Ecological Indicators 144 (2022) 109446

Contents lists available at ScienceDirect

Ecological Indicators

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

Prediction of forest nutrient and moisture regimes from understory vegetation with random forest classification models

Lisein Jonathan [a,*], Fayolle Adeline [a], Legrain Andyne [a], Prévot Céline [b], Claessens Hugues [a]

[a] *Liège University – Faculty of Gembloux Agro-Bio Tech – Unit of Forest Resources Management, Belgium*
[b] *Forêt.Nature asbl, Belgium*

ARTICLE INFO

Keywords:
Floristic relevé
Bioindicator
Forest site
Nutrient regime
Moisture regime
Western Europe
Temperate forest
Ecological group
Forest management
Random forest classification
Ecogram matrix

ABSTRACT

The proper choice of the tree species to be grown in a specific forest site requires a good knowledge of the tree species autecology and a comprehensive description of the local environmental conditions. In Belgium (Western Europe), ecological forest site are classified according to three major gradients: climate, soil nutrient (fertility) and soil moisture regimes. Understory indicator species are used by practitioners to determine nutrient and moisture regimes, but requires a significant expertise of forest ecosystems. The present work aims in a first instance at modelling the nutrient and moisture regimes based on species composition. Secondly, a practical decision support tool is developped and made available in order to predict forest nutrient and moisture regime starting from a floristic relevé. To do so, we collected floristic relevés representing understory vegetation diversity in Belgium and covering all the nutrient and moisture gradient. The combination of soil and topographic measurements with the indicator plants presence/absence support forest scientists in inferring a nutrient and moisture regime to each relevé. The resulting dataset was balanced along the different nutrient or moisture regimes and Random Forest classification models were trained in order to predict the forest site characteristic from indicator species presence (or absence). One model was fitted for the prediction of the nutrient regime, exclusively based on the floristic information. A second one was trained to classify the moisture regime. Accurate predictions confirms the appropriate use of indicator species for the Belgian forest site classification. The two models are intregrated in a web application dedicated to forest practionners. This website enables the automatic determination of nutrient and moisture regimes from the species list of a floristic relevé.

1. Introduction

Indicator species are commonly used to characterise a forest site (climate, topography and soil) (Bartoli et al., 2000; Gégout et al., 2003; Gégout et al., 2005; Pinto et al., 2016), which is of great interest to forest practitioners, and for a long time, understory vegetation have played a fundamental role in many forest site classifications (Cajander, 1909; Braun-Blanquet, 1928; Wilson et al., 2001; Bergès et al., 2006). Recently, like almost all forest sciences, phytosociology have benefited from the widespread use of Geographical Information Systems (Biondi, 2011) and from the constitution of large vegetation plots databases (Douda et al., 2016; Chytrỳ et al., 2016). On the one hand, geographic layers describing the environmental conditions have supported the representation of forest site at the scale of entire regions (Van der Perre et al., 2015). On the other hand, applied phytosociology have been used to map forest habitats (Noirfalise, 1984) and to develop the Natura 2000 conservation network (european Habitats Directive 92/43/EEC) across Europe (Biondi, 2011). Besides, GIS analyses support the modelization of species distribution (Coudun et al., 2006; Coudun and Gégout, 2007), as well as floristic association distribution maps (Fayolle et al., 2014). Beyond the characterization of forest site with bioindicator, dynamic of the succession of vegetation association is also of great help for assessing the state of ancienticy, restauration or ecological integrity of forest ecosystem (Carignan and Villard, 2002; Baeten et al., 2009; Verstraeten et al., 2013). Traditionnaly, the autecology of indicator species is described through indicator value arbitrarily defined by experts (Diekmann, 2003). Ellenberg published in 1974 a famous list of indicator values (Ellenberg, 1974), which since has been extensively used and adapted in many temperate countries. In Ellenberg 's approach, environmental conditions are summarized by seven gradients, that include soil moisture and soil acidity, and every indicator species is assigned a value between one and nine representing the position of its ecological

* Corresponding author.
E-mail address: liseinjon@hotmail.com (L. Jonathan).

https://doi.org/10.1016/j.ecolind.2022.109446
Received 20 July 2022; Received in revised form 6 September 2022; Accepted 10 September 2022
Available online 20 September 2022
1470-160X/

optimum. Positionning a forest site on an environmental gradient is carried out rather simply by computing the average of all indicator values of the present species (Bartoli et al., 2000; Wilson et al., 2001). In Belgium, Noirfalise made its own list of indicator values (Noirfalise, 1984). In France, Rameau et al. (1989) later described the ecological niche with a two-dimentionnal graphical representation of the forest sites. Environmental conditions are summarized by a moisture-acidity matrix, the ecogram, in which the ecological niche is drawn for every single species. Different indicator species occupying the same niche on the ecogram are commonly grouped in *ecological groups* (Noirfalise, 1984; Rameau et al., 1989; Duvigneaud, 1946; Godart, 1989) which are more operationnal than species separately. In the Wallon region (southern Belgium), many dedicated decision-support tools are available to forest managers. Indeed, at the scale of the entire region, representing 17 000 square kilometers, the environmental conditions (climate, soil and topography) of forest site have been studied for a long time. Mapping of the soil properties has been carried out between 1950 and 1990. In average, two hand augering per hectare were performed, plus multiple additionnal field observations that were reported on 1:20 000 maps, which were subsequently digitalized (Legrain et al., 2014). Furthermore, the topography has been finely described by a digital terrain model issued from an aerial laser survey (LiDAR) realized in 2013 and 2014. Eventually, ten different bioclimatic areas which cover the 550 000 ha of forested land have been defined (Van der Perre et al., 2015). Forest scientists have thus a considerable documentation of forest environmental conditions at a regional scale, all being available in the form of geographical information layers. In parallel, forest practitioners have setted up a field-based methodology to determine the ability of forest tree species to form resilient and productive forest stands, while taking into account changes of both climate and society. A set of decision-making tools, named the *forest trees autecology guide* (Petit et al., 2017), have been developped starting from 1984 by a consortium of universities. This toolbox (fichierecologique.be), updated in 2017 to include ecosystem services and global warming risks, has been structured in two parts. On the one hand, forest tree species autecology has been gathered and synthetized in technical documents. On the other hand, a forest site classification has been proposed, along with identification keys that support the determination of any forest site based on field measurements (understory vegetation or soil information). This forest site classification consists in a systematic breakdown of the three predominant ecological factors that impact forest tree development: climate, nutrient and moisture gradients (see Pyatt, 1995; Wilson et al., 2001 for a similar work in Great Britain). Climate conditions are known through the ten bioclimatic areas of Wallonia (Van der Perre et al., 2015). The nutrient gradient, split in six regimes from hyperacidic to carbonated forest soils, and the moisture gradient, divided into ten regimes, are represented side by side in a two dimensionnal table, the ecogram matrix (Fig. 1). Each forest tree species autecology is representated in the ecogram by colored fitness classes reflecting the species equilibrium with its environment and in particular its capacity to produce quality wood. The nutrient regime varies from −3 to +2, with hyper-oligotrophic soil (−3 NR) and different degrees of oligotrophic soils (−2 and −1), mesotrophic soils with a balanced nutrient supply (0 NR), with virtually no restriction for forest tree growth, eutrophic (+1 NR)) and carbonated (+2 NR) soils with an excess of calcium carbonates following an increasing risk of low and imbalanced nutrient supply. Along with the nutrient gradient, the pH of the 0–20 cm soil layer (A-horizon) varies from less than 3.8 (−3 NR) to more than 7.5 (+2 NR). The pH value reflects the soil nutrient supply and has therefore a central role in the abiotic identification key for nutrient regimes (Gégout et al., 2003). Among the other soil characteristics that are required in the process of nutrient regime determination, there are the soil depth, the presence or absence of podzolic profile, of calcareous bedrock and of alluvial or colluvial sediments. Besides, an HCl test is performed on fine soil particles if any active calcium carbonates is suspected. Concerning the moisture regimes, it is a comprehensive combination of two distinct but interrelated gradients: water and oxygen availability. Ten regimes depict the moisture gradient (y-axis), ranging from −4 MR to +5 MR (Fig. 1). Moisture regimes with positive value (+1 to +5) represent sites with a lack of water and negative MR are representative of forest sites with an excess of water implying a lack of oxygen (hypoxia). Moisture regime 0 represent soil with a constant and proper oxygen and water supply: neither too dry nor too wet. The determination of moisture regime by means of the abiotic identification key requires soil (texture, depth, gleyic coloration) and topographic information (thermal sector and water supply from lateral flow). Thus, forest site classification in Wallonia is structured, for a specific bioclimatic area, by the combination of nutrient and moisture regimes. Until now, nutrient and moisture regimes have been mostly determined by the use of identification keys which are based on abiotic information, e.g., soil depth, soil granulometry, hydromorphic indices, stone content, pH and topographic position of the forest site. However, the inference from understory indicator species has recently reached a state of operationnal development (Claessens et al., 2021). Indeed, the concept of ecological groups (*indicator plant groups*) used for a long times in Belgium (Duvigneaud, 1946; Noirfalise, 1984; Godart, 1989) have been revised and the ecological niche of each ecological group has been modelled in the ecogram matrix. The new classification of *223* understory indicator species into 32 ecological groups (Claessens et al., 2021) includes *floristic* identification keys that enable one to determine nutrient and moisture regime starting from a vegetation relevé. As the prediction of moisture regimes from indicator species gives results that are not fine enough to comply with all the ten distinct moisture regimes, further grouping has been realized resulting into four new moisture regimes: dry forest site (+5,+4 and +3 MR), mesic (+2 and +1 MR), fresh (0 and −1 MR) and humid or marshy forest sites (−2, −3 and −4 MR). Although the choice of indicator species and the definition of ecological groups have benefited from statistical analysis, they are ultimately the result of an extensive expertise of

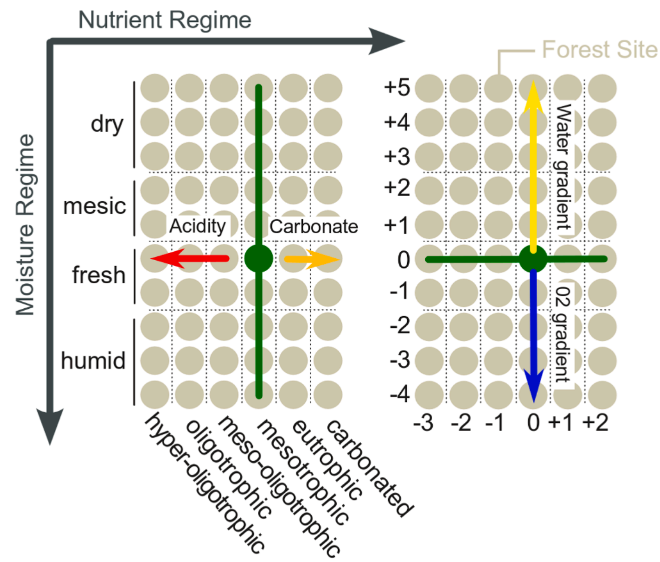

Fig. 1. Forest site classification with the ecogram, with nutrient regime as x-axis and moisture regime as y-axis. **On the left**: the ecogram with the trophic gradient is illustrated. The green line indicates forest sites featured by perfect nutrient availability. The red arrow shows the acidity gradient, with −3 nutrient regime corresponding to the hyper-oligotrophic (or hyperacidic) soil. In contrast, the orange arrow indicates the carbonate gradient, ending on forest sites featured by nutrient imbalance due to an excess of calcium carbonates. **On the right**, the ecogram matrix with moisture gradient is illustrated. The green line represents forest sites with good water and oxygen availability, in other words forest sites without water supply restrictions. The yellow arrow indicates the lack of water gradient, with very dry forest sites on the top of the ecogram. The blue arrow indicates the lack of oxygen availability, due to an excess of water in the soil.

forest ecosystems.

Determining forest nutrient and moisture regimes from a vegetation survey still requires a significant level of botanical expertise and knowledge of species ecology. This applied research aims at replacing part of this expertise by artificial intelligence, through the use of random forest classification algorithms. First, we gather and harmonize thousands of georeferenced floristic relevés from various vegetation surveys in order to cover the nutrient and moisture gradients. We determine the nutrient and moisture regimes from geographic information systems and we conduct an expert-based verification. Second, we train two random forest classifiers in order to predict the six nutrient regimes and the four moisture regimes separately. Both models mimic the expertise of the phytosociologist using exclusively the presence of indicator species as explanatory variables. Third, we release a dedicated web application integrating these models, which supports the encoding and reporting of vegetation data in order to facilitate and promote the use of indicator species for the determination of forest site.

2. Materials and methods

2.1. Study area and environmental layers

The study area corresponds to the south part of Belgium, i.e., Wallonia. The Digital Soil Map of Wallonia and the LiDAR Digital Terrain Model offer a proper description of forest site accross the study area. Although abiotic identification keys that discriminate moisture and nutrient regimes were primarily designed for a use in the field, they have been automated for a use in GIS analysis (Wampach et al., 2017). The goal was to produce maps of nutrient and moisture regimes that cover the whole region, which will be used for the attribution of a moisture and nutrient regime to each georeferenced vegetation relevé. First, forest site topographic description has been modelled from the Digital Terrain Model. An appropriate use of multiple topographic position indexes, slope and radiation indexes, all of them computed from the digital terrain model, enabled to discrimitate the three global radiation sectors required for the determination of the moisture regimes (cold north-east facing slope, warm south-west facing slope and neutral sites) (Wampach et al., 2017). Secondly, the occasional supply of water from lateral water flow was modelled and computed. Lateral water supplies and radiation sectors were then complementarily used with the soil description of the Digital Soil Map of Wallonia in order to determine the moisture regime at every location in Wallonia. Mapping of the nutrient regimes, in constrast, was not easily realized: the nutrient status depends on chemical processes for which the few observations we had were not compatible with the goal of generating an accurate map of the entire region. This is why it has been hypothetized that a nutrient regime could be determined for a georeferenced relevé only if a field or laboratory pH measurement of the top soil layer has been performed.

2.2. Forest understory vegetation relevés

2.2.1. Compilation of relevés

We combined relevés of forest understory vegetation accross Wallonia, including a few large vegetation surveys (> 500 relevés) that were complemented with additional surveys (~ 100 relevés) targeting underrepresented environemental conditions. As two environmental gradients are studied separately (nutrient and moisture regimes), two distinct floristic databases have been set up. In comparison to the moisture floristic database, the nutrient floristic database contains a smaller number of relevés, those associated with a soil pH measurement (Table 1). The common methodology of the floristic relevé consisted in the identification of all understorey vascular plants in a homogenous forest site. In most of the surveys, the sampled area was not strictly delimitated, but the relevé was stopped when no new species were found and identified. Species were inventoried in three vegetation layers: herb, shrub (maximum heigth of 3 meters) and tree layers and ground cover fraction was estimated using the Braun-Blanquet cover-abundance scale (Braun-Blanquet, 1932). For the prediction of forest site, only herb and shrub layers were used, as composition of the forest stands are not necessary natural vegetation that reflect the environmental conditions, depending on forest management, and only species presence (or absence) are considered, as cover estimation from historical and recent relevés surveyed by different person are not necessarily comparable (Diekmann, 2003).

Table 1
Vegetation surveys gathered in two databases, one devoted to nutrient regime (**NR**) analysis and the second for moisture regime (**MR**). Figures represents the number of utilized relevés.

Surveys	NR	MR	Description	Reference
Regional permanent forest inventory	847	6771	Systematic sampling (south Belgium)	(Sanchez et al., 2007)
Phytosociological mapping center	615	2576	All vegetation types in Belgium	(Noirfalise, 1984)
Belgian Natura 2000 network		522	European Habitats Directive 92/43/EEC	
Tournibus forest	139		Handles local variations in NR	(Van Calster et al., 2008)
Flanders (North of Belgium)	82		Sandy soil, Atlantic climate – no georeferencing	(Lameire et al., 2000; Baeten et al., 2009; Baeten et al., 2010)
Gembloux Agro-Bio tech	200	144	Mesotrophic, marshy and riverside forest site	(Claessens et al., 1999; Claessens et al., 2010) (De Jaegere et al., 2016; Dubois et al., 2021)
Additionnal relevés in carbonated soils	135		Calcareous bedrock and shallow soils forest sites	(Piqueray et al., 2011)
Total number of relevés	2055	10 013		
Number of species	798	826		
Number of species (⩾5 occ.)	443	588		

A total of 2055 vegetation plots (relevés) were collected from a dozen of different vegetation surveys (Table 1) to cover the nutrient gradient. In order to support forest experts in the process of determining the nutrient regime of every sample plot, only relevés for which a pH measure has been performed are used. Two surveys have made a predominant contribution. The first one is the vegetation survey realized by the regional forest inventory (Sanchez et al., 2007). The regional forest inventory uses a systematic sampling scheme, with a rectangle grid of 1000 meters of west-east side and 500 meters of sourth-north side. Permanent sample plots are setted up on each point of the spatial grid which are located in a forested zone. Vegetation relevés are performed for trees and understorey plants on circular plot of 12 meters radius in order to characterise forest biodiversity and habitat (Sanchez et al., 2007). A small fraction of these 11 000 forest sample plots are monitored by soil scientists. In addition to the traditionnal field soil survey by hand-augering, laboratory measurements are carried out for some soil parameters, including pH. Thus, pH measurements for 847 floristic relevés were collected. Due to the systematic sampling approach, these surveys adequately cover the Walloon surface. Unfortunately, vegetation relevés are strongly imbalanced along the nutrient gradient, as 68% of Walloon forest sites are oligotrophic or meso-oligotrophic (−2 and −1 nutrient regime), typically featured in a shale bedrock context. The second predominant vegetation survey that has been gathered was issued from the phytosociological mapping research center. This former research lab has been active from 1948 to 1978 under the direction of Noirfalise

(Noirfalise, 1984), which has been of major importance for the development of vegetation association maps in Wallonia. The set of relevé samples from the phytosociological mapping research center covers very well the nutrient gradient, as particular attention has been made on sampling all different vegetation and forest sites. This vegetation survey has been carried out by phytosociologists in optimal phenological conditions: during spring or at the beginning of summer, when the understory vegetation is well developped. Topsoil water pH measurements have been performed on the 0–20 cm soil layer (A-horizon), mainly on the field, by means of Hellige pH meter (McCoy and Donohue, 1979), for a total of 615 vegetation plots. Eventually, additionnal relevés performed in Flanders, in calcareous forest sites and in mesotrophic sites were integrated in the nutrient floristic database (Table 1) in order to cover the best all the nutrient gradient.

Once the standardization and compilation of the nutrient set of relevés has been finalized, nutrient regimes were automatically determined based on the combination of GIS-derived edaphic information and field pH measurement. In the special case of relevés from the North of Belgium for which the exact geographic position is not known, field edaphic observations replace the GIS-derived information. The process of assigning a nutrient regime to a vegetation plot is performed according to the edaphic identification key for nutrient regime. In comparison with the nutrient dataset, an extensive number of relevés was compiled to cover the moisture gradient, as shown in Table 1. Indeed, considering the hypothesis that GIS-extracted informations are sufficient for the prior determination of the moisture regime, the sole required information apart from the vegetation relevé itself is the geographic coordinates of the plot. First, all the 6771 relevés from the regional forest inventory were collected. Secondly, 2576 plots issued from the former phytosociological mapping research center surveys were added. Thirdly, field surveys that were made prior to the definition of Natura 2000 zones in Belgium were collected. Eventually, 144 relevés from marshy or riverside forests from Claessens et al. (2010) have filled in the moisture floristic dataset in order to properly represent humid forest sites. Then, extraction of moisture regime from geograpic layers have been performed for each sample plot. The resulting comprehensive floristic database contains 10,013 relevés, all of them associated with site moisture (MR). All the vegetation communities of Wallonia were well represented in this extensive dataset.

2.2.2. Expert-based verification and determination of nutrient and moisture regimes

For both the nutrient and the moisture floristic databases, a carefull inspection was made by a phytosociological expert to control the consistency of the relevé association with the nutrient or moisture regimes (Fig. 2). Considering the high number of vegetation relevés (2055 for the nutrient regime and 10,013 relevés for the moisture regime), a specific graphical user interface and processing application has been developped (in C++). This application was developped to track relevés with a high likelihood of presenting a poor fit between species presence and a determined forest site (combination of nutrient and moisture regime). For example, the calcareaous ecological group of Bear's Garlic (*Allium ursinum*) is never present on oligotrophic site and only exeptionnaly on meso-oligotrophic site (ecological group has the name of its most representative indicator plant). All relevés with the presence of Bear's Garlic ecological group on a nutrient regime of −3 (hyper-oligotrophic) or −2 (oligotrophic) were easily identified and a proper nutrient regime was deduced by the expert on the basis of indicator species presence. For the nutrient database, every single relevé was checked by an expert. For the moisture database, verification was performed only for outliers (e.g relevés classified as humid but with the presence of xerophilous species). The detection of suspicious relevés was also facilitated by the use of a first set of random forest models.

2.3. Automatic classification of forest site from indicator species

We used non-parametric approach of Random Forest algorithms to model moisture and nutrient regimes of the relevés. Random Forest is a supervised method of classification that has gained prominence in ecology (Prasad et al., 2006; Cutler et al., 2007). The Random Forest method has proven its efficiency in managing high dimensional problems (Genuer et al., 2010) and consists in a collection of Classification and Regression Trees (CARTs) (Breiman et al., 2017). Individual classification trees are trained on a bootstrap sample of floristic and forest site observations by randomly selecting a subset of explanatory variables (presence/absence of plants) at each node. All the decision trees do not predict the same class (NR or MR), due to the sampling of both the in-bag samples (floristic observations used for training the decision tree) and of the explanatory variables used at each tree nodes. Prediction from a Random Forest corresponds to the aggregation of different individual tree predictions. For example, each of the six nutrient regimes received thus a certain number of votes, with one vote corresponding to the prediction of one decision tree for this regime. The nutrient regimes which received the highest number of votes is the most probable one. Out-of-bag error is a prediction error estimate, based on the out-of-bag sample. This sample corresponds to a set of observations which are not used to build the current individual decision trees (Genuer et al., 2010). This classification performance was used as a quality measurement for the nutrient and moisture regimes prediction.

Random Forest are subject to severe overfit (Evans et al., 2011): an exagerated fitting of the training dataset, along with a poor performance in practical application. In order to reduce overfitting, a two-step approach of data augmentation and balancing was implemented (Fig. 3). Data augmentation is a technique used to increase the number of observations of a dataset. A preliminary inspection of the floristic relevés have revealed that most of them present a high number of different undestorey species. Half of the relevés from the nutrient database present at least 24 understory plants, including minimum 13 indicators species. For comparison, Hawkes et al. (1997) consider that half a dozen indicator species is sufficient for providing a sound description of the forest site. We thus artificially decreased the completeness of the training set by sampling indicator species of relevés. Subsampled (*truncated*) floristic relevés were computed and added to the training dataset. The subsampling was applied exclusively on floristic relevés including more than four different ecological groups to create truncated relevés of three ecological groups. Three ecological groups seem indeed to be a proper minimum number for a correct prediction of the forest site, considering that most of relevés in hyper-oligotrophic contains only two or three different hyper-acidophiles ecologigal groups[1]. To this end, a large relevé was artificially broken down in a certain number of truncated relevés. First of all, a list of ecological groups present in the raw relevé was prepared. Three ecological groups were then allocated to a first truncated relevé, by random selection and retrieval from this list. All the indicator species of the raw relevé related to these ecological groups were utilized to create this truncated relevé. If the list of remaining ecological groups contained more than four ecological groups, another truncated relevé was generated the same way, and so on. The complete training dataset contained thus the raw relevés plus all the subsampled relevés. Eventually, data balancing was realized in order to balance the number of relevés amongs nutrient and moisture regimes. The same numbers of relevés for the six nutrient regimes (n = 500) and for the four moisture regimes (n = 2000) were randomly selected (with replacement) from the dataset prior to model training. An ordination was performed in order to check the representativeness of the floristic database after data augmentation and balancing. The projection of species scores on the ordination axes also

[1] Ecological groups of blueberry (*Vaccinium myrtillus*), whirtleberry (*Molinia caerulea*), Purple Moor-grass and sphagnum mosses

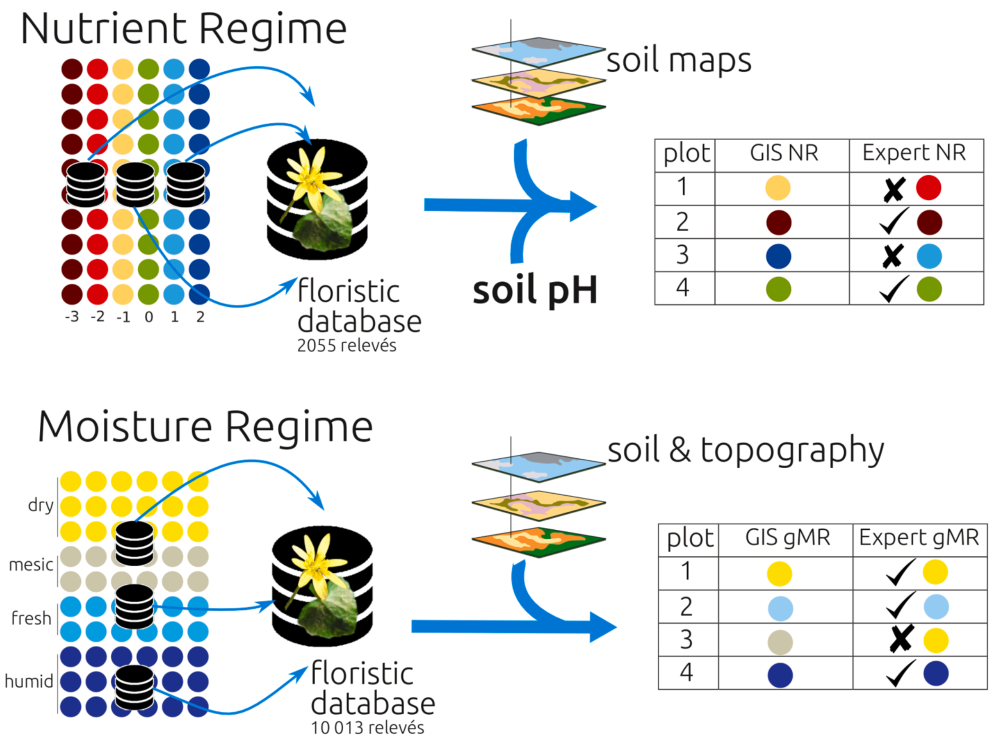

Fig. 2. Preparation of vegetation and forest site datasets. **Upper part**: setting up data for nutrient regimes. Understory vegetation surveys were compiled in order to cover at the best the nutrient gradient. Vegetation sample plots were kept only if they are localized and if a measure of soil pH has been performed (field or laboratory measurement). The nutrient regime was determined based on edaphic properties (known from GIS extraction) and soil pH measurement. An expert inferred an individual vegetation plot and performs, if required, a correction of the nutrient regime based on the presence of indicator plant and ecological groups. **Lower part**: preparation of the dataset for moisture regime. Different vegetation surveys were gathered in one database that properly covers the moisture gradient. Topographic and soil description geographic layers are used to determine the moisture regime of each plot and an expert carried out the verification and correction of the moisture regime based on the presence of indicator species.

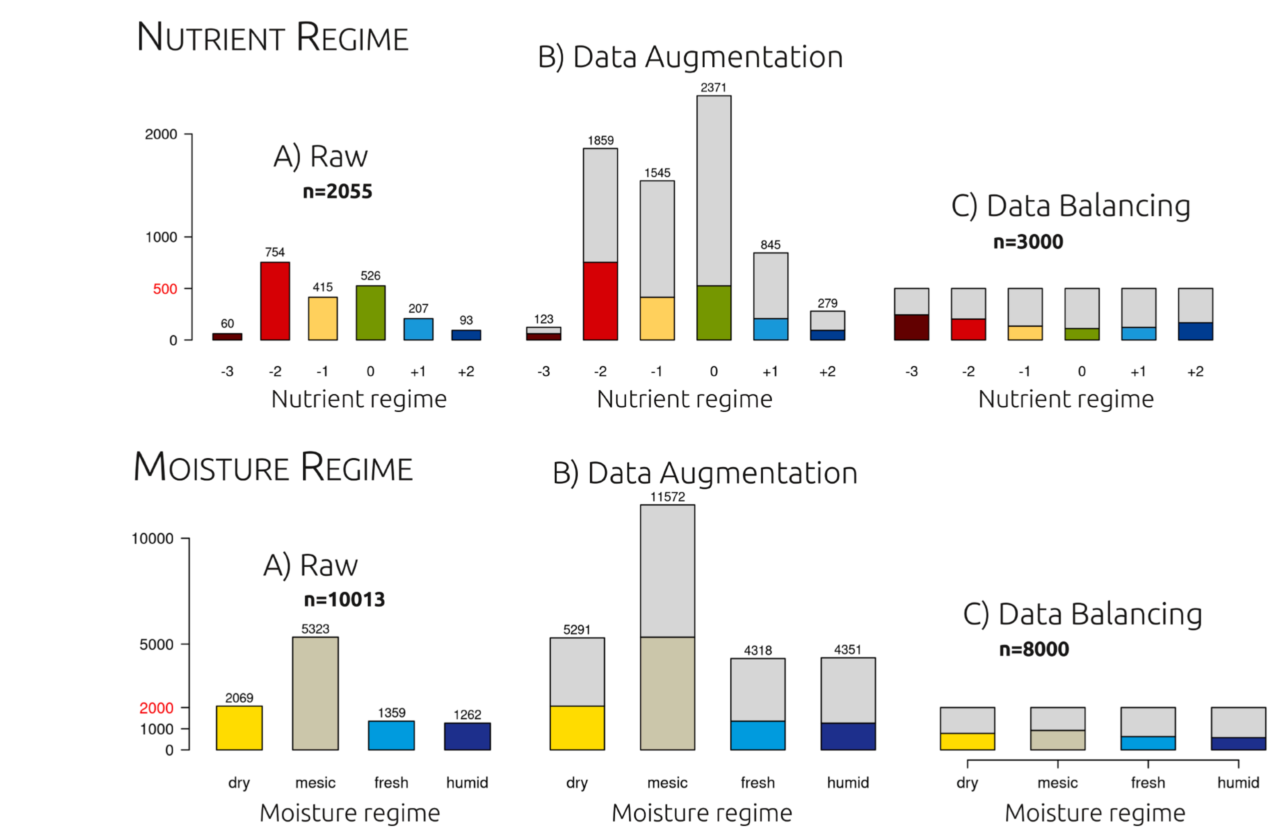

Fig. 3. Number of floristic relevés per nutrient regime (Upper part) and moisture regime (Lower part) after data augmentation and balancing. **Upper part**: compilation of multiple floristic inventories resulted in 2055 floristic samples (left histogram). Data augmentation created truncated relevés, displayed in the middle by grey rectangles. Then, data were balanced along the nutrient gradient (on the right) in order to sample 500 relevés per nutrient regime. **Lower part**: a total of 10013 relevés were collected along the moisture gradient. Truncated relevés were generated (grey rectangles in the middle) and balancing was performed in order to keep 2000 relevés per moisture regimes.

allowed verifying the relevance of ecological groups. To do so, we performed a correspondance analysis on the presence matrix of 262 understory species (with at least 100 occurences) in 12,068 relevés (nutrient and moisture floristic databases).

A Random Forest classifier of 1000 decision trees was then trained on the presence/absence matrix (3000 relevés after data augmentation and balancing) in order to classify the six nutrient regimes. Similarly, another Random Forest classifier was trained on the presence/absence matrix (8000 relevés after data augmentation and balancing) in order to classify the four moisture regimes. We use the C++ version of the Random Forest software *ranger* (Wright and Ziegler, 2015). In addition to the presence/absence of the 223 indicator species, explanatory variables included also the presence/absence of the 32 ecological groups. Once a single indicator species of a specific ecological group is present, we consider that this ecological group is present in the relevé. Ecological group act thus as redundant but synthetic information about the presence of indicator species on a particular forest site, exactly the same way they are used by field foresters.

In order to make these floristic-based classification of nutrient and moisture regimes available to all forest practitionners accross Wallonia, the two Random Forest algorithms have been implemented in a website (http://phytospy.gembloux.ulg.ac.be/) which includes a form to encode a floristic relevé, as well as a view of the predicted nutrient and moisture regimes in the hydro-trophic ecogram matrix. This website is free and exclusively developped from open-source C++ libraries, among which the web framework *Wt* library (Emweb, 2022). It presents the 32 ecological groups of Belgium and includes a photo database for species description.

3. Results

3.1. Vegetation relevés along the nutrient and moisture gradients

We have assembled a large database of vegetation relevés covering the variability of environmental conditions and plant association. To our knowledge, no previous research has led to an understory vegetation survey that covers all forest sites in Wallonia with associated abiotic variables. Especially considering the fact that some forest sites are extremely marginal in terms of surface. Indeed, calcareous forest sites (nutrient regime +2) in Wallonia represent only 0.8% of the forested area. Swampy forest soils, for their part (corresponding to a moisture regime of −4) cover 1.8% of forest sites. Many understorey species are specific to these marginal environmental conditions. A huge number of relevés were collected: 2055 plots for the nutrient gradient, 10,013 for the moisture gradient (Fig. 3). Despite the balancing of floristic relevés among the six nutrient regimes or among the four moisture regimes, marginal forest sites are still largely underrepresented. For the nutrient gradient, only 60 relevés correspond to hyper-oligotrophic regime (−3 NR) and 93 to carbonated (calcareous) regime (+2 NR). In comparison, oligotrophic sites (−2 NR) are predominant, with 754 relevés (Fig. 3) in accordance with their large distribution accross the region. For the moisture regime, the raw dataset is less imbalanced due to the initial grouping into four moisture regimes (dry, mesic, fresh and humid). We confidently suppose that the floristic variation is well represented in these diverse floristic samples. A correspondence analysis on all the relevés has been realized in order to investigate if nutrient and moisture gradients are properly covered (Fig. 4). The first ordination axis (4.0% variance explained) corresponds to the nutrient gradient, from the poorest regime (−3 NR) with negative scores and up to the most calcareous regime (+2 NR) with positive scores. The second axis (3.3% variance explained) correspond to the moisture gradient from the most humid relevés with negative scores up to the driest relevés with positive scores. The second ordination axis shows grassland vegetation communities with negative scores which correspond to relevés located on forest edge, featured by light-demanding plant species, e.g., *Achillea millefolium*, *Plantago lanceolata* and *Trifolium pratense*. The ordination also illustrates how the nutrient and moisture gradients are interdependent in the south of Belgium: rich soils (+1 and +2 NR) are associated with a dry moisture regime. The dataset used for model training is represented on the ordination (Fig. 4) in orange for nutrient regime and in purple for moisture regime and illustrates the proper coverage of the floristic variation after data augmentation and balancing. Examination of species scores in the ordination space confirmed the relevance of ecological groups that are properly spreaded on the nutrient and

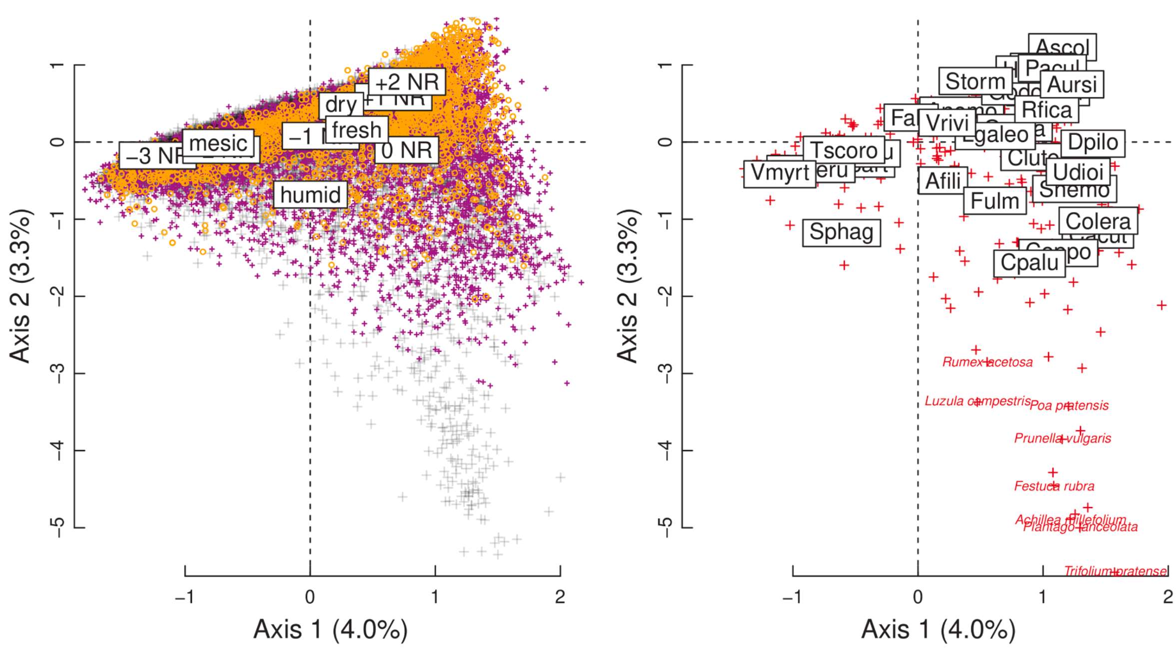

Fig. 4. Axes 1 and 2 of floristic relevés ordination by correspondence analysis based on the presence of 262 understory species in 12068 relevés. Left figure shows positions of the relevés (in black, all relevés, in purple relevés that were used to train the moisture regime model and in orange relevés utilized for the nutrient regime model). Right figure illustrates the position of the species in red and the position of ecological groups (labelled with the first letter of the genius and the four first letters of the species.)

moisture gradient. For example, Blueberry (Vmyr on Fig. 4) with the most negative scores on the first ordination axis (nutrient gradient) is associated with an hyper-oligotrophic nutrient regime (−3 NR), and Bear's Garlic (Aursi) with positive score is associated with a carbonated nutrient regime (+2 NR).

3.2. Prediction of the nutrient and moisture regimes

Prediction of nutrient and moisture regimes exclusively from understory vegetation showed a fine performance. The overall out-of-bag error for the nutrient Random Forest fitted on the 3000 relevés after data augmentation and balancing was 29.7 %. It means that 29.7% of cross-validated relevés were misclassified by the model. The prediction model for the moisture regimes fitted on the 8000 relevés performed less well, with an overall out-of-bag which reached 44 %.

Concerning the predictions of the nutrient regime from indicator species, the confusion matrix shows a good distribution of relevés on the diagonal (Table 2) confirming the good model quality. The excellent match for carbonated relevés was highlighted with 95% of proper automatic classification. The perfect fit for the hyper-oligotrophic site (−3 NR) is interesting, but should be taken with caution as it could be a sign of model overfit. The classification of eutrophic sites (+1 NR) was complex. A total of 284 of the 500 eutrophic relevés were correctly classified (57%), and most of the confusions came from the similarities between carbonated (+2 NR) and eutrophic floristic species. Oligotrophic and meso-oligotrophic sites (−2 and −1 NR), for their part, were subject to slight confusion. A total of 91 meso-oligotrophic (−1 NR) relevés were classified as oligotrophic (−2 NR), and 82 oligotrophic samples were predicted as meso-oligotrophic forest site. Forest sites on mesotrophic soil (0 NR) presented the lowest performance because these conditions suit plenty of understorey plants: these conditions are optimal for a high variety of plants, resulting in a rich mix of neutrophilous and acidiclin indicator species. Many of indicator plants that thrive in this nutrient regime are present in meso-oligotrophic and eutrophic sites. 186 of the 500 mesotrophic (0 NR) relevés were correctly classified (37%). 30 % of 0 NR were predicted as meso-oligotrophic sites, and 24% as eutrophic soils. This classifier is pessimistic about the forest site, as it tend to underestimated the nutrient regime quality. Preference is given to a model which is pessimistic and thus which does not predict easily an uncontrained forest site like the 0 NR (mesotrophic).

Concerning the prediction of the moisture regimes, the Random Forest (RF) model showed a lower performance, even though the grouping of moisture regimes intended to adress the problem of somehow limited indicator values of understorey plant to moisture conditions. The confusion matrix for the 2000 floristic samples per moisture regimes shows that humid sites are quite well discriminated with the Random Forest model (Table 3). With 69.5 % of correct classification (n = 1390), humid forest sites are the preffered habitat for a bunch of indicator species that are dependant of a good water supply. Hygrophilous species show a narrow ecological amplitude, and are therefore accurate indicator species. Their well-known fidelity is illustrated by the rate of disagreement by two or more classes. Disagreement by two classes is the misclassification of humid soils into another moisture regimes that is not adjacent (e.g. mesic and dry MR are not adjacent to the humid MR). The fresh MR is adjacent to the humid MR and the RF model have some confusion with it, but mesic and dry MR, with 201 and 76 misclassified floristic relevés out of a total of 2000, are strongly differentiated from humid MR. Disagreement by two classes is thus extremely low, with only 14 % (n = 276) for humid MR. More surprisingly, the dry MR sites are also featured by a low two-classes disagreement rate, as most of the confusions are made with mesic MR. This good performance is certainly attributable to the interaction between the trophic gradient and the moisture gradient which is particular to Belgium. Indeed, calcareous species thrive mainly in dry forest sites in Belgium (see left ordination of Fig. 4), as soil with calcareous bedrock are rarely found on humid site but very often in the direct surrounding of cliffs or on steep terrain. Thermophilic, xerophilic and calcareous flora are thus interrelated and mostly confused in our analyses, which explain why dry MR sites are so well differenciated by the Random Forest model. Similarly to the nutrient gradient, the performance were lower for the optimum forest site, here the fresh moisture regimes. These soils are indeed featured by a proper supply of water and oxygen during the whole vegetation period. Again, in these conditions, a high number of indicator species were present: plants that transgress either to mesic forest site or to humid soil.

3.3. Integration of the predictive models into a web application

The RF models have been implemented into a website dedicated to forest practitioners. For a relevé, predictions of the nutrient and moisture regimes are displayed on barplots, adjacent to the corresponding axis of the ecogram (Fig. 5) and representing the likelihood scores. The regime with the highest score/bar, which is filled with a dark grey color, is considered as the predicted regime regarding the floristic relevé. In the ecogram, intersection of the predicted nutrient regime with the predicted moisture regime are the most likely forest site. It is highlighted by a yellow color (Fig. 5). The first relevé on the left is made up of eigth understory species including five indicator species, which belong to two ecological groups: Wood Sage and Blueberry. These ecological groups are both acidiphilic and mesophilic, which is in accordance with the prediction of the Random Forest models. An oligotrophic nutrient regime (−2 NR) is predicted with 72% of likelihood and a mesic moisture regime with a probability of 69%. The second relevé on the rigth is composed of 17 forest understory species: all of them are indicator species related to 12 distinct ecological groups. Although the ecological niche of these groups varies along the nutrient and moisture gradients, the presence of numerous calcareous indicator species (e.g., *Acer campestre*, *Allium ursinum*, *Mercurialis perennis* and *Primula veris*) mixed with wide-amplitude neutrocline (e.g., ecological group of Stinging Nettle) lead the Random Forest model to assign an eutrophic nutrient regime (+1 NR). Regarding the moisture regime, the prediction is not very strong with a likelihood of 37% for the dry regime. The mix of xeroclinic

Table 2
Confusion matrix from out-of-bag cross validation for the prediction of nutrient regime from indicator species presence/absence by Random Forest (**RF**). Overall out-of-bag error is 29.7%, and most of the confusion are related to mesotrophic soil (0 NR).

			Nutrient regimes infered by expert					
			−3	−2	−1	0	+1	+2
RF predictions	hyper-oligotrophic	−3	**500** (100%)	84 (17%)	11 (2%)	2	1	0
	oligotrophic	−2	0	**324** (65%)	91 (18%)	11 (2%)	4 (1%)	0
	meso-oligotrophic	−1	0	82 (16%)	**316** (63%)	153 (31%)	45 (9%)	1
	mesotrophic	**0**	0	4 (1%)	55 (11%)	**186** (37%)	61 (12%)	3
	eutrophic	+1	0	5 (1%)	24 (5%)	121 (24%)	**284** (57%)	20 (4%)
	carbonated	+2	0	1	3	27 (5%)	105 (21%)	**476** (95%)
	> 1 class disagreement		0%	2%	7%	7%	10%	4%

Table 3
Confusion matrix from out-of-bag cross validation for the prediction of the moisture regimes from indicator species presence/absence by Random Forest (**RF**).

		Moisture regime infered by expert or by GIS analysis			
		dry	**mesic**	**fresh**	**humid**
RF predictions	**dry**	**1176** (59%)	474 (24%)	307 (15%)	76 (4%)
	mesic	551 (28%)	**1080** (54%)	384 (19%)	201 (10%)
	fresh	206 (10%)	288 (14%)	**807** (40%)	333 (17%)
	humid	67 (3%)	158 (8%)	502 (25%)	**1390** (69%)
	> 1 class disagreement	13%	8%	15%	14%

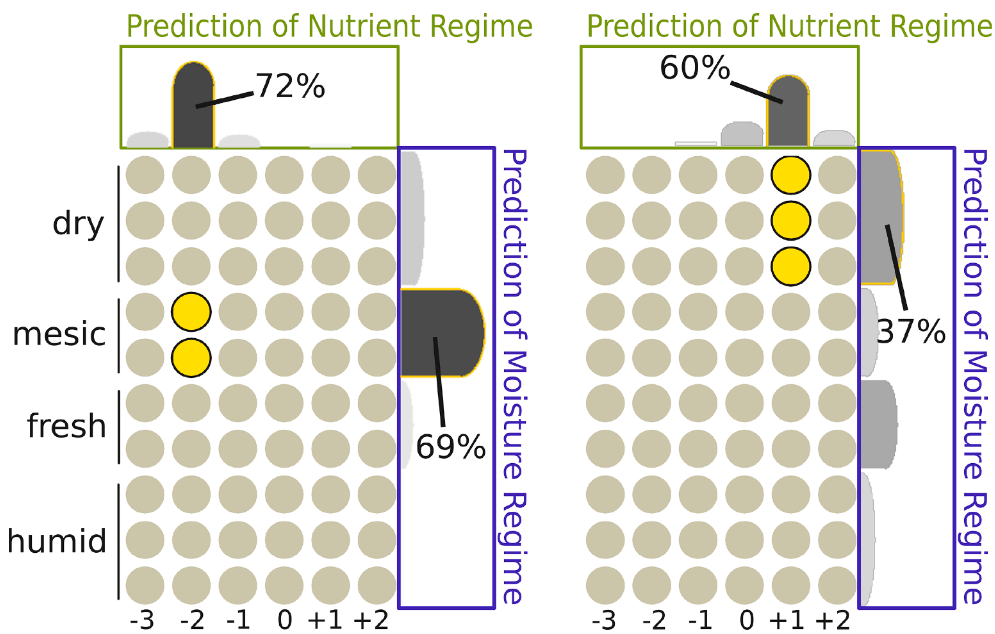

Fig. 5. Illustration of the prediction models integrated in a website for two floristic relevés from Sougnez (1978). On the top, the two floristic relevés. On the bottom, the ecogram matrix with the barplot of the prediction for nutrient regime on the top and of the moisture regime on the right. The two barplots provide the result from Random Forest models, which attribute a likelihood for each regimes. Intersection of the most likely nutrient and moisture regimes is highlighted in yellow.

species as *Brachypodium sylvaticum* or the ecological group of Fingered Sedge with different mesophilic species lead to a poor prediction performance for the moisture regime, which is brought to the attention of forest practionners by means of the prediction likelihood.

4. Discussion

In this study, we aimed to predict the forest site, and specifically the nutrient and moisture regimes, from understory vegetation in order to make available the floristic expertise of phytosociologists to a large audience. Thousands of floristic relevés scattered across Belgian forests were gathered into two databases, respectively covering the nutrient and the moisture gradients. Each relevé was associated with a nutrient regime or with a moisture regime by a floristic expert, based on the presence of indicator species and on the soil and topographic description of the forest site obtained from geographical layers. The *223* understory indicator species representing 32 ecological group were used as explanatory variables in two Random Forest models that predicted nutrient and moisture regimes from the presence of indicator species and ecological groups. To give a similar importance to each regime, we performed a meticulous preparation of the training datasets consisting in data augmentation and balancing. Data augmentation was performed to counterbalance the risk of model overfitting, while data balancing was implemented to compensate the skewed class distribution of relevés along environmental gradients. The Random Forest model predicting nutrient regimes achieved a rate of error of 30% on cross-validation, while the one predicting the moisture regime performed slightly less good with an out-of-bag error of 44%. Our Random Forest models were implemented in a decision support tool for forest practionners in a free website application (http://phytospy.gembloux.ulg.ac.be/).

The good global performance of our models confirmed the usefullness of understory indicator species for describing forest sites (Wang, 2000; Wilson et al., 2001; Bergès et al., 2006). The Belgian forest trees autecology guide methodology, which focus on the three major environmental gradients (climate, nutrient and moisture), is a simplification of the forest site conditions. In particular, moisture regime summarizes both water availability, hypoxia and thermal microclimate conditions, although thermal conditions have by themselves a strong impact on species distribution (Sewerniak and Puchałka, 2020). The seasonality of water availability also hinders the use of indicator species to characterize the moisture regime since hygrophilous species can growth and close their life cycle during spring even if the soils become dryer during summer. Nonetheless, the performance of our models is adequate for an operationnal use by Belgian foresters. Here, to summarize the phytosociological expertise, we used a non-parametric machine learning algorithm which handles non-linear species-environment relationships (Cutler et al., 2007; Evans et al., 2011). Our approach did not focus on the interaction between one species and the forest environment, but operated at the scale of the entire floristic relevé. Nutrient and moisture regimes were directly predicted from the presence/absence matrix of indicator species and ecological groups in constrast to most applied phytosociology studies which use the indicator value of discriminating species. Indicator values were proposed by phytosociologists as a mean to represent the ecological niche of indicator plant species: they are an extremely succinct and subjective summary of the expert knowledge of a complex ecosystem-plant interaction (Diekmann, 2003; Biondi, 2011). In order to objectify these expert-interpreted indicator values, statistical indices of both fidelity and specificity of species to specific forest site characteristics have been developped, e.g., *IndVal* (Dufr and Legendre, 1997). Although *IndVal* has a nice mathematical design, which has been improved by De Céceres and Legendre (2009), it fails to grasp all the complex nature of ecological systems. Average of species indicator values is generally performed to describe forest site of a vegetation relevé (Bartoli et al., 2000; Wilson et al., 2001; Tichỳ, 2002; Pinto et al., 2016) eventually weighed with abundance (cover). Another approach of forest site description with indicator species consists in the developpement of determination keys, based on the cumulative presence of species grouped in indicator species groups (Wang, 2000) or in ecological groups (Claessens et al., 2021). These determination keys are used manually or are automated in the form of an expert system algorithm (Noble, 1987).

In this study, we used a powerfull supervised classification method which handles the multiple interactions between understory species and their environmental conditions (Evans et al., 2011). This was made possible by the compilation of numerous floristic surveys that end up with huge relevé databases (sup 10 000 relevés). Large and shared vegetation plots databases are becoming more and more common, especially in Europe, offering new opportunities for future phytosocological research (Douda et al., 2016; Chytrỳ et al., 2016). Although Random Forest models cannot be visually inspected due to the high number of decision trees and thus fails to provide a proper understanding of ecological mecanisms (Evans et al., 2011), we believe that they are more powerfull than any expert system, e.g, classification keys, and outperform as well the classical averaging of indicator values. By nature, Random Forest algorithms are highly suitable to develop efficient and complex classifiers and to model relationships that are unknown to the phytosociologists, or at least too difficult to formalize. In addition, their use by forest practitionners is easy, and the prediction for a new floristic relevé has the advantage of providing a likelihood score for each regime, which adds an usefull nuance related to the confidence of a prediction. Our website application (http://phytospy.gembloux.ulg.ac.be/) allows the practical use of our research results by anyone. Once nutrient and moisture regimes are automatically determined from a understory species list, recommandations about the forest species that fit this forest site are also provided.

Finally, this work opens the door to both theoretical and practical perspectives. The approach is relatively easy to extend at a larger scale, notably across Western Europe, and will allow the testing of theoretical questions in community and evolutionnary ecology. Practically, the numerous vegetation relevés, in combination with existing environmental layers, will improve the current mapping of forest site.

CRediT authorship contribution statement

Lisein Jonathan: Conceptualization, Methodology, Software, Formal analysis, Writing – original draft. **Fayolle Adeline:** Conceptualization, Methodology, Writing – review & editing. **Legrain Andyne:** Validation, Resources. **Prévot Céline:** Validation, Resources. **Claessens Hugues:** Conceptualization, Supervision, Resources.

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 research was funded thanks to the *Plan quinquennal de recherches forestières* of the Service Public de Wallonie (forest administration) and was undertaken thanks to the compilation of thousands of floristic relevés. We kindly acknowledge every individual who has contributed to these many surveys and all those who have shared this information. Thanks to Kris Verheyen of the Faculty of Bioscience Engineering of Ghent University for kindly providing vegetation relevés from various studies. Thanks to Gilles Colinet who suppervise soils analysis for the regional forest inventory and to Lionel Wibail from the Natura 2000 European Commission for sharing floristic relevés database. A thank-you to the forest technicians who have digitally recorded the ancient vegetation relevé issued by the phytosociological mapping research center, Benoit Maekels and Frederic Henrotay. We thank all the former students of Gembloux Agro-Bio Tech who have completed the nutrient floristic dataset; Frédéric Mouchet, Julien Piqueray, Laureline Claessens and Augustin Therasse. Thanks also to Géraldine Le Mire for english editing. Eventually, a special thought to François Ridremont who started this research in 2013.

References

Baeten, L., Bauwens, B., De Schrijver, A., De Keersmaeker, L., Van Calster, H., Vandekerkhove, K., Roelandt, B., Beeckman, H., Verheyen, K., 2009. Herb layer changes (1954–2000) related to the conversion of coppice-with-standards forest and soil acidification. Appl. Veg. Sci. 12 (2), 187–197.
Baeten, L., Hermy, M., Van Daele, S., Verheyen, K., 2010. Unexpected understorey community development after 30 years in ancient and post-agricultural forests. J. Ecol. 98 (6), 1447–1453.
Bartoli, M., Tran-ha, M., Largier, G., Dumé, G., Larrieu, L., 2000. Ecoflore, un logiciel simple de diagnostic écologique. Revue forestière française 6, 530–547.
Bergès, L., Gégout, J.-C., Franc, A., 2006. Can understory vegetation accurately predict site index? A comparative study using floristic and abiotic indices in sessile oak (Quercus petraea Liebl.) stands in northern France. Ann. For. Sci. 63 (1), 31–42.
Biondi, E., 2011. Phytosociology today: Methodological and conceptual evolution. Plant Biosyst. 145 (sup1), 19–29.
Braun-Blanquet, J., 1928. Pflanzensoziologie: grundzüge der vegetationskunde. Springer-Verlag, Berlin.
Braun-Blanquet, J., 1932. Plant sociology. The study of plant communities.
Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 2017. Classification and regression trees. Routledge.
Cajander, A., 1909. On forest types. Über Waldtypen. Acta For. Fenn 1 (1), 1–175.
Carignan, V., Villard, M.-A., 2002. Selecting indicator species to monitor ecological integrity: a review. Environ. Monit. Assessment 78 (1), 45–61.
Chytrỳ, M., Hennekens, S.M., Jiménez-Alfaro, B., Knollové, I., Dengler, J., Jansen, F., Landucci, F., Schaminée, J.H., Acic, S., Agrillo, E., and others, 2016. European Vegetation Archive (EVA): an integrated database of European vegetation plots. Appl. Veg. Sci. 19(1): 173–180. Publisher: Wiley Online Library.
Claessens, H., Pauwels, D., Thibaut, A., Rondeux, J., 1999. Site index curves and autecology of ash, sycamore and cherry in Wallonia (Southern Belgium). Forestry 72 (3), 171–182.
Claessens, H., Oosterbaan, A., Savill, P., Rondeux, J., 2010. A review of the characteristics of black alder (Alnus Glutinosa (L.) Gaertn.) and their implications for silvicultural practices. Forestry 83 (2), 163–175.
Claessens, H., Prévot, C., Lisein, J., 2021. Guide d'interprétation de la flore indicatrice en forêt. Forêt.nature edition.
Coudun, C., Gégout, J.-C., 2007. Quantitative prediction of the distribution and abundance of Vaccinium myrtillus with climatic and edaphic factors. J. Veg. Sci. 18 (4), 517–524.
Coudun, C., Gégout, J.-C., Piedallu, C., Rameau, J.-C., 2006. Soil nutritional factors improve models of plant species distribution: an illustration with Acer campestre (L.) in France. J. Biogeogr. 33 (10), 1750–1763.
Cutler, D.R., Edwards Jr, T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., Lawler, J.J., 2007. Random forests for classification in ecology. Ecology 88 (11), 2783–2792.
De Céceres, M., Legendre, P., 2009. Associations between species and groups of sites: indices and statistical inference. Ecology 90 (12), 3566–3574.
De Jaegere, T., Hein, S., Claessens, H., 2016. A review of the characteristics of small-leaved lime (Tilia Cordata Mill.) and their implications for silviculture in a changing climate. Forests 7 (3), 56.
Diekmann, M., 2003. Species indicator values as an important tool in applied plant ecology-a review. Basic Appl. Ecol. 4 (6), 493–506.
Douda, J., Boublík, K., Slezák, M., Biurrun, I., Nociar, J., Havrdová, A., Doudová, J., Acic, S., Brisse, H., Brunet, J., et al., 2016. Vegetation classification and biogeography of European floodplain forests and alder carrs. Appl. Veg. Sci. 19 (1), 147–163.
Dubois, H., Claessens, H., Ligot, G., 2021. Towards Silviculture Guidelines to Produce Large-Sized Silver Birch (Betula Pendula Roth) Logs in Western Europe. Forests 12 (5), 599.

Dufrêne, M., Legendre, P., 1997. Species assemblages and indicator species: the need for a flexible asymmetrical approach. Ecol. Monogr. 67 (3), 345–366.
Duvigneaud, P., 1946. La variabilité des associations végétales. Bulletin de la Société royale de Botanique de Belgique 78, 107–134.
Ellenberg, H., 1974. Zeigerwerte der Gefaesspflanzen Mitteleuropas. Göttingen, publisher erich goltze kg edition, Scripta Geobotanica IX.
Emweb, 2022. Wt – web GUI library in modern C++. URL: https://www.webtoolkit.eu.
Evans, J.S., Murphy, M.A., Holden, Z.A., Cushman, S.A., 2011. Modeling species distribution and change using random forest. In: Predictive species and habitat modeling in landscape ecology. Springer, pp. 139–159.
Fayolle, A., Swaine, M.D., Bastin, J.-F., Bourland, N., Comiskey, J.A., Dauby, G., Doucet, J.-L., Gillet, J.-F., Gourlet-Fleury, S., Hardy, O.J., et al., 2014. Patterns of tree species composition across tropical African forests. J. Biogeogr. 41 (12), 2320–2331.
Genuer, R., Poggi, J.-M., Tuleau-Malot, C., 2010. Variable selection using random forests. Pattern Recogn. Lett. 31 (14), 2225–2236.
Godart, M.-F., 1989. Ecological Species Groups in Forest Communities in South Belgium. Vegetatio 81 (1/2): 127–135. ISSN 00423106.
Gégout, J.-C., Hervé, J.-C., Houllier, F., Pierrat, J.-C., 2003. Prediction of forest soil nutrient status using vegetation. J. Veg. Sci. 14 (1), 55–62.
Gégout, J.-C., Coudun, C., Bailly, G., Jabiol, B., 2005. EcoPlant: a forest site database linking floristic data with soil and climate variables. J. Veg. Sci. 16 (2), 257–260.
Hawkes, J.C., Pyatt, D.G., and White, I.M.S., 1997. Using Ellenberg Indicator Values to Assess Soil Quality in British Forests from Ground Vegetation: A Pilot Study. J. Appl. Ecol. 34(2): 375–387. ISSN 00218901, 13652664. doi: 10.2307/2404883.
Lameire, S., Hermy, M., Honnay, O., 2000. Two decades of change in the ground vegetation of a mixed deciduous forest in an agricultural landscape. J. Veg. Sci. 11 (5), 695–704.
Legrain, X., Bock, L., Colinet, G., 2014. Suitability of the soil map and legacy data in Wallonia (BE) to support the GSM initiative. In: GlobalSoilMap. CRC Press, pp. 99–102.
McCoy, D., Donohue, S., 1979. Evaluation of commercial soil test kits for field use. Commun. Soil Sci. Plant Anal. 10 (4), 631–652.
Noble, I., 1987. The role of expert systems in vegetation science. Vegetatio 69(1): 115–121. Publisher: Springer.
Noirfalise, A., 1984. Forêts et stations forestières en Belgique. Les Presses Agronomiques, Gembloux.
Petit, S., Claessens, H., Vincke, C., Ponette, Q., Marchal, D., 2017. Le Fichier écologique des essences, version 2.0. Forêt. Nature 143, 12–19.
Pinto, P.E., Dupouey, J.-L., Hervé, J.-C., Legay, M., Wurpillot, S., Montpied, P., Gégout, J.-C., 2016. Optimizing the bioindication of forest soil acidity, nitrogen and mineral nutrition using plant species. Ecol. Ind. 71, 359–367.
Piqueray, J., Bottin, G., Delescaille, L.-M., Bisteau, E., Colinet, G., and Mahy, G., 2011. Rapid restoration of a species-rich ecosystem assessed from soil and vegetation indicators: The case of calcareous grasslands restored from forest stands. Ecol. Indic. 11(2): 724–733. ISSN 1470–160X. doi: 10.1016/j.ecolind.2010.06.007.
Prasad, A.M., Iverson, L.R., Liaw, A., 2006. Newer classification and regression tree techniques: bagging and random forests for ecological prediction. Ecosystems 9 (2), 181–199.
Pyatt, G., 1995. Ecological Site Classification for forestry in Great Britain. Forestry Commission (260): 6. Farnham.
Rameau, J.-C., Mansion, D., and Dumé, G., 1989. Flore forestière française: Plaines et collines, volume 1. Forêt privée française.
Sanchez, C., Claessens, H., Puissant, T., Lecomte, H., Rondeux, J., et al., 2007. Vegetation diversity assessment in Southern Belgium's permanent forest inventory. CAB International. Sustainable Forestry: from Monitoring and Modelling to Knowledge Management and Policy Science, pages 208–223.
Sewerniak, P., Puchałka, R., 2020. Topographically induced variation of microclimatic and soil conditions drives ground vegetation diversity in managed Scots pine stands on inland dunes. Agricultural and Forest Meteorology, 291: 108054. Publisher: Elsevier.
Sougnez, N., 1978. Les chênaies-charmaies du district calcaire mosan. Communications du Centre d'écologie forestière et rurales (I.R.S.I.A.),, page 85 pp.
Tichỳ, L., 2002. Juice, software for vegetation classification. Journal of vegetation science, 13 (3): 451–453. Publisher: Wiley Online Library.
Van Calster, H., Baeten, L., Verheyen, K., De Keersmaeker, L., Dekeyser, S., Rogister, J.E., and Hermy, M., 2008. Diverging effects of different overstorey conversion scenarios on the understorey vegetation in a former coppice-with-standards forest. Forest Ecology and Management, 256 (4): 519–528. ISSN 0378–1127. doi: 10.1016/j.foreco.2008.04.042.
Van der Perre, R., Bythell, S., Bogaert, P., Claessens, H., Ridremont, F., Tricot, C., Vincke, C., Ponette, Q., 2015. La carte bioclimatique de Wallonie: un nouveau découpage écologique du territoire pour le choix des essences forestières. Forêt-Nature 135, 47–58.
Verstraeten, G., Baeten, L., Van den Broeck, T., De Frenne, P., Demey, A., Tack, W., Muys, B., Verheyen, K., 2013. Temporal changes in forest plant communities at different site types. Appl. Veg. Sci. 16 (2), 237–247.
Wampach, F., Lisein, J., Cordier, S., Ridremont, F., Claessens, H., 2017. Cartographie de la disponibilité en eau et en éléments nutritifs des stations forestières de Wallonie. Forêt. Nature 143, 47–60.
Wang, G.G., 2000. Use of understory vegetation in classifying soil moisture and nutrient regimes. For. Ecol. Manage. 129 (1–3), 93–100.
Wilson, S.M., Pyatt, D., Malcolm, D., Connolly, T., 2001. The use of ground vegetation and humus type as indicators of soil nutrient regime for an ecological site classification of British forests. For. Ecol. Manage. 140 (2–3), 101–116.
Wright, M.N., Ziegler, A., 2015. ranger: A fast implementation of random forests for high dimensional data in C++ and R. arXiv preprint arXiv:1508.04409.