty quantification. In M. Aubinet, T. Vesala, & D. Papale (Eds.), *Eddy covariance. A practical guide to measurement and data analysis, Springer Atmospheric Sciences* (pp. 173–209). New York: Springer Dordrecht Heidelberg London. https://doi.org/10.1007/978-94-007-2351-1

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., et al. (2016). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, *40*, 913–929. https://doi.org/10.1111/ecog.02881

Safa, B., Arkebauer, T. J., Zhu, Q., Suyker, A., & Irmak, S. (2019). Net Ecosystem Exchange (NEE) simulation in maize using artificial neural networks. *IFAC Journal of Systems and Control*, *7*, 100036. https://doi.org/10.1016/j.ifacsc.2019.100036

Schimel, D. S., House, J. I., Hibbard, K. A., Bousquet, P., Ciais, P., Peylin, P. H., et al. (2001). Recent patterns and mechanisms of carbon exchange by terrestrial ecosystems. *Nature*, *414*, 169–172.

Schmidt, M. (2014). Post processed fluxes of the roving EC station RUEC004 in Niederzier. *CRC/TR32 Database (TR32DB), TEODOOR Portal*. Retrieved from https://teodoor.icg.kfa-juelich.de/ibg3searchportal/dispatch?stations=http%3A%2F%2Fibg3wradar.ibg.kfa-juelich.de%3A8080%2Feifelrur_public%2Fsos,RU_EC_004on2020-04-14

Schmidt, M. (2019). *TERENO data from station(s) RU_EC_002 with parameter(s) AirFlux for time period 2012-08-17 to 2012-10-25*. Retrieved from https://hdl.handle.net/20.500.11952/TERENO.RU_EC_002.1573739908305

Schmidt, M., Reichenau, T. G., Fiener, P., & Schneider, K. (2012). The carbon budget of a winter wheat field: An eddy covariance analysis of seasonal and inter-annual variability. *Agricultural and Forest Meteorology*, *165*, 114–126. https://doi.org/10.1016/j.agrformet.2012.05.012

Schmidt, M., & Schween, J. (2018). Post processed fluxes (2017) of the permanent EC station RUEC001 near Selhausen. *CRC/TR32 Database (TR32DB), TEODOOR Portal*. Retrieved from http://teodoor.icg.kfa-juelich.de/ibg3searchportal/dispatch?stations=http%3A%2F%2Fibg3wradar.ibg.kfa-juelich.de%3A8080%2Feifelrur_public%2Fsos,RU_EC_001on2020-04-14

Schmitt, M., Bahn, M., Wohlfahrt, G., Tappeiner, U., & Cernusca, A. (2010). Land use affects the net ecosystem CO_2 exchange and its components in mountain grasslands. *Biogeosciences*, *7*, 2297–2309. https://doi.org/10.5194/bg-7-2297-2010

Schulzweida, U. (2019). *CDO user guide (Version 1.9.8)*. http://doi.org/10.5281/zenodo.3539275

Sun, G., Caldwell, P., Noormets, A., McNulty, S. G., Cohen, E., Moore Myers, J., Domec, J. C., et al. (2011). Upscaling key ecosystem functions across the conterminous United States by a water-centric ecosystem model. *Journal of Geophysical Research*, *116*, G00J05. https://doi.org/10.1029/2010JG001573

Tammi, I., Mustajärvi, K., & Rasinmäki, J. (2017). Integrating spatial valuation of ecosystem services into regional planning and development. *Ecosystem Services*, *26*(B), 329–344. https://doi.org/10.1016/j.ecoser.2016.11.008

Torgo, L., Branco, P., Ribeiro, R. P., & Pfahringer, B. (2013). Resampling strategies for regression. *Expert Systems*, *32*, 465–476. https://doi.org/10.1111/exsy.12081

Tramontana, G., Jung, M., Schwalm, C. R., Ichii, K., Camps-Valls, G., Ráduly, B., et al. (2016). Predicting carbon dioxide and energy fluxes across global FLUXNET sites with regression algorithms. *Biogeosciences*, *13*, 4291–4313. https://doi.org/10.5194/bg-13-4291-2016

Vitale, L., Di Tommasi, P., D'Urso, & Magliulo, V. (2016). The response of ecosystem carbon fluxes to LAI and environmental drivers in a maize crop grown in two contrasting seasons. *International Journal of Biometeorology*, *60*, 411–420. https://doi.org/10.1007/s00484-015-1038-2

Wang, D. (2017). *MODIS/Terra+Aqua surface radiation Daily/3-Hour L3 Global 5km SIN Grid V006* [Data set]. NASA EOSDIS Land Processes DAAC. https://doi.org/10.5067/MODIS/MCD18A1.006

Wang, H., Jiang, F., Wang, J., Ju, W., & Chen, J. M. (2019). Terrestrial ecosystem carbon flux estimated using GOSAT and OCO-2 XCO_2 retrievals. *Atmospheric Chemistry and Physics*, *19*, 12067–12082. https://doi.org/10.5194/acp-19-12067-2019

Wiekenkamp, I., Huisman, J. A., Bogena, H. R., Graf, A., Lin, H. S., Drüe, C., & Vereecken, H. (2016). Changes in measured spatiotemporal patterns of hydrological response after partial deforestation in a headwater catchment. *Journal of Hydrology*, *542*, 648–661. https://doi.org/10.1016/j.jhydrol.2016.09.037

Wohlfahrt, G., Hammerle, A., Haslwanter, A., Bahn, M., Tappeiner, U., & Cernusca, A. (2008). Seasonal and inter-annual variability of the net ecosystem CO_2 exchange of a temperate mountain grassland: Effects of weather and management. *Journal of Geophysical Research*, *113*, D08110. https://doi.org/10.1029/2007JD009286

Wutzler, T., Lucas-Moffat, A., Migliavacca, M., Knauer, J., Sickel, K., Sigut, L., et al. (2018). Basic and extensible post-processing of eddy covariance flux data with REddyProc. *Biogeosciences*, *15*, 5015–5030. https://doi.org/10.5194/bg-15-5015-2018

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

Xiao, J., Davis, K. J., Urban, N. M., Keller, K., & Saliendra, N. Z. (2011). Upscaling carbon fluxes from towers to the regional scale: Influence of parameter variability and land cover representation on regional flux estimates. *Journal of Geophysical Research*, *116*, G00J06. https://doi.org/10.1029/2010JG001568

Xiao, J., Zhuang, Q., Baldocchi, D. D., Law, B. E., Richardson, A. D., Chen, J., et al. (2008). Estimation of net ecosystem carbon exchange for the conterminous United States by combining MODIS and AmeriFlux data. *Agricultural and Forest Meteorology*, *148*, 1827–1847. https://doi.org/10.1016/j.agrformet.2008.06.015

Xu, X., Yang, G., Tan, Y., Tang, X., Jiang, H., Sun, X., et al. (2017). Impacts of land use changes on net ecosystem production in the Taihu Lake Basin of China from 1985 to 2010. *Journal of Geophysical Research: Biogeosciences*, *122*, 690–707. https://doi.org/10.1002/2016JG003444

Yuan, W., Liu, S., Yu, G., Bonnefond, J. M., Chen, J., Davis, K., et al. (2010). Global estimates of evapotranspiration and gross primary production based on MODIS and global meteorology data. *Remote Sensing of Environment*, *114*, 1416–1431. https://doi.org/10.1016/j.rse.2010.01.022

Zacharias, S., Bogena, H., Samaniego, L., Mauder, M., Fuß, R., Pütz, T., et al. (2011). A network of terrestrial environmental observatories in Germany. *Vadose Zone Journal*, *10*, 955–973. https://doi.org/10.2136/vzj2010.0139

Zhang, L., Wylie, B. K., Ji, L., Gilmanov, T. G., Tieszen, L. L., & Howard, D. M. (2011). Upscaling carbon fluxes over the Great Plains grasslands: Sinks and sources. *Journal of Geophysical Research*, *116*, G00J03. https://doi.org/10.1029/2010JG001504

Zhou, Q., Fellows, A., Flerchinger, G. N., & Flores, A. N. (2019). Examining interactions between and among predictors of net ecosystem exchange: A machine learning approach in a semi-arid landscape. *Scientific Reports*, *9*, 2222. https://doi.org/10.1038/s41598-019-38639