Remote Sensing of Environment 247 (2020) 111933

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Mapping canopy nitrogen in European forests using remote sensing and environmental variables with the random forests method

Yasmina Loozen[a,b,*], Karin T. Rebel[a], Steven M. de Jong[b], Meng Lu[b], Scott V. Ollinger[c], Martin J. Wassen[a], Derek Karssenberg[b]

[a] Copernicus Institute of Sustainable Development, Environmental Sciences, Faculty of Geosciences, Utrecht University, Utrecht, the Netherlands
[b] Physical Geography, Faculty of Geosciences, Utrecht University, Utrecht, the Netherlands
[c] Earth Systems Research Center and Department of Natural Resources and the Environment, University of New Hampshire, Durham, NH, USA

ARTICLE INFO

Keywords:
Canopy nitrogen
Foliar nitrogen
Plant traits
ICP Forests
Remote sensing
MODIS
MERIS
Environmental predictors
Random forests
Vegetation indices

ABSTRACT

Canopy nitrogen (N) influences carbon (C) uptake by vegetation through its important role in photosynthetic enzymes. Global Vegetation Models (GVMs) predict C assimilation, but are limited by a lack spatial canopy N input. Mapping canopy N has been done in various ecosystems using remote sensing (RS) products, but has rarely considered environmental variables as additional predictors. Our research objective was to estimate spatial patterns of canopy N in European forests and to investigate the degree to which including environmental variables among the predictors would improve the models compared to using remotely sensed products alone. The environmental variables included were climate, soil properties, altitude, N deposition and land cover, while the remote sensing products were vegetation indices and NIR reflectance from MODIS and MERIS sensors, the MOD13Q1 and MTCI products, respectively. The results showed that canopy N could be estimated both within and among forest types using the random forests technique and calibration data from ICP Forests with good accuracy (r^2 = 0.62, RRMSE = 0.18). The predicted spatial pattern shows higher canopy N in mid-western Europe and relatively lower values in both southern and northern Europe. For all subgroups tested (All plots, Evergreen Needleleaf Forest (ENF) plots and Deciduous Broadleaf Forest (DBF) plots), including environmental variables improved the predictions. Including environmental variables was especially important for the DBF plots, as the prediction model based on remotely sensed data products predicted canopy N with the lowest accuracy.

1. Introduction

In recent years, mapping canopy nitrogen (N), defined here as the N concentration in plant foliage (g N / 100 g dry matter, %N), has been studied at different scales and in a variety of natural environments (Martin et al., 2008; Ollinger et al., 2008; Ramoelo et al., 2012; Wang et al., 2016). This interest in canopy N can be attributed to the role N plays in physiological and ecosystem processes. N is an essential nutrient for plant growth (Zhao and Zeng, 2009). Leaf nitrogen concentration is linked to several leaf traits associated with plant photosynthesis (Hikosaka, 2004), i.e. photosynthetic capacity (Evans, 1989), light use efficiency (Kergoat et al., 2008), specific leaf area and leaf life span (Reich et al., 1999), as shown in the leaf economic spectrum (Wright et al., 2005; Wright et al., 2004) as well as whole-ecosystem net primary productivity (Reich, 2012).

Global vegetation models (GVMs) are designed to simulate ecosystem functioning and carbon (C) assimilation by terrestrial ecosystems. Several DGVMs explicitly include a representation of the N cycle, which allows them to analyze the influence of the N cycle on the terrestrial carbons sink (Xu-Ri and Prentice, 2008). Spatially explicit data about the N cycle are needed to validate these models. Canopy N mapping through remote sensing could be useful for this purpose.

Mapping canopy N using remote sensing evolved from benchtop studies aiming to identify specific wavelengths related to leaf N concentration using spectroradiometers (Kumar et al., 2006). The red-edge and near infra-red (NIR) have since then been identified as key spectral regions for canopy N estimation (Clevers and Gitelson, 2013; Li et al., 2014; Ollinger et al., 2008). The role of the red-edge region, located between 680 and 750 nm (Horler et al., 1983), for canopy N estimation is based on the link between foliar N and chlorophyll through the observed correlation between the red-edge region and leaf chlorophyll content (Clevers and Gitelson, 2013; Homolová et al., 2013; Horler et al., 1983; Kokaly et al., 2009; Schlemmer et al., 2013). The NIR

* Corresponding author at: Copernicus Institute of Sustainable Development, Environmental Sciences, Faculty of Geosciences, Utrecht University, Utrecht, the Netherlands.
E-mail address: y.m.a.loozen@uu.nl (Y. Loozen).

https://doi.org/10.1016/j.rse.2020.111933
Received 30 December 2019; Received in revised form 14 May 2020; Accepted 30 May 2020

spectral region was also identified to correlate with canopy N (Ollinger et al., 2008; Wang et al., 2016). This was observed in temperate and boreal North American forests where canopy N was correlated to both the NIR spectral region as well as NIR-based vegetation indices, including NDVI and EVI (Lepine et al., 2016; Ollinger et al., 2008). Similar relationships were also observed in a mixed European temperate forest (Wang et al., 2016). Although the exact mechanism behind the relationship between canopy N and the NIR reflectance is still unclear, it likely stems from associations between canopy N and the structural properties influencing the NIR scattering .

Among the existing techniques employed for canopy N mapping, creating and using vegetation indices (VIs) is a method that relies on a combination of several spectral regions or bands. Initially developed for crops and local scale applications (Chen et al., 2010; Clevers and Gitelson, 2013; Hansen and Schjoerring, 2003; Li et al., 2014; Mutanga et al., 2004; Schlemmer et al., 2013; Serrano et al., 2002; Tian et al., 2011), red-edge and, to a lesser extent, NIR-based VIs have been used for canopy N estimation at larger scales in various ecosystems. Several studies focused on grasslands and forest at local scale (Ling et al., 2014; Mirik et al., 2005; Wang et al., 2016) while other studies focused on regional areas such as savannah (Ramoelo et al., 2012) and Mediterranean forests (Loozen et al., 2018).

More recently, environmental variables were used together with remote sensing products to predict canopy N. This approach was suggested by McNeil et al. (2012) who observed an influence of N deposition on the spatial variability of leaf N concentration. Including environmental variables to predict canopy N is thus based on the fact that foliage biochemical concentration is influenced by several environmental factors. In particular, canopy N has been documented to be influenced by climate in Mediterranean forests (Sardans et al., 2011), in Europe (Sardans et al., 2015) and at the global scale (Reich and Oleksyn, 2004). Similarly, N deposition affects canopy N (McNeil et al., 2007, 2012; Sardans et al., 2016b; Sardans et al., 2015) as does plant functional type (PFT) (Han et al., 2011; Sardans et al., 2016a; Sardans et al., 2015). Soil properties, i.e. soil pH and nutrients, were also found to correlate with canopy N (Han et al., 2011). This approach, i.e. including environmental to predict canopy N using remote sensing, was implemented in a study mapping canopy N in savannah grass using red edge VIs as well as several environmental variables: soil, climate, geology and altitude (Ramoelo et al., 2012). In a recent study, Moreno-Martínez et al. (2018) used the random forests algorithm to map canopy N at global scale for several PFTs. As predictor variables, they used both bands and VI products from the MODIS sensor as well as environmental variables, i.e. bioclimatic variables, surface temperature and elevation.

The random forests algorithm used by Moreno-Martínez et al. (2018), is a machine learning technique based on regression trees which allows to model nonlinear relationships using several types of explanatory variables. It was found to be among the best techniques to predict foliar traits (Moreno-Martínez et al., 2018). Random forests have mainly been implemented in grasslands, at local (Adjorlolo et al., 2014; Mutanga et al., 2015) and regional scales (Ramoelo et al., 2015), but also in a coffee plantation (Chemura et al., 2018) and the miombo woodlands (Mutowo et al., 2018). These studies included either all reflectance bands available or several VIs as predictor variables for canopy N.

In this context, although several studies attempted to develop a methodology to map canopy N mapping over large spatial extents (Lepine et al., 2016; Martin et al., 2008; Moreno-Martínez et al., 2018), no study so far investigated the feasibility of mapping the spatial patterns of canopy N in European forests. In this study, our research objective was (*i*) to predict canopy N and its spatial pattern over European forests and (*ii*) to test whether including environmental variables as predictors improves canopy N predictions compared to approaches that rely on remotely sensed data alone. To do so, we mapped canopy N in European forests using the canopy N data from the ICP network as calibration data. We related canopy N plot data to the NDVI, EVI and NIR obtained from the MODIS MOD13Q1 product and the MTCI from the MERIS sensor, and environmental variables. The environmental variables included were elevation, climate, soil properties, N deposition and land-cover. We used the random forests machine learning technique to relate canopy N to the predictor variables. To evaluate the influence of including environmental variables on the results, we evaluated nine different random forests model settings: models using all predictor variables (*All pred*), using only remote sensing variables (*RS only*), and using only environmental variables (*Env only*). Each model was parameterized on three subgroups: all available plots (All plots), only Evergreen Needleleaf Forest (ENF) plots and only Deciduous Broadleaf Forest (DBF) plots. Including these three subgroups provided insights about the feasibility of mapping canopy N at European scale on all available plots, ENF plots and DBF plots. The results of the models were evaluated on each subgroup separately.

2. Material and methods

2.1. Canopy N data

2.1.1. ICP Forests

Canopy N data used in this analysis were obtained from the ICP Forests program (International Co –operative Program on Assessment and Monitoring of Air Pollution Effects on Forests, www.icp-forests.net). ICP Forests is a European biomonitoring network of forest conditions. The intensive monitoring program (level II network) includes more than 800 permanent forest plots sampled regularly across European countries. The forest plots are in homogeneous forest sites selected such that the diversity in European forests is represented. The forest plots have a minimum size of 0.25 ha, which corresponds to 56 m diameter for a circular plot (Ferreti et al., 2017). The foliar chemistry survey, including canopy N measurements, has been repeated every two years. The plots were sampled following a standard and consistent sampling design. Minimum five trees of each species belonging to the dominant class were selected. The sampling was repeated on the same sampled trees over the years. The leaves or needles were collected from the upper third part of the crown. If several species composed the dominant forest class, the foliar chemistry analysis was done separately for each species. Deciduous species plots were sampled during the second half of the growing season, before the onset of autumn, while evergreen plots were sampled during winter months, in the dormancy period. Quality control of the foliar concentration measurement was ensured by means of regular interlaboratory comparisons (Rautio et al., 2016).

2.1.2. Canopy N data analysis

Annual plot canopy N measurements data were obtained from the ICP Forests website for the period 1990–2014. Missing and duplicate entries as well as rare tree species, i.e. species that were sampled in less than six plot measurements, were excluded from the analysis. Canopy N outlier values, defined as those that were outside of the species-specific 5–95% percentile, were also removed from the dataset. 5207 annual plot measurements were left for analysis. Canopy N annual plot measurements were averaged by plot over all the sampling years to produce long-term averages of plot canopy N. This represented 818 plots, for which we obtained a long-term average canopy N value. Long-term averages canopy N will be called canopy N in the rest of the article. Plots were labelled according to their PFT. Plots with trees belonging to different PFTs were labelled as mixed PFT. Descriptive statistics of the canopy N data were performed.

2.2. Environmental variables

We chose to include as predictor variables the following environmental variables for their known influences on the N cycle and ecosystem properties in general.

Table 1
List of all the predictors variables, both environmental and remote sensing variables, included in the analysis. For the remote sensing variables MOD13Q1 and MTCI, there is a long-term monthly average variable for each month. There are not listed here for the sake of length.

Type of variable	Variable set	Variable name	Unit	Original spatial resolution	Source
Environmental	Bioclimatic variable	Annual mean temperature	°C	1 km	Fick and Hijmans, 2017
		Mean diurnal range (mean of monthly (max temp - min temp))	°C	1 km	Fick and Hijmans, 2017
		Temperature seasonality (standard deviation *100)	°C	1 km	Fick and Hijmans, 2017
		Max temperature of warmest month	°C	1 km	Fick and Hijmans, 2017
		Min temperature of coldest month	°C	1 km	Fick and Hijmans, 2017
		Temperature annual range	°C	1 km	Fick and Hijmans, 2017
		Isothermality (mean diurnal range /temp annual range)	%	1 km	Fick and Hijmans, 2017
		Mean temperature of wettest quarter	°C	1 km	Fick and Hijmans, 2017
		Mean temperature of driest quarter	°C	1 km	Fick and Hijmans, 2017
		Mean temperature of warmest quarter	°C	1 km	Fick and Hijmans, 2017
		mean temperature of coldest quarter	°C	1 km	Fick and Hijmans, 2017
		Annual precipitation	mm	1 km	Fick and Hijmans, 2017
		Precipitation of wettest month	mm	1 km	Fick and Hijmans, 2017
		Precipitation of driest month	mm	1 km	Fick and Hijmans, 2017
		Precipitation seasonality (coefficient of variation)	%	1 km	Fick and Hijmans, 2017
		Precipitation of wettest quarter	mm	1 km	Fick and Hijmans, 2017
		Precipitation driest quarter	mm	1 km	Fick and Hijmans, 2017
		Precipitation warmest quarter	mm	1 km	Fick and Hijmans, 2017
		Precipitation of coldest quarter	mm	1 km	Fick and Hijmans, 2017
	Altitude	EU DEM altitude	m	30 m	European Environment Agency, 2013
	Soil properties	Absolute depth to bedrock (BDTICM)	cm	250 m	Hengl et al., 2017
		Bulk density (fine earth, BLDFIE)	kg m^{-3}	250 m	Hengl et al., 2017
		Soil pH in water (PHIHOX)	pH	250 m	Hengl et al., 2017
		Cation exchange capacity of soil (CECSOL)	cmolc kg^{-1}	250 m	Hengl et al., 2017
		Coarse fragments volumetric (CRFVOL)	%	250 m	Hengl et al., 2017
		Sand content (50–2000 micro meter) mass fraction (SNDPPT)	%	250 m	Hengl et al., 2017
		Silt content (2–50 micro meter) mass fraction (SLTPPT)	%	250 m	Hengl et al., 2017
		Clay content (0–2 micro meter) mass fraction (CLYPPT)	%	250 m	Hengl et al., 2017
		Soil organic carbon stock (OCSTHA)	tons ha^{-1}	250 m	Hengl et al., 2017
		Soil organic carbon density (OCDENS)	kg m^{-3}	250 m	Hengl et al., 2017
		Soil organic carbon content (fine earth fraction)	g kg^{-1}	250 m	Hengl et al., 2017
	Nitrogen deposition	Oxidized nitrogen deposition (NOy)	g N m^{-2} yr^{-1}	50 km	Lamarque et al., 2013a
		Reduced nitrogen deposition (NHx)	g N m^{-2} yr^{-1}	50 km	Lamarque et al., 2013a
		Total nitrogen deposition	g N m^{-2} yr^{-1}	50 km	Lamarque et al., 2013a
	CCI Land cover	CCI land cover map (epoch 2010)	factorial	300 m	Defourny et al., 2016
Remote sensing	MOD13Q1	MODIS NDVI long-term monthly averages	/	250 m	Didan, 2015
		MODIS EVI long-term monthly averages	/	250 m	Didan, 2015
		MODIS NIR long-term monthly averages	/	250 m	Didan, 2015
	MTCI	MTCI long-term monthly averages	/	1 km	Dash and Curran, 2004

2.2.1. Bioclimatic variables

Climate was found to be related to canopy N (Reich and Oleksyn, 2004; Sardans et al., 2015; Sardans et al., 2011). Bioclimatic variables from the WorldClim2 dataset (Fick and Hijmans, 2017) were used. The bioclimatic variables were computed from monthly temperature and precipitation over the period 1972–2000. The bioclimatic variables consist in annual mean, seasonality, minimum or maximum values. The complete list of bioclimatic variables (19) included in the analysis is presented in Table 1. The initial spatial resolution was 1 km.

2.2.2. Altitude

We have decided to include elevation as a predictor variable because of both the correlation found between canopy N and temperature at global scale (Reich and Oleksyn, 2004) and the relationship between altitude and temperature. The digital elevation model over Europe (EU-DEM) was used for altitude data (European Environment Agency, 2013). The EU-DEM is a digital surface model based on both SRTM and ASTER GDEM as source data. The EU-DEM was produced using Copernicus data. The EU-DEM was obtained from the European Environmental agency website for the extent of Europe. The original spatial resolution was 30 m.

2.2.3. Soil properties

We chose to include soil properties as predictor variables because soil is an important component of the ecosystem that influences vegetation. More specifically, canopy N has been shown to be correlated to soil pH and soil mineral content (Han et al., 2011). Soil property maps were obtained from Soilgrids250m Global Soil Information (Hengl et al., 2017). The Soilgrids250m soil properties maps were predicted from a large collection of soil profile samples and globally available remote sensing products using machine learning prediction techniques. The list of Soilgrids250m variables included in this analysis (e.g. soil pH, soil cation exchange capacity, CEC, soil sand, clay and silt content, soil depth) is presented in Table 1. The original spatial resolution was 250 m.

2.2.4. N deposition

Canopy N has been found to be correlated with N deposition in various ecosystems including in *Pinus sylvestris* forests and mixed European forests (McNeil et al., 2007, 2012; Sardans et al., 2016b; Sardans et al., 2015). N deposition was included as predictor variable. The N deposition maps used in this analysis were aggregated from three atmospheric chemistry models (GISS-E2-R, CCSM-CAM3.5 and GFDL-AM3, Lamarque et al. (2013a)) within the Atmospheric Chemistry and Climate Model Intercomparison Project (ACCMIP, Lamarque et al. (2013b)). The maps were obtained from the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP) website (ISIMIP, 2019). We used reduced (NHx), oxidized (NOy) and total N deposition maps for the year 2006. The total N deposition was calculated as the sum of oxidized and reduced N depositions. The initial spatial resolution was 0.5 × 0.5 decimal degrees or approximately 50 km.

2.2.5. Land cover

We included the land cover as predictor variable in this analysis because the plant functional type has been shown to be related to leaf N (Kattge et al., 2011; Sardans et al., 2016a; Sardans et al., 2015). We used the land cover (LC) map from the ESA climate change initiative (CCI) over the epoch 2008–2012 (v.1.6.1, Defourny et al. (2016)). Following the UN land cover classification system, the land cover includes 22 land cover classes, which are compatible with the classification used in GVMs. We chose the ESA CCI LC from the epoch 2008–2012 because it is the period with the highest number of annual plot canopy N measurements. The LC spatial resolution was 300 m.

2.3. Remote sensing variables

2.3.1. MOD13Q1 product

MODIS data included in this analysis were NDVI and EVI VIs as well as NIR reflectance from the MOD13Q1 product (Didan, 2015). We chose to include these remote sensing products because the relationship between canopy N and NIR, either as a stand-alone reflectance product or included in the calculation of the NDVI and EVI, is well documented (Chemura et al., 2018; Mutowo et al., 2018; Ollinger et al., 2008). The MOD13Q1 product is available globally for every 16 days period at 250 m spatial resolution. One MODIS image for each 16 days period between the 1st January 2002 and the 31st December 2014 was obtained for each product considered (NIR, NDVI and EVI) as well as for the pixel reliability quality layer (QA). The MODIS images were downloaded from the AppEEARS website (AppEEARS Team, 2019) for the extent of Europe.

2.3.2. MTCI product

The MTCI was originally developed to monitor chlorophyll content in vegetation (Dash and Curran, 2004). It has been related to canopy N in various types of ecosystem and for various spatial extents, from local to regional studies (Cho et al., 2013; Loozen et al., 2018; Ramoelo et al., 2012; Tian et al., 2011). MTCI is a red-edge based VI that is computed using three MERIS sensor's reflectance bands located near the red-edge region (Eq. 1).

$$MTCI = \frac{R_{band10} - R_{band9}}{R_{band9} - R_{band8}} = \frac{R_{753.75} - R_{708.75}}{R_{708.75} - R_{681.25}} \quad (1)$$

The MTCI level 3 product is available almost globally as a monthly average at 1 km spatial resolution. The original reflectance data were provided by the European Space agency and were processed by Airbus Defense and space. The MTCI imagery is distributed by the Natural Environment Research Council (NERC) Earth Observation Data Centre (NEODC, 2015). One MTCI image was downloaded for the extent of Europe for each month between June 2002 and December 2011, except for October 2003, as no valid product was available.

2.4. Data preprocessing

Both MODIS and MTCI imagery were averaged by month to produce 12 long-term monthly averaged images for each remote sensing product considered. Before averaging, pixel-based quality information from the QA layers were applied on MODIS images. All pixels for which the quality value was not labelled as “good” were excluded from the computation of the long-term average. While there is no quality layer available for MTCI imagery, MTCI values lower than 1 are not valid. We ensured that no pixel with invalid values were included in the long-term MTCI monthly averages. However, by doing so, virtually no MTCI pixel were excluded from the calculation. The obtained long-term monthly averages of the MODIS products in January, February and December as well as the MTCI product in January and December contained a high number of pixels with missing values. These long-term monthly averages were excluded from the analysis.

Among the 22 LC classes initially present in the CCI LC variable, some were rarely represented in our study area. We grouped together similar LC classes and obtained 15 LC classes. Each of the obtained LC class was converted to a binary layer, in which the pixel values corresponded to the presence or absence of the specified LC class. All of the binary layers obtained were used as variables in the analysis.

All the predictor variables layers, including MODIS and MTCI long-term monthly averages, were resampled to a common grid and spatial resolution of 300 m using the bilinear interpolation of the resample function of the raster package in the R environment (Hijmans, 2018; R Core Team, 2019).

2.5. Random forests

Random forests is a machine learning method that is built on the classification and regression trees (CART) with the ensemble method. It was developed by Breiman (2001) and has been applied to map canopy N using VIs and other remote sensing products at different scales and in different ecosystems (Chemura et al., 2018; Moreno-Martínez et al., 2018; Mutowo et al., 2018; Ramoelo et al., 2015).

Random forests avoids overfitting by randomly sampling the predictor space. It is can model non-linear relationships without being constrained by the assumptions of the variable distributions and dependency. In a recent study, random forests was also found to give better leaf trait predictions compared to regularized linear regression, neural networks and kernel methods (Moreno-Martínez et al., 2018).

Random forests works by training many regression trees and reporting the mean response over all the trees. We implemented the random forests analysis in the R environment (R Core Team, 2019) using the randomForest package (Liaw and Wiener, 2002). The random forests algorithm is governed by three parameters, the number of trees (ntree), the number of sampled variables (mtry) and the minimum number of terminal seeds (nodesize). The random forests algorithm and the settings of the models are as follows:

- The regression tree is grown by iteratively splitting the bootstrap sample into two groups using the best predictor from a randomly selected subsample of all the available predictors. The mtry was set to one third of the total number of predictor variables.
- To build each regression tree, a bootstrap sample including two third of the training data is randomly selected. The remaining third of the training data (called the out-of-bag data (OOB)) is used to evaluate this specific tree.
- The tree is grown until the nodesize is reached. We set the node size to 5.
- This process repeated for ntree number of times. The ntree parameter was set to 2500 trees.

The random forests algorithm provides the mean square errors and r-square values assessed using the OOB samples, MSE_{OOB} and $r_{OOB}{}^2$, respectively.

We implemented the random forests model to predict canopy N in European forests. We fitted the random forests model to long-term average plot canopy N (section 2.1.2) using the predictor variables. We tested nine different models. A first type of model included all the predictor variables (*All pred*), a second type of model included only the remote sensing variables (*RS only*) and a third type of model included only the environmental variables (*Env only*). Each type of model was tested separately on three subgroups: All plots, only ENF plots and only DBF plots. We included these three subgroups in this study because each group provides different insight about canopy N spatial patterns and the feasibility of mapping canopy N at European scale. We did not develop a separate model for either EBF or mixed plots because of the restricted number of plots for these two PFTs (29 and 11, respectively).

The subsequent workflow was applied for all the models tested in this analysis:

- A first random forests model was fitted to the canopy N data. The predictor variables included in the model were selected using a recursive backward elimination (Brungard et al., 2015; Mutanga et al., 2015). The model was first fitted with all the predictors. The least important predictor was removed from the model. This process was repeated until only one predictor variable was left. The model selected was the one with lowest MSE_{OOB} value. $r_{OOB}{}^2$and $RRMSE_{OOB}$ are reported.
- We used 10-fold cross validation (C-Val) as independent validation to assess the accuracy of the selected model. The $r_{C-Val}{}^2$and $RRMSE_{C-Val}$ are reported. The C-Val is calculated using the R caret package (Kuhn, 2018) .

Table 2
Descriptive statistics of long-term average canopy nitrogen concentration (%N) plot data calculated over all forest plots (All plots) and grouped by plant functional type (PFT): Deciduous Broadleaf Forest (DBF), Evergreen Broadleaf Forest (EBF), Evergreen Needleleaf Forest (ENF) and mixed forest plots (mixed), with minimun (min), maximum (max), mean, and standard deviation (sd) values.

PFT	Number of plots (%)	Canopy N (%N)			
		min	max	mean	sd
All plots	818 (100%)	0.6	3.0	1.8	0.5
DBF	265 (32%)	1.5	3.0	2.4	0.2
EBF	29 (4%)	1.2	1.6	1.4	0.1
ENF	513 (63%)	0.6	2.2	1.4	0.2
mixed	11 (1%)	1.4	2.4	1.7	0.3

The coefficient of determination (r^2) is calculated as $1 - \frac{MSE}{Var(CN)}$ (Eq. 2), where MSE is the mean squared error, Var is the variance and CN, the canopy nitrogen (%N). The Relative Root Mean Squared Error (RRMSE) is calculated as $\sqrt{\frac{1}{n} \times \sum_{i=1}^{n} (P_i - O_i)^2} \times \frac{1}{\overline{o}_i}$ (Eq. 3), where $i = 1, 2, \ldots, n$ are distinct values, n is the total number of values, P_i is the predicted value, O_i is the observed value and $\overline{O}_i$ is the mean of all observed values.

The importance of the predictor variables is assessed by randomly permuting each predictor variable and calculating the subsequent decrease in OOB accuracy. The importance measure is expressed as the mean decrease in MSE_{OOB} (Liaw and Wiener, 2002).

2.6. Mapping canopy N

The best model for each subgroup (All plots, DBF and ENF), assessed using $r_{CV}{}^2$ and $RMSE_{CV}$, is used as predictive model to map canopy N in European forests.

3. Results

3.1. Descriptive analysis of canopy N plot data

Table 2 gives the canopy N plot data descriptive statistics. Among the 818 forest plots included in the analysis, the majority belongs to the ENF forest type (63%), while the second most occurring PFT, i.e. DBF, is present in 32% of the plots. Forest plots where several PFTs were sampled represent only 1% of the total. As expected, ENF plots have on average lower canopy N compared to DBF plots (1.4 and 2.4%N, respectively). The locations of the canopy N plot data are presented in Fig. 1. Higher canopy N values are observed in Midwestern Europe while lower canopy N values occur in the southern and northern part of the study region (Spain and south of Sweden, respectively, Fig. 1 - a). Regarding the PFTs, DBF and ENF occur in the whole study area, while EBF and mixed plots are only found in certain regions, i.e. in Southern Europe and in central Western Europe, respectively. The observed difference in canopy N values between ENF and DBF plots is reflected in the respective PFTs maps (Fig. 1 – b and c).

3.2. Results of the random forests analysis

Table 3 shows results from the random forests analysis. Among the nine models tested, the model including all predictor variables and all plots performed best, with a r^2 of 0.62 for the validation. The models including all predictors showed higher r^2 and lower RRMSE, both for OOB and validation, compared to models including either remote sensing or environmental variables. This was the case for all groups considered (All plots, ENF and DBF). When considering models for either

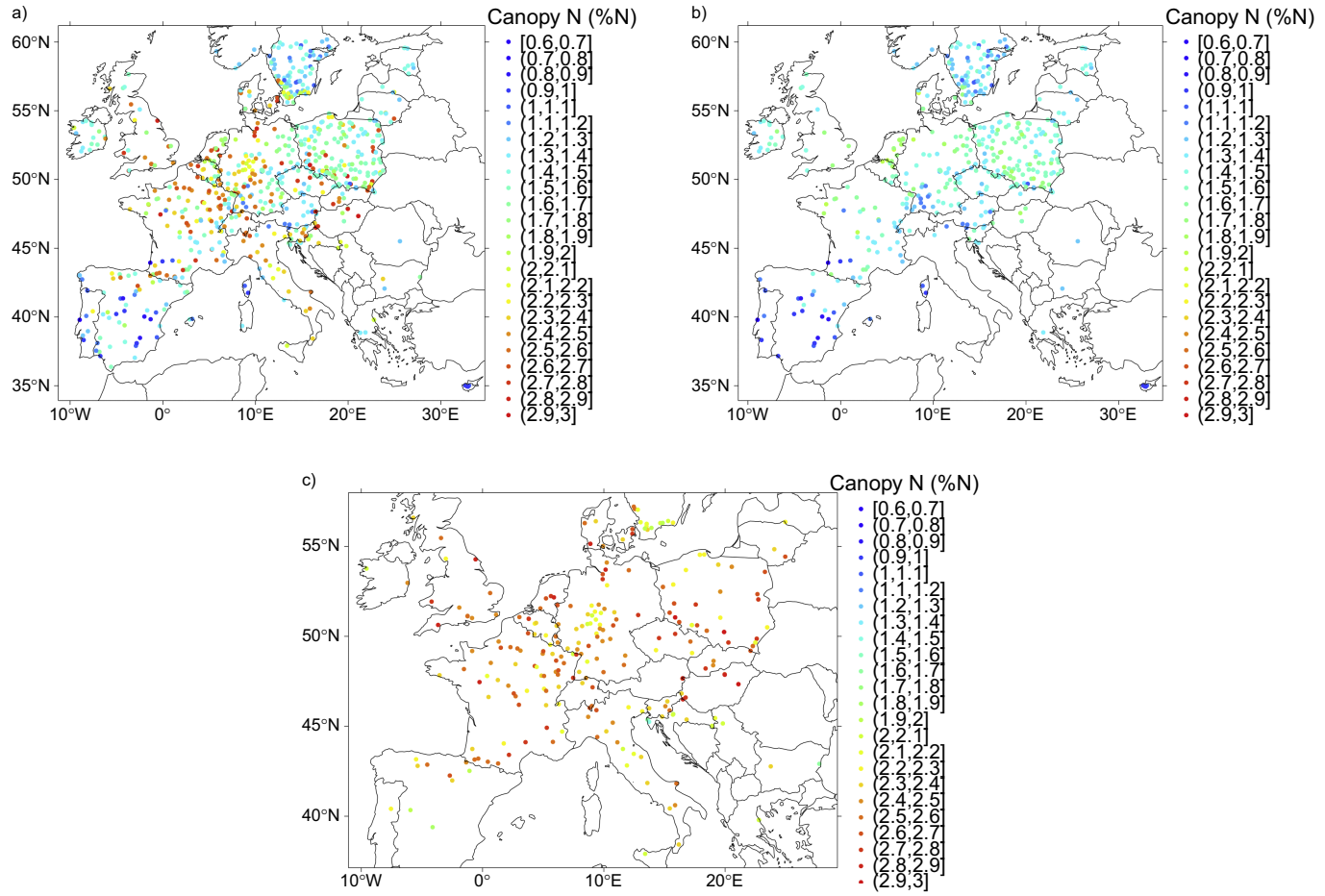

Fig. 1. Map of the forest plot locations for a) All plots, b) Evergreen Needleleaf Forest (ENF) plots and c) Deciduous Broadleaf Forest (DBF) plots. The color scale gives the long-term canopy nitrogen (%N).

Table 3
Results of the random forests analysis for the models including all predictor variables (All pred), only remote sensing variables (RS only) and only environmental variables (Env only), and for each plant functional type (PFT) group: all plots (All plots), Evergreen Needleleaf Forest (ENF) and Deciduous Broadleaf Forest (DBF). The initial number of predictors represent the number of predictor variables available to build the model before variables selection, the selected number of predictors represents the number of variables selected to build the model. r^2, Relative Root Mean Square error (RRMSE) are presented for both out-of-bag data (OOB) and cross-validation (C-Val).

PFT	Number of plots	Model	Initial number of predictors	Selected number of predictors	Calibration (OOB)		Validation (C-Val)	
					r^2	RRMSE	r^2	RRMSE
All plots	818	All pred	86	17	0.63	0.18	0.62	0.18
		RS only	37	16	0.61	0.18	0.60	0.18
		Env only	49	15	0.55	0.19	0.54	0.20
ENF	513	All pred	86	22	0.50	0.11	0.49	0.11
		RS only	37	9	0.45	0.12	0.44	0.12
		Env only	49	8	0.48	0.11	0.47	0.11
DBF	265	All pred	86	17	0.40	0.08	0.39	0.08
		RS only	37	17	0.12	0.09	0.09	0.09
		Env only	49	10	0.39	0.08	0.39	0.08

all plots or ENF plots, the type of model (*All pred*, *RS only* or *Env only*) did not have a strong effect on the models fit. For the DBF subgroup, on the contrary, the *RS only* type of model showed lower r^2 compared to the other two (r^2 = 0.09 and r^2 = 0.39, for *RS only* and both *All pred and Env only*, respectively). The *RS only* model was thus not able to predict canopy N for the DBF subgroup. If we compare the r^2 for different subgroups included in the analysis, the models with all plots performed always better ($r^2_{C\text{-}Val}$ 0.54–0.62) always performed better than those of the ENF ($r^2_{C\text{-}Val}$ 0.47–0.49) or the DBF ($r^2_{C\text{-}Val}$ 0.09–0.39) subgroups. However, the opposite is observed when comparing the RRMSE for three subgroups. The RRMSE for the DBF subgroup (RRMSE 0.08–0.09) is lower than the RRMSE of the ENF (RRMSE 0.11–0.12) or all plots (RRMSE 0.18–0.20) subgroups.

Scatterplots of predicted vs observed values for canopy N for the All plots, ENF plots and DBF plots are presented in Fig. 2. Regression lines were fitted between the predicted vs observed values for each group of predictor variables studied. Chow tests (Chow, 1960) were done to assess whether the sets of coefficients between different linear regressions were equal. The tests showed that the differences between each group-wise pairs of regression lines were not significant.

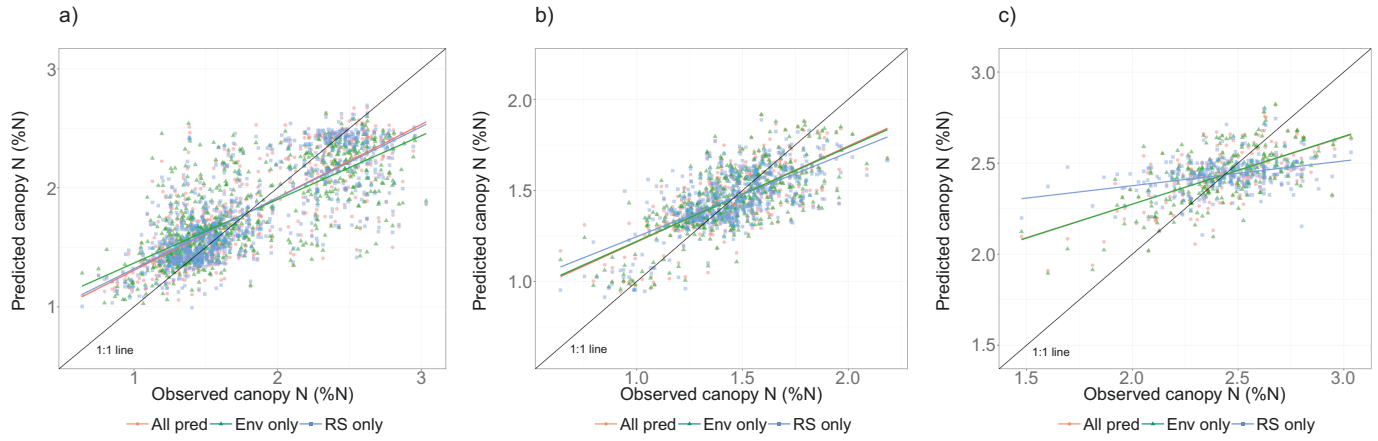

Fig. 2. Scatterplots between observed and predicted canopy N values (%N) based on OOB prediction for a) All plots, b) ENF plots and c) DBF plots. The color scale refers to the predictor variables tested: all predictor variables (*All pred*), only remote sensing variables (*RS only*) and only environmental variables (*Env only*).

3.3. Variable importance

The 10 most important variables for predicting canopy N are presented in Table 4 for all models considered. For the *All pred* model, when all plots were included, nine out of the 10 most important variables for predicting canopy N were remote sensing variables, with the two most important ones being EVI long-term average in May and June. These two variables are also the most important for canopy N prediction in the *All plots RS only* model. More generally, the important variables for the *All pred* model showed large agreement with the *RS only* model. In the *Env only* model, the two most important variables were the binary variables for presence or absence of broadleaf

Table 4
List of the 10 most important variables for canopy nitrogen prediction and importance values for all models tested: all predictor variables (All pred), only remote sensing variables (RS only) and only environmental variables (Env only), and for each plant functional type (PFT) group: All plots, Evergreen Needleleaf Forest (ENF) and Deciduous Broadleaf Forest (DBF).

All pred – All plots	Importance	Env only – All plots	Importance	RS only – All plots	Importance
MODIS EVI long-term monthly average May	6.9	CCI land cover broadleaf deciduous	7.9	MODIS EVI long-term monthly average May	6.7
MODIS EVI long-term monthly average June	5.7	CCI land cover needleleaf evergreen	7.4	MODIS EVI long-term monthly average June	5.8
MODIS EVI long-term monthly average July	4.3	Annual Mean Temperature	4.7	MTCI long-term monthly average February	5.7
MTCI long-term monthly average February	4.2	Oxidized nitrogen deposition	4.0	MODIS EVI long-term monthly average July	4.4
MODIS NDVI long-term monthly average June	3.3	Sand content mass fraction	3.7	MODIS EVI long-term monthly average August	3.6
MODIS NDVI long-term monthly average August	2.2	Mean temperature coldest quarter	3.7	MODIS NDVI long-term monthly average June	3.2
MODIS NDVI long-term monthly average November	2.2	Silt content mass fraction	3.2	MODIS NDVI long-term monthly average November	2.5
MTCI long-term monthly average June	2.1	Minimum temperature coldest month	3.0	MTCI long-term monthly average March	2.4
Oxidized nitrogen deposition	1.9	Mean temperature warmest quarter	2.7	MODIS NDVI long-term monthly average August	2.2
MODIS NDVI long-term monthly average March	1.9	Clay content mass fraction	2.6	MTCI long-term monthly average June	2.2

All pred - ENF plots	Importance	Env only - ENF plots	Importance	RS only - ENF plots	Importance
Sand content mass fraction	0.8	Annual mean temperature	1.4	MTCI long-term monthly average February	1.7
Annual mean temperature	0.7	Mean temperature warmest quarter	1.4	MTCI long-term monthly average September	1.2
Coarse fragments volumetric	0.7	Coarse fragments volumetric	1.2	MTCI long-term monthly average March	1.1
Mean temperature warmest quarter	0.6	Precipitation warmest quarter	1.0	MTCI long-term monthly average August	1.1
Precipitation warmest quarter	0.5	Mean temperature coldest quarter	1.0	MTCI long-term monthly average October	0.8
Mean temperature driest quarter	0.5	Silt content mass fraction	0.9	MODIS NDVI long-term monthly average June	0.7
Silt content mass fraction	0.5	Temperature seasonality	0.7	MTCI long-term monthly average May	0.6
Maximum temperature warmest month	0.3	Soil bulk density	0.6	MTCI long-term monthly average April	0.5
Oxidized nitrogen deposition	0.3	/	/	MODIS NIR long-term monthly average November	0.4
Mean temperature coldest quarter	0.3	/	/	/	/

All pred - DBF plots	Importance	Env only - DBF plots	Importance	RS only - DBF plots	Importance
Oxidized nitrogen deposition	1.1	Oxidized nitrogen deposition	1.6	MODIS NIR long-term monthly average September	0.8
Mean temperature warmest quarter	0.6	Mean temperature warmest quarter	0.8	MODIS EVI long-term monthly average October	0.8
Sand content mass fraction	0.5	Sand content mass fraction	0.7	MODIS EVI long-term monthly average June	0.7
Silt content mass fraction	0.5	Silt content mass fraction	0.7	MODIS NIR long-term monthly average July	0.7
Maximum temperature warmest month	0.5	Total nitrogen deposition	0.7	MODIS EVI long-term monthly average July	0.6
Annual mean temperature	0.5	Mean temperature driest quarter	0.6	MODIS NIR long-term monthly average October	0.6
Mean temperature driest quarter	0.4	Clay content mass fraction	0.6	MODIS EVI long-term monthly average August	0.5
Total nitrogen deposition	0.4	Soil bulk density	0.5	MTCI long-term monthly average June	0.5
Clay content mass fraction	0.4	Temperature seasonality	0.5	MODIS NDVI long-term monthly average August	0.5
MTCI long-term monthly average May	0.4	Altitude	0.5	MODIS NDVI long-term monthly average October	0.4

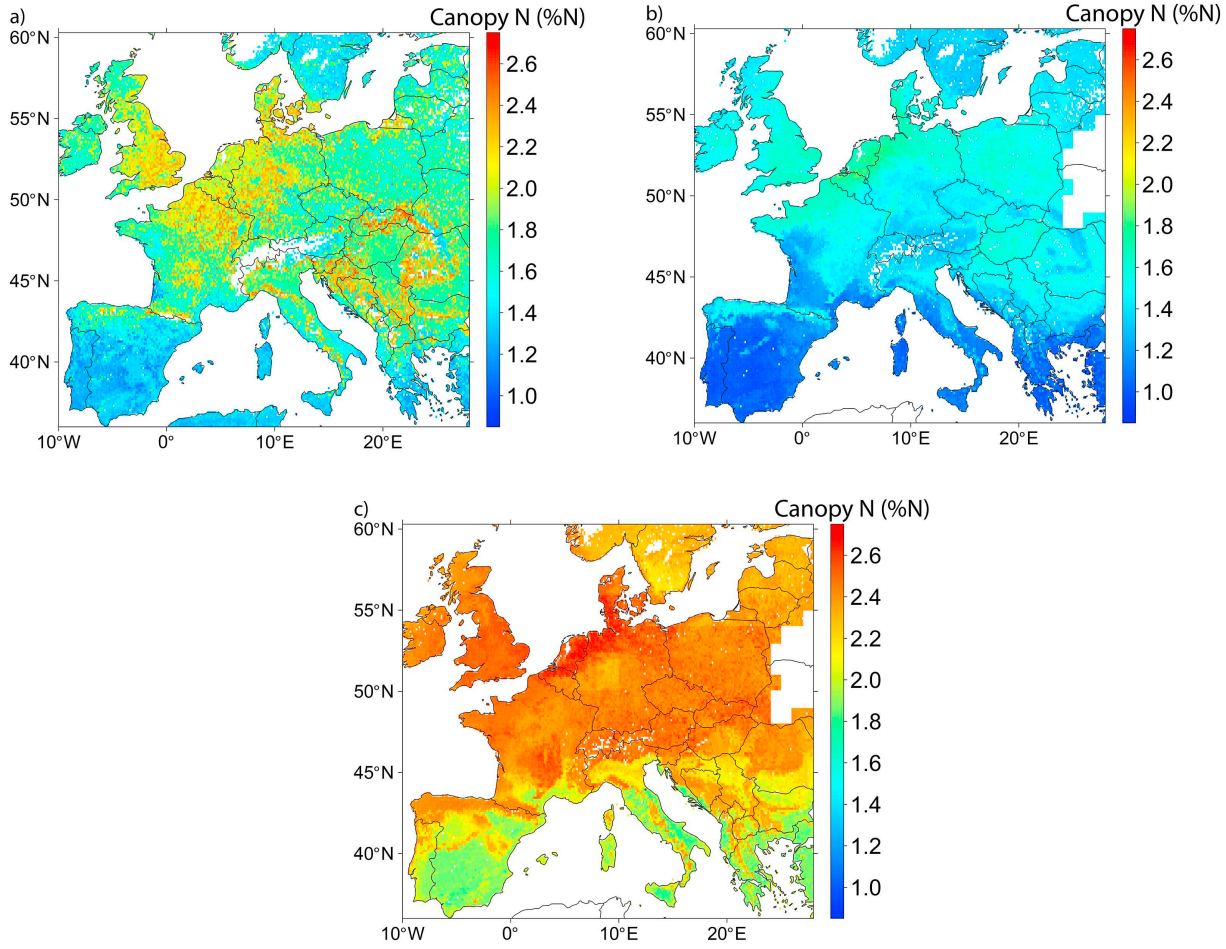

Fig. 3. Predicted canopy nitrogen maps (%N) calibrated using a) All plots, b) Evergreen Needleleaf Forests (ENF) and c) Deciduous Broadleaf Forest (DBF).

deciduous forest and needleleaf evergreen forest. These variables were obtained by modifying the CCI land cover map.

For the PFTs specific models, both *ENF and DBF All pred* models showed large similarities with their *Env only* model counterpart. For the *ENF All pred* model no RS variable was among the 10 most important predictor variables. When comparing *ENF and DBF All pred* models, soil properties and climate variables are important predictors of canopy N. For ENF plots, the soil sand content and the annual mean temperature were more important while for DBF plots, the oxidized nitrogen deposition as well as the mean temperature of the warmest quarter were important predictors of canopy N. For both DBF and ENF plots, the granulometry of the soil influenced the prediction of canopy N.

3.4. Canopy N map for European forests

The best predicted canopy N maps for each group considered, i.e. All plots, ENF plots and DBF plots, are presented in Fig. 3. The range of canopy N values of the predicted map corresponded to the range observed from forest samplings for each subgroup. The broad-scale spatial patterns show similarities between the three maps considered: in the southern and northern regions of Europe, i.e. the Mediterranean region and the south of Sweden, the predicted canopy N was lower than in the mid-western region of Europe. Local-scale patterns are also present. For example, in Netherlands and in the north west of Germany, the predicted canopy N is relatively higher than average for both the ENF and DBF maps. However, this is not observed when all plots are considered. The All plots model also predicts relatively higher values of canopy N, approximately between 2.4 and 2.6%N, in Eastern Europe and Slovakia, in particular. The observed fine-scale pattern corresponds to the location of the Carpathian Mountains. This pattern is also present, although less clearly, in the ENF map, but not in the DBF map.

4. Discussion

4.1. Canopy N spatial pattern

The aim of this study was to estimate spatial patterns in canopy N over European forests, which we were able to do with an r^2 for the validation of 0.62 for the All plots subgroup. The broad-scale spatial pattern (Fig. 3) of the predicted canopy N maps showed similarities between the three subgroups considered: lower canopy N concentration in the south and in the north of Europe, higher values in mid-western Europe. This pattern was also similar to that observed in the forest plot data (Fig. 1a), which is expected as this data was used to train the model. This indicates that the developed model was able to represent the broad-scale canopy N pattern present in the data. More specifically, however, the three predicted maps show finer scale variations. Both the DBF and the ENF predicted maps show relatively high values in the Netherlands and in the northwestern part of Germany. The forest plot data (Fig. 1a) also included several forest plots with relatively high long-term canopy N in this region. This trend, however, is not present in the predicted canopy N map for all plots. Another dissimilarity was noticeable in mid-southern part of France with higher canopy N values in both the All Plots and DBF predicted maps. This region corresponds to the location of the Massif Central mountain area and the canopy N values of forest plots located there were not higher than average.

4.2. Comparison with published studies

As mapping canopy N in forests has seldom been done at the European scale, there is not a large body of literature to compare our results with. A recent study published a forest leaf N map at global scale (Moreno-Martínez et al., 2018). When visually comparing this map with the results of the present study for the All plots model, the global scale map also presents relatively lower concentration in the south of Sweden, while showing relatively higher values in the center of Europe. However, the range of values is different between the two maps: while most values are between 1.4 and 2%N in the published study, in our analysis the range of values is larger, between 1.2 and 2.4%N. The accuracy for predicting canopy N (r^2 = 0.62 for the best model) was similar for both studies. The published study used the published data from the TRY database (Kattge et al., 2011) for calibration. Although the data in the TRY database are numerous, the data were sampled for various research purposes and the sampling methods do not follow a standard guideline. The ICP Forests data, on the contrary, were sampled following a consistent process, which ensures of the good quality of the dataset.

4.3. The role of environmental variables

The second aim of this study was to test whether including environmental variables as predictors improves canopy N predictions compared to approaches that rely on remotely sensed data alone. The results showed that including environmental variables as predictors improved the explanatory power of the models for all groups considered as the models for *All pred* always show higher r^2 and lower RRMSE compared to the *Env only* and *RS only* models (Table 3). In a previous study in savannah grass, including environmental variables in a stepwise multilinear regression also improved canopy N prediction compared to using VIs only (Ramoelo et al., 2012). It is interesting to note that the *RS only* model performed better than the *Env only* model for the all plots subgroup, while it was the opposite for both the ENF and DBF subgroups. This shows that, in our study, RS variables were useful to distinguish between different PFTS in the *All pred All plots* model. Moreover, the influence of including environmental variables can also be seen in the most important variables selected for the *All pred* models. For both the *ENF and DBF All pred* models, the most important variables are environmental variables, RS variables being among the least important predictors (Table 4). For the DBF subgroup, the *RS only* model showed the lowest observed accuracy of all models tested (r^2 = 0.09). The remote sensing variables were not able to predict canopy N and including environmental variables was essential to predict canopy N for DBF plots, in our dataset. Including environmental variables to predict canopy N was thus more beneficial for separate PFTs, and even more so for the DBF subgroup.

4.4. Variables importance

Regarding the difference in RS products selected, it is interesting to note that, although the NIR spectral region was shown to be important for canopy N prediction in previous studies (Moreno-Martínez et al., 2018; Mutowo et al., 2018; Ollinger et al., 2008), it was seldom selected as predictor variable in the models tested. It was among the most important variables for the *DBF Env only* model, which showed the lowest r^2 of all the models tested. On the contrary, although being tested in few studies for canopy N prediction (Lepine et al., 2016; Ramoelo et al., 2012; Wang et al., 2012), EVI was the most important variable for the *All pred and all plots* models. A remote sensing product derived from EVI, the maximum EVI, was also found to be the most important variable in a study mapping leaf N at global scale (Moreno-Martínez et al., 2018). In a recent study in the Miombo woodlands, the results showed that NIR VIs, among which NDVI and EVI, are complementary to the NIR spectral region for canopy N mapping (Mutowo et al., 2018). However, in our study, NDVI was not selected as most important variable for any of the developed models. The MTCI index, based on the relationship between canopy N and chlorophyll (Dash and Curran, 2004) was only selected as the most important variable for the *ENF RS only* model. The stronger relationship between canopy N and the NIR region compared to the MTCI, a red-edge based index, was previously observed as well. In a mixed temperate forest, Wang et al. (2016) showed that the relationship between canopy N and MTCI was weak. What is also surprising is that the RS products selected as important variables are from different months than the forest plots sampling months. For DBF plots, the two most important RS variables for the *DBF RS only* model were from the months September and October while the DBF plots were mainly sampled in July and August. For the ENF plots, it is not so clear as the RS products from winter months, during which the ENF plots are sampled, were excluded from the analysis.

Among the environmental predictors tested, oxidized N deposition was the most important variable for the DBF plots for both the *All pred and the Env only* model (Table 4). For the ENF subgroup, although previous studies showed that canopy N was correlated to N deposition in needleleaf forests (Fleischer et al., 2013; Sardans et al., 2016b), N deposition was only selected among the least important predictors for the *All pred* model and it was not selected at all for the *Env only* model. This might be related to an observed stronger response of deciduous species to N deposition compared to coniferous species (Crowley et al., 2012). The annual mean temperature was among the most important variables for predicting canopy N for the ENF subgroup for both the *All pred and the Env only* models. Mean annual temperature was also among the selected predictor variables for the *All pred and all plots* model. This is consistent with previous findings showing an influence of mean annual temperature on leaf N (Reich and Oleksyn, 2004; Sardans et al., 2015).

For the *Env only* models including all plots, the land cover binary variables indicating the presence of broadleaf deciduous forest and the presence of needleleaf evergreen forest were the most important variables. This is not surprising, as forest type is well known to influence foliage N (Sardans et al., 2016a; Sardans et al., 2015; Sardans et al., 2011). In our study too, the DBF and ENF plots also show different mean canopy N concentration, with 1,4%N and 2,4%N for ENF and BDF plots, respectively (Table 2).

4.5. Source of errors

When we look critically at the results obtained, it is interesting to note that the r^2 values from the 10 fold cross-validation (r_{CV}^2) are close to the r_{OOB}^2 values. This shows that the developed models were robust for validation. The r^2 obtained for the ENF and DBF models are lower than the value obtained for the *All pred* models. This could be related to the smaller number of plots used to calibrate the ENF and DBF models, n = 513 and n = 265, respectively, compared to the *All pred* model, n = 818. The limited range of canopy N values for the ENF and DBF models, might also explain the observed decrease in model fit.

4.6. Future perspectives

In this study, we showed that combining vegetation indices with environmental variables can contribute to canopy N mapping over large spatial extents. Although, as showed by the comparison and differences with a recent published study (section 4.2), this still needs to be further developed, this study contributes to the discussion about the feasibility of canopy N mapping over large spatial extents. The resulting canopy N map could provide spatial indicators of canopy N in European forests.

In this analysis, we worked with MODIS and MERIS remote sensing data to insure that the period during which the remote sensing data were acquired was overlapping with the sampling period of the forest plots. Although more recent satellite sensors such as Sentinel 2, Sentinel 3 or RapidEye have either higher spatial or spectral resolutions that

would probably improve the accuracy of the obtained canopy N map, the time series were too short to allow for a long enough overlapping period with canopy N sampling data. As the ICP Forest monitoring network is an ongoing project, a future perspective of this study would be to compare the results with a similar analysis, including more recent satellite sensors once the satellites' data time series are long enough.

We focused on canopy N in European forests. While this is a common land use type across Europe (42% of total land area), it would be valuable to further develop this analysis by including other natural PFTs (like grasses) and non-natural land use types like agricultural land. However, the ICP Forests database we used for this analysis was very valuable, and these high quality long-term data are not yet available for all land use types. The number of sites, but also the consistency in the way the forest plots are sampled and the %N are measured in ICP Forests is unique. We would like to emphasize that this would be an important necessary step to extrapolate to other land use types.

Finally, an envisioned result of this project is to improve GVMs by providing large-scale information about canopy N and its spatial pattern. In the future, we therefore foresee to compare our results with canopy N modelled with GVMs. However, canopy N values are not static over time, and in our study we averaged the canopy plot data as well as the RS data over a long time period. This gave us more data to work with, as for each year much less data was available. However, if we want to optimize GVMs using the predicted maps, including temporal variations, e.g. yearly or bi-yearly data, would make the predicted maps more compatible with GVMs output. Another future development of the present study is thus to include temporal changes in canopy N values.

5. Conclusion

In this study, our objective was to characterize spatial patterns of canopy N in forests across Europe. We showed that we could map canopy N using the random forests technique and calibration data from ICP Forests with good accuracy (r^2 = 0.62, RRMSE = 0.18, for validation). Among the RS products tested (EVI, MTCI, NIR and NDVI), EVI was the most important predictor for canopy N prediction when all plots were included, while MTCI the most important predictor for the *ENF RS only* model. We also investigated whether including environmental variables as predictors would improve the prediction models. For all subgroups tested (All plots, ENF plots and DBF plots), including environmental variables improved the predictions. Moreover, in our dataset, including environmental variables was especially essential for the DBF plots, as the prediction model based on remotely sensed data products only was not able to predict canopy N with sufficient accuracy. Finally, including environmental variables together with RS products to predict canopy N showed promising results and could be tested in other regions and with different land use types. A future outcome of this analysis is to compare the predicted canopy N map to GVMs outputs.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the Netherlands Organization for Scientific Research (NWO) [NWO ALW-GO-AO/14-12]. We would like to acknowledge Álvaro Moreno-Martínez for sharing his high resolution leaf nitrogen map with us as well as Gerard Heuvelink for his help with the random forests model.

References

AppEEARS Team, 2019. Application for Extracting and Exploring Analysis Ready Samples (AppEEARS). In: USGS/Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota, USA.: NASA EOSDIS Land Processes Distributed Active Archive Center (LP DAAC).

European Environment Agency, 2013. Digital Elevation Model over Europe (EU-DEM). www.eea.europa.eu/data-and-maps/data/eu-dem.

Adjorlolo, C., Mutanga, O., Cho, M.A., 2014. Estimation of canopy nitrogen concentration across c3 and c4 grasslands using worldview-2 multispectral data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 7, 4385–4392.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Brungard, C.W., Boettinger, J.L., Duniway, M.C., Wills, S.A., Edwards, T.C., 2015. Machine learning for predicting soil classes in three semi-arid landscapes. Geoderma 239-240, 68–83.

Chemura, A., Mutanga, O., Odindi, J., Kutywayo, D., 2018. Mapping spatial variability of foliar nitrogen in coffee (Coffea arabica L.) plantations with multispectral Sentinel-2 MSI data. ISPRS J. Photogramm. Remote Sens. 138, 1–11.

Chen, P., Haboudane, D., Tremblay, N., Wang, J., Vigneault, P., Li, B., 2010. New spectral indicator assessing the efficiency of crop nitrogen treatment in corn and wheat. Remote Sens. Environ. 114, 1987–1997.

Cho, M.A., Ramoelo, A., Debba, P., Mutanga, O., Mathieu, R., van Deventer, H., Ndlovu, N., 2013. Assessing the effects of subtropical forest fragmentation on leaf nitrogen distribution using remote sensing data. Landsc. Ecol. 28, 1479–1491.

Chow, G.C., 1960. Tests of equality between sets of coefficients in two linear regressions. Econometrica 28, 591–605.

Clevers, J.G.P.W., Gitelson, A.A., 2013. Remote estimation of crop and grass chlorophyll and nitrogen content using red-edge bands on sentinel-2 and-3. Int. J. Appl. Earth Obs. Geoinf. 23, 344–351.

Core Team, R., 2019. R: A Language and Environment for Statistical Computing. In. R Foundation for Statistical Computing, Vienna, Austria.

Crowley, K.F., McNeil, B.E., Lovett, G.M., Canham, C.D., Driscoll, C.T., Rustad, L.E., Denny, E., Hallett, R.A., Arthur, M.A., Boggs, J.L., Goodale, C.L., Kahl, J.S., McNulty, S.G., Ollinger, S.V., Pardo, L.H., Schaberg, P.G., Stoddard, J.L., Weand, M.P., Weathers, K.C., 2012. Do nutrient limitation patterns shift from nitrogen toward phosphorus with increasing nitrogen deposition across the northeastern United States? Ecosystems 15, 940–957.

Dash, J., Curran, P.J., 2004. The MERIS terrestrial chlorophyll index. Int. J. Remote Sens. 25, 5403–5413.

Defourny, P., Kirches, G., Brockmann, C., Boettcher, M., Peters, M., Bontemps, S., Lamarche, C., Schlerf, M., Santoro, M., 2016. Land Cover CCI Product User Guide Version 2. UCL, Louvain-La-Neuve, pp. 91.

Didan, K., 2015. MOD13Q1 MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid V006 In NASA EOSDIS Land Processes DAAC (Ed.). USGS Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota. https://doi.org/10.5067/MODIS/MOD13Q1.006.

Evans, J.R., 1989. Photosynthesis and nitrogen relationships in leaves of C3 plants. Oecologia 78, 9–19.

Ferreti, M., Fischer, R., Mues, V., Granke, O., Lorenz, M., Seidling, W., 2017. Part II: Basic design principles for the ICP Forests Monitoring Networks. In: UNECE ICP Forests Programme Co-ordinating Centre (Ed.), *Manual on methods and criteria for harmonized sampling, assessment, monitoring and analysis of the effects of air pollution on forests* (p. 21 p + Annex). Thünen Institute of Forest Ecosystems, Eberswalde, Germany.

Fick, S.E., Hijmans, R.J., 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. Int. J. Climatol. 37, 4302–4315.

Fleischer, K., Rebel, K.T., Van Der Molen, M.K., Erisman, J.W., Wassen, M.J., Van Loon, E.E., Montagnani, L., Gough, C.M., Herbst, M., Janssens, I.A., Gianelle, D., Dolman, A.J., 2013. The contribution of nitrogen deposition to the photosynthetic capacity of forests. Glob. Biogeochem. Cycles 27, 187–199.

Han, W.X., Fang, J.Y., Reich, P.B., Ian Woodward, F., Wang, Z.H., 2011. Biogeography and variability of eleven mineral elements in plant leaves across gradients of climate, soil and plant functional type in China. Ecol. Lett. 14, 788–796.

Hansen, P.M., Schjoerring, J.K., 2003. Reflectance measurement of canopy biomass and nitrogen status in wheat crops using normalized difference vegetation indices and partial least squares regression. Remote Sens. Environ. 86, 542–553.

Hengl, T., Mendes de Jesus, J., Heuvelink, G.B.M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M.N., Geng, X., Bauer-Marschallinger, B., Guevara, M.A., Vargas, R., MacMillan, R.A., Batjes, N.H., Leenaars, J.G.B., Ribeiro, E., Wheeler, I., Mantel, S., Kempen, B., 2017. SoilGrids250m: global gridded soil information based on machine learning. PLoS One 12, e0169748.

Hijmans, R.J., 2018. Raster: Geographic Data Analysis and Modeling. https://CRAN.R-project.org/package=raster.

Hikosaka, K., 2004. Interspecific difference in the photosynthesis-nitrogen relationship: patterns, physiological causes, and ecological importance. J. Plant Res. 117, 481–494.

Homolová, L., Malenovský, Z., Clevers, J.G.P.W., García-Santos, G., Schaepman, M.E., 2013. Review of optical-based remote sensing for plant trait mapping. Ecol. Complex. 15, 1–16.

Horler, D.N.H., Dockray, M., Barber, J., 1983. The red edge of plant leaf reflectance. Int. J. Remote Sens. 4, 273–288.

ISIMIP, 2019. Input Data Set: Nitrogen Deposition. https://www.isimip.org/gettingstarted/details/24/.

Kattge, J., Díaz, S., Lavorel, S., Prentice, I.C., Leadley, P., Bönisch, G., Garnier, E., Westoby, M., Reich, P.B., Wright, I.J., Cornelissen, J.H.C., Violle, C., Harrison, S.P., Van Bodegom, P.M., Reichstein, M., Enquist, B.J., Soudzilovskaia, N.A., Ackerly,

D.D., Anand, M., Atkin, O., Bahn, M., Baker, T.R., Baldocchi, D., Bekker, R., Blanco, C.C., Blonder, B., Bond, W.J., Bradstock, R., Bunker, D.E., Casanoves, F., Cavender-Bares, J., Chambers, J.Q., Chapin III, F.S., Chave, J., Coomes, D., Cornwell, W.K., Craine, J.M., Dobrin, B.H., Duarte, L., Durka, W., Elser, J., Esser, G., Estiarte, M., Fagan, W.F., Fang, J., Fernández-Méndez, F., Fidelis, A., Finegan, B., Flores, O., Ford, H., Frank, D., Freschet, G.T., Fyllas, N.M., Gallagher, R.V., Green, W.A., Gutierrez, A.G., Hickler, T., Higgins, S.I., Hodgson, J.G., Jalili, A., Jansen, S., Joly, C.A., Kerkhoff, A.J., Kirkup, D., Kitajima, K., Kleyer, M., Klotz, S., Knops, J.M.H., Kramer, K., Kühn, I., Kurokawa, H., Laughlin, D., Lee, T.D., Leishman, M., Lens, F., Lenz, T., Lewis, S.L., Lloyd, J., Llusià, J., Louault, F., MA, S., Mahecha, M.D., Manning, P., Massad, T., Medlyn, B.E., Messier, J., Moles, A.T., Müller, S.C., Nadrowski, K., Naeem, S., Niinemets, Ü., Nöllert, S., Nüske, A., Ogaya, R., Oleksyn, J., Onipchenko, V.G., Onoda, Y., Ordoñez, J., Overbeck, G., Ozinga, W.A., Patiño, S., Paula, S., Pausas, J.G., Peñuelas, J., Phillips, O.L., Pillar, V., Poorter, H., Poorter, L., Poschlod, P., Prinzing, A., Proulx, R., Rammig, A., Reinsch, S., Reu, B., Sack, L., Salgado-Negret, B., Sardans, J., Shiodera, S., Shipley, B., Siefert, A., Sosinski, E., Soussana, J.-F., Swaine, E., Swenson, N., Thompson, K., Thornton, P., Waldram, M., Weiher, E., White, M., White, S., Wright, S.J., Yguel, B., Zaehle, S., Zanne, A.E., Wirth, C., 2011. TRY – a global database of plant traits. Glob. Chang. Biol. 17, 2905–2935.

Kergoat, L., Lafont, S., Arneth, A., Le Dantec, V., Saugier, B., 2008. Nitrogen controls plant canopy light-use efficiency in temperate and boreal ecosystems. Journal of Geophysical Research: Biogeosciences 113.

Kokaly, R.F., Asner, G.P., Ollinger, S.V., Martin, M.E., Wessman, C.A., 2009. Characterizing canopy biochemistry from imaging spectroscopy and its application to ecosystem studies. Remote Sens. Environ. 113, S78–S91.

Kuhn, M., 2018. Package 'caret' Classification and Regression Training. https://CRAN.R-project.org/package=caret.

Kumar, L., Schmidt, K., Dury, S., Skidmore, A., 2006. Imaging spectrometry and vegetation science. In: Meer, F.D.V.D., de Jong, S.M. (Eds.), Imaging Spectrometry: Basic Principles and Prospective Applications. Springer Netherlands, Dordrecht, pp. 111–155.

Lamarque, J.F., Dentener, F., McConnell, J., Ro, C.U., Shaw, M., Vet, R., Bergmann, D., Cameron-Smith, P., Dalsoren, S., Doherty, R., Faluvegi, G., Ghan, S.J., Josse, B., Lee, Y.H., MacKenzie, I.A., Plummer, D., Shindell, D.T., Skeie, R.B., Stevenson, D.S., Strode, S., Zeng, G., Curran, M., Dahl-Jensen, D., Das, S., Fritzsche, D., Nolan, M., 2013a. Multi-model mean nitrogen and sulfur deposition from the atmospheric chemistry and climate model Intercomparison project (ACCMIP): evaluation of historical and projected future changes. Atmos. Chem. Phys. 13, 7997–8018.

Lamarque, J.F., Shindell, D.T., Josse, B., Young, P.J., Cionni, I., Eyring, V., Bergmann, D., Cameron-Smith, P., Collins, W.J., Doherty, R., Dalsoren, S., Faluvegi, G., Folberth, G., Ghan, S.J., Horowitz, L.W., Lee, Y.H., MacKenzie, I.A., Nagashima, T., Naik, V., Plummer, D., Righi, M., Rumbold, S.T., Schulz, M., Skeie, R.B., Stevenson, D.S., Strode, S., Sudo, K., Szopa, S., Voulgarakis, A., Zeng, G., 2013b. The atmospheric chemistry and climate model Intercomparison project (ACCMIP): overview and description of models, simulations and climate diagnostics. Geosci. Model Dev. 6, 179–206.

Lepine, L.C., Ollinger, S.V., Ouimette, A.P., Martin, M.E., 2016. Examining spectral reflectance features related to foliar nitrogen in forests: implications for broad-scale nitrogen mapping. Remote Sens. Environ. 173, 174–186.

Li, F., Miao, Y., Feng, G., Yuan, F., Yue, S., Gao, X., Liu, Y., Liu, B., Ustin, S.L., Chen, X., 2014. Improving estimation of summer maize nitrogen status with red edge-based spectral vegetation indices. Field Crop Res. 157, 111–123.

Liaw, A., Wiener, M., 2002. R News. 2. pp. 18–22.

Ling, B., Goodin, D.G., Mohler, R.L., Laws, A.N., Joern, A., 2014. Estimating canopy nitrogen content in a heterogeneous grassland with varying fire and grazing treatments: Konza Prairie, Kansas, USA. Remote Sens. 6, 4430–4453.

Loozen, Y., Rebel, K.T., Karssenberg, D., Wassen, M.J., Sardans, J., Peñuelas, J., De Jong, S.M., 2018. Remote sensing of canopy nitrogen at regional scale in Mediterranean forests using the spaceborne MERIS terrestrial chlorophyll index. Biogeosciences 15, 2723–2742.

Martin, M.E., Plourde, L.C., Ollinger, S.V., Smith, M.L., McNeil, B.E., 2008. A generalizable method for remote sensing of canopy nitrogen across a wide range of forest ecosystems. Remote Sens. Environ. 112, 3511–3519.

McNeil, B.E., Read, J.M., Driscoll, C.T., 2007. Foliar nitrogen responses to elevated atmospheric nitrogen deposition in nine temperate forest canopy species. Environ. Sci. Technol. 41, 5191–5197.

McNeil, B.E., Read, J.M., Driscoll, C.T., 2012. Foliar nitrogen responses to the environmental gradient matrix of the Adirondack Park, New York. Ann. Assoc. Am. Geogr. 102, 1–16.

Mirik, M., Norland, J.E., Crabtree, R.L., Biondini, M.E., 2005. Hyperspectral one-meter-resolution remote sensing in Yellowstone National Park, Wyoming: I. forage nutritional values. Rangeland Ecology and Management 58, 452–458.

Moreno-Martínez, Á., Camps-Valls, G., Kattge, J., Robinson, N., Reichstein, M., van Bodegom, P., Kramer, K., Cornelissen, J.H.C., Reich, P., Bahn, M., Niinemets, Ü., Peñuelas, J., Craine, J.M., Cerabolini, B.E.L., Minden, V., Laughlin, D.C., Sack, L., Allred, B., Baraloto, C., Byun, C., Soudzilovskaia, N.A., Running, S.W., 2018. A methodology to derive global maps of leaf traits using remote sensing and climate data. Remote Sens. Environ. 218, 69–88.

Mutanga, O., Skidmore, A.K., Prins, H.H.T., 2004. Predicting in situ pasture quality in the Kruger National Park, South Africa, using continuum-removed absorption features. Remote Sens. Environ. 89, 393–408.

Mutanga, O., Adam, E., Adjorloloa, C., Abdel-Rahmanw, E.M., 2015. Evaluating the robustness of models developed from field spectral data in predicting African grass foliar nitrogen concentration using WorldView-2 image as an independent test dataset. Int. J. Appl. Earth Obs. Geoinf. 34, 178–187.

Mutowo, G., Mutanga, O., Masocha, M., 2018. evaluating the applications of the near-infrared region in mapping foliar N in the Miombo Woodlands. Remote Sensing 10.

NEODC, 2015. NEODC - NERC Earth Observation Data Centre. Natural Environment Research Council. http://neodc.nerc.ac.uk/.

Ollinger, S.V., Richardson, A.D., Martin, M.E., Hollinger, D.Y., Frolking, S.E., Reich, P.B., Plourde, L.C., Katul, G.G., Munger, J.W., Oren, R., Smith, M.L., Paw, U., Bolsta, P.V., Cook, B.D., Day, M.C., Martin, T.A., Monson, R.K., Schmid, H.P., 2008. Canopy nitrogen, carbon assimilation, and albedo in temperate and boreal forests: Functional relations and potential climate feedbacks. Proceedings of the National Academy of Sciences of the United States of America 105, 19336–19341.

Ramoelo, A., Skidmore, A.K., Cho, M.A., Schlerf, M., Mathieu, R., Heitkönig, I.M.A., 2012. Regional estimation of savanna grass nitrogen using the red-edge band of the spaceborne rapideye sensor. Int. J. Appl. Earth Obs. Geoinf. 19, 151–162.

Ramoelo, A., Cho, M.A., Mathieu, R., Madonsela, S., van de Kerchove, R., Kaszta, Z., Wolff, E., 2015. Monitoring grass nutrients and biomass as indicators of rangeland quality and quantity using random forest modelling and WorldView-2 data. Int. J. Appl. Earth Obs. Geoinf. 43, 43–54.

Rautio, P., Fürst, A., Stefan, K., Raitio, H., Bartels, U., 2016. In: UNECE ICP Forests Programme co-Ordinating Centre (Ed.), Sampling and Analysis of Needles and Leaves. Manual on methods and criteria for harmonized sampling, assessment, monitoring and analysis of the effects of air pollution on forests (P. 19 P + Annex) Thünen Institute of Forest Ecosystems, Eberswalde, Germany.

Reich, P.B., 2012. Key canopy traits drive forest productivity. Proceedings of the Royal Society B: Biological Sciences *279*, 2128–2134.

Reich, P.B., Oleksyn, J., 2004. Global patterns of plant leaf N and P in relation to temperature and latitude. Proc. Natl. Acad. Sci. U. S. A. 101, 11001.

Reich, P.B., Ellsworth, D.S., Walters, M.B., Vose, J.M., Gresham, C., Volin, J.C., Bowman, W.D., 1999. Generality of leaf trait relationships: A test across six biomes. Ecology 80, 1955–1969.

Sardans, J., Rivas-Ubach, A., Peñuelas, J., 2011. Factors affecting nutrient concentration and stoichiometry of forest trees in Catalonia (NE Spain). For. Ecol. Manag. 262, 2024–2034.

Sardans, J., Janssens, I.A., Alonso, R., Veresoglou, S.D., Rillig, M.C., Sanders, T.G.M., Carnicer, J., Filella, I., Farré-Armengol, G., Peñuelas, J., 2015. Foliar elemental composition of European forest tree species associated with evolutionary traits and present environmental and competitive conditions. Glob. Ecol. Biogeogr. 24, 240–255.

Sardans, J., Alonso, R., Carnicer, J., Fernández-Martínez, M., Vivanco, M.G., Peñuelas, J., 2016a. Factors influencing the foliar elemental composition and stoichiometry in forest trees in Spain. Perspectives in Plant Ecology, Evolution and Systematics 18, 52–69.

Sardans, J., Alonso, R., Janssens, I.A., Carnicer, J., Vereseglou, S., Rillig, M.C., Fernández-Martínez, M., Sanders, T.G.M., Peñuelas, J., 2016b. Foliar and soil concentrations and stoichiometry of nitrogen and phosphorous across European Pinus sylvestris forests: relationships with climate, N deposition and tree growth. Funct. Ecol. 30, 676–689.

Schlemmer, M., Gitelson, A., Schepers, J., Ferguson, R., Peng, Y., Shanahan, J., Rundquist, D., 2013. Remote estimation of nitrogen and chlorophyll contents in maize at leaf and canopy levels. Int. J. Appl. Earth Obs. Geoinf. 25, 47–54.

Serrano, L., Peñuelas, J., Ustin, S.L., 2002. Remote sensing of nitrogen and lignin in Mediterranean vegetation from AVIRIS data: decomposing biochemical from structural signals. Remote Sens. Environ. 81, 355–364.

Tian, Y.C., Yao, X., Yang, J., Cao, W.X., Hannaway, D.B., Zhu, Y., 2011. Assessing newly developed and published vegetation indices for estimating rice leaf nitrogen concentration with ground- and space-based hyperspectral reflectance. Field Crop Res. 120, 299–310.

Wang, W., Yao, X., Yao, X., Tian, Y., Liu, X., Ni, J., Cao, W., Zhu, Y., 2012. Estimating leaf nitrogen concentration with three-band vegetation indices in rice and wheat. Field Crop Res. 129, 90–98.

Wang, Z., Wang, T., Darvishzadeh, R., Skidmore, A.K., Jones, S., Suarez, L., Woodgate, W., Heiden, U., Heurich, M., Hearne, J., 2016. Vegetation indices for mapping canopy foliar nitrogen in a mixed temperate forest. Remote Sensing 8.

Wright, I.J., Reich, P.B., Westoby, M., Ackerly, D.D., Baruch, Z., Bongers, F., Cavender-Bares, J., Chapin, T., Cornelissen, J.H.C., Diemer, M., Flexas, J., Garnier, E., Groom, P.K., Gulias, J., Hikosaka, K., Lamont, B.B., Lee, T., Lee, W., Lusk, C., Midgley, J.J., Navas, M.-L., Niinemets, U., Oleksyn, J., Osada, N., Poorter, H., Poot, P., Prior, L., Pyankov, V.I., Roumet, C., Thomas, S.C., Tjoelker, M.G., Veneklaas, E.J., Villar, R., 2004. The worldwide leaf economics spectrum. Nature 428, 821–827.

Wright, I.J., Reich, P.B., Cornelissen, J.H.C., Falster, D.S., Garnier, E., Hikosaka, K., Lamont, B.B., Lee, W., Oleksyn, J., Osada, N., Poorter, H., Villar, R., Warton, D.I., Westoby, M., 2005. Assessing the generality of global leaf trait relationships. New Phytol. 166, 485–496.

Xu-Ri, Prentice, I.C., 2008. Terrestrial nitrogen cycle simulation with a dynamic global vegetation model. Glob. Chang. Biol. 14, 1745–1764.

Zhao, Q., Zeng, D.H., 2009. Diagnosis methods of N and P limitation to tree growth: A review. Chinese Journal of Ecology 28, 122–128.