Modeling potential natural vegetation: A new light on an old concept to guide nature conservation in fragmented and degraded landscapes.

Axel Bourdouxhe[1]*, Lionel Wibail[2] , Hugues Claessens[3] and Marc Dufrêne[1]

[1]Biodiversity and Landscape Unit, Gembloux Agro-Bio Tech, Université de Liège, Gembloux, Belgium;

[2]SPW/DGARNE/DEMNA, Gembloux, Belgium;

[3]Gestion des ressources forestières et des milieux naturels, Gembloux Agro-Bio Tech, Université de Liège, Gembloux, Belgium;

*Corresponding author

Biodiversity and Landscape Unit, Gembloux Agro-Bio Tech, 2 Passage des Déportés, 5030 GEMBLOUX, Belgium

e-mail: axel.bourdouxhe@uliege.be

Accepted author version

Abstract

Modeling biotope distributions is of paramount importance to monitor species habitats and guide conservation and restoration actions to decrease population extinction rates. However, modeling biotopes as independent landscape units, as is current practice, has some limitations. Vegetation communities that define biotopes evolve through different stages and associations until they reach an equilibrium. To consider these temporal dynamics, we developed a modeling approach based on potential natural vegetation (PNV) corresponding to ecological contexts supporting vegetation succession. The

Abbreviations
PNV: Potential Natural Vegetation
BDM: Biotope Distribution Modeling

assumption made is that modeling PNV better distinguishes biotope ecological niches, improving prediction accuracy. Results of the final prediction map were excellent, with an overall accuracy of 0.95 and a kappa coefficient of 0.91. The proposed method was also compared with a classic biotope model and our approach showed 29% mean improvement in accuracy. Our results produced a good distinction between the different ecological niches of potential natural vegetation. However, some areas of confusion were identified but these are mainly explained by imprecision and incompleteness of the reference biotope dataset and long-term human management. Using potential natural vegetation is therefore recommended for further studies on biotope mapping.

Keywords: Potential natural vegetation, biotope, species distribution model, vegetation dynamics, vegetation communities

1. Introduction

The conservation and restoration of functional species habitat networks became a priority to tackle landscape fragmentation, one of the key factor of biodiversity loss (Hilty et al., 2020; IPBES, 2019; Resasco, 2019). This strategy implies the use of multiple information sources involving the recording and mapping of species of interest, vegetation, ecological conditions, land use, pressures, management, dispersal barriers and many more. One key source of information is vegetation mapping using the biotope concept. A biotope is a landscape unit characterized by specific environmental conditions and supporting a characteristic community of species. A species habitat is

Accepted author version

considered as the set of biotopes it needs to complete its lifecycle (Forman, 1995; Löfvenhaft et al., 2002). Biotope level can therefore be considered as the finest unit of landscape management for conservation purpose. Biotopes are commonly described by the vegetation species assemblages depending on more or less precise ecological conditions (Davies et al., 2004). Generally, these ecological conditions are defined by topography, hydrography, climate, soil type and human management, justifying that biotope units are significant landscape features. However, as biotope surveys require skilled teams that sometimes have to cover large territories (Lillesand et al., 2008), biotope mapping is therefore limited by time-consuming surveys, which restricts their capacity to be up-to-date in a land-use change context. To circumvent these technical limitations, biotope modeling using available environmental predictors and remote-sensing data is increasingly being developed (Horvath et al., 2019; Maggini et al., 2006). This approach, derived from species distribution modeling (SDM), produces relatively accurate maps of the current or the potential distribution of biotopes when, taking into account land cover or not (Álvarez-Martínez et al., 2018; Horvath et al., 2019). Current and potential biotope distributions are also strategic information to design ecological networks. Biotope restoration can then target areas to increase species habitat patches and limit fragmentation impacts to compensate habitat loss due to human activities (Jones et al., 2021).

The interest and quality of biotope distribution modeling (BDM) is however limited because several biotopes can occupy similar ecological conditions (soil, climate and topography) during the vegetation successions. Indeed, plant communities characterizing biotopes evolve naturally from pioneer open biotopes to mature forest

Accepted author version

biotopes reaching one or several stages of equilibrium called “climax” (Niering, 1987). Different types of successional patterns are possible with dynamics of natural or artificial disturbances that, for example, restart the process of succession, regressive dynamics, lock-in at certain stages, etc (Blasi et al., 2004; Rees et al., 2001; Woodward, 2009). BDM accuracy is thus disturbed by the competition between different biotopes belonging to the same succession in similar ecological conditions and the difficulty of having descriptive variables to discriminate the different stages. Some authors consider this issue by focusing only on climax stage biotopes (Horvath et al., 2019; Maggini et al., 2006). The presence of some stages/biotopes of a vegetation succession is also impacted by human management that restricts them to less productive areas such as the steepest slopes or the most superficial soils (Hall, 1988). The observed relationship between some biotope distribution and environmental conditions, therefore, can only be one part of the potential biotope environmental niche.

One pragmatic solution to consider temporal dynamics of ecosystems is to use the concept of “potential natural vegetation” (PNV). PNV concept was defined by Tuxen (1956) as ‘the vegetation that would develop in a particular ecological zone or environment, assuming the conditions of flora and fauna to be natural, if the action of man on the vegetation mantle stopped and in the absence of substantial alteration in present climatic conditions’ (Gallizia Vuerich et al., 2001). By convention, PNV therefore refer to climax vegetation but it designs the ecological conditions necessary for the development of the main stages of the associated succession (Loidi and Fernández-González, 2012; Prach et al., 2016). The concept of PNV allows to define landscape units that share similar ecological contexts, regardless of the stage of natural vegetation

Accepted author version

succession (Leguédois et al., 2011; Stumpel and Kalkhoven, 1978). This concept was used in the past to classify landscapes to guide nature oriented sylviculture, conservation and landscape planning (De Keersmaeker et al., 2013). Such PNV maps can also be used as one of multiple approaches to determine the conservation status or naturalness of areas by measuring the difference between PNV and current land cover(Hemsing and Bryn, 2012) or to identify gaps and completion of networks of protected areas such as Natura 2000 (Bohn and Gollub, 2006). PNV mapping can be seen as an integrative approach that better considers vegetation dynamics, which is important for the design and connectivity of ecological networks and their management. It has the advantage of better delineating the maximum potential distribution of biotopes depending on the ecological context and therefore ensures network coherence. It also helps to guide the management of conservation and restoration actions based on natural disturbance to create a more heterogeneous landscape.

However, the expected link between PNV typology and ecological conditions is well debated among experts (Chiarucci et al., 2010). Pioneer or intermediate biotopes are sometimes shared by different vegetation succession series. The dynamic drivers influencing a biotope to develop to a specific vegetation succession may be subtle variations in ecological factors and vegetation composition that are undetected in the field, the result of biological processes (competition and predation favoring certain species) or the impact of direct (management) or indirect (pollution, eutrophication, etc.) effects of present and past human activities. However, a lot of example of vegetation communities evolving to the expected climax exist thanks to a better understanding of these successions (Loidi and Fernández-González, 2012; Prach et al., 2016).

Accepted author version

The most important prerequisites to map PNV are therefore (Loidi and Fernández-González, 2012):

1. a PNV typology that is sufficiently validated with quite exclusive biotopes belonging to each PNV,
2. a validated mapping of the different biotopes belonging to each PNV
3. a precise mapping of ecological factors which are supposed to play an essential and proximal role in biotope distribution.

PNV mapping also benefited from the development of distribution modeling approaches and multiple examples of PNV modeling exist (Reger et al., 2014; Somodi et al., 2017). Recent studies addressed different topics such as the best modeling approach (Hemsing and Bryn, 2012), the representation of multiple overlapping modeled PNV (Somodi et al., 2017), the definition of PNV (De Keersmaeker et al., 2013) or the modeling of numerous PNV at larger (Liu et al., 2009) and smaller scales (Longcore et al., 2018). When PNV and related vegetation communities are well identified, they are mainly based on national classification systems and models are calibrated on stable/climax forest or open areas stages reference data (Hemsing and Bryn, 2012). Building PNV based on species assemblage using floristic occurrence data is far more complex and raise many methodological questions (De Keersmaeker et al., 2013). The presence of characteristic species is highly dependent of natural sites condition and some PNV may not have any highly characteristic species (De Keersmaeker et al., 2013). The integration of multiple PNV predictions into a single map allows to better apprehend such useful results but current aggregation method are complex (Somodi et al., 2017).

Accepted author version

On the other hand, single biotope modeling – i.e., peat bogs, calcareous grasslands, moors, etc. – is still being applied (Horvath et al., 2019). In the literature, PNV concept seems to be a closed study field as many biotopes modeling study cases do not even mention PNV concept (Álvarez-Martínez et al., 2018; Horvath et al., 2019).However and as mentioned earlier, modeling PNV present many advantages to produce accurate and useful prediction map of biotopes.

In this paper, we propose a methodological framework which aims to be as most replicable as possible using biotope classification systems and mapping used in all European countries. In contrast to last PNV modeling applications using last stable forest stages, we suggest modeling PNV using multiples biotopes from the same vegetation succession to better cover the ecological niches of PNV in disturbed landscapes. Our methodological framework also proposes a simpler aggregation technique to deliver a unique map of all PNV considering also potential overlaps. We will also compare PNV modeling approach to individual biotope modeling and assess performance of both approaches. Finally, we also suggest an independent qualitative assessment of modeled PNV distributions using floristic occurrence data.

2. Materials and methods

2.1. Study area

Wallonia is an administrative region covering southern Belgium (17 000 km^2). It is characterized by a temperate climate with precipitation distributed throughout the year. However, the precipitation can double within the territory (from 800 to 1600 mm/year) and its rate increases with altitude. The summit of the region is a high peaty plateau that culminates at 694 m above sea level. The Walloon region can be split into six principal

Accepted author version

parallel biogeographic zones that are well defined by a large gradient from recent to oldest geological substrates going from NW to SE. This explains a wide diversity of very different ecological contexts, ranging from boreo-alpine peatlands to Mediterranean limestone lawns less than 30 km away (Dufrêne and Legendre, 1991).

2.2. Methodological framework

The proposed approach was carried out in four steps:

1. Biotopes developing in the same ecological context were identified and referred to one specific PNV by using historic and recent expert knowledge to maximize the difference between PNV ecological contexts.
2. PNV distributions were individually modeled using environmental predictors including topographic information, soil conditions and climatic parameters.
3. To obtain a unique typology for the territory, possible PNV prediction maps were used as predictors to produce a classification to test the potential presence of the different PNV.
4. Finally, overlaps between individually modeled PNV distributions were identified. These mixed PNV areas allowed us to identify where the proposed classification is more uncertain.

The general workflow of our method is summarized in Figure 1.

Accepted author version

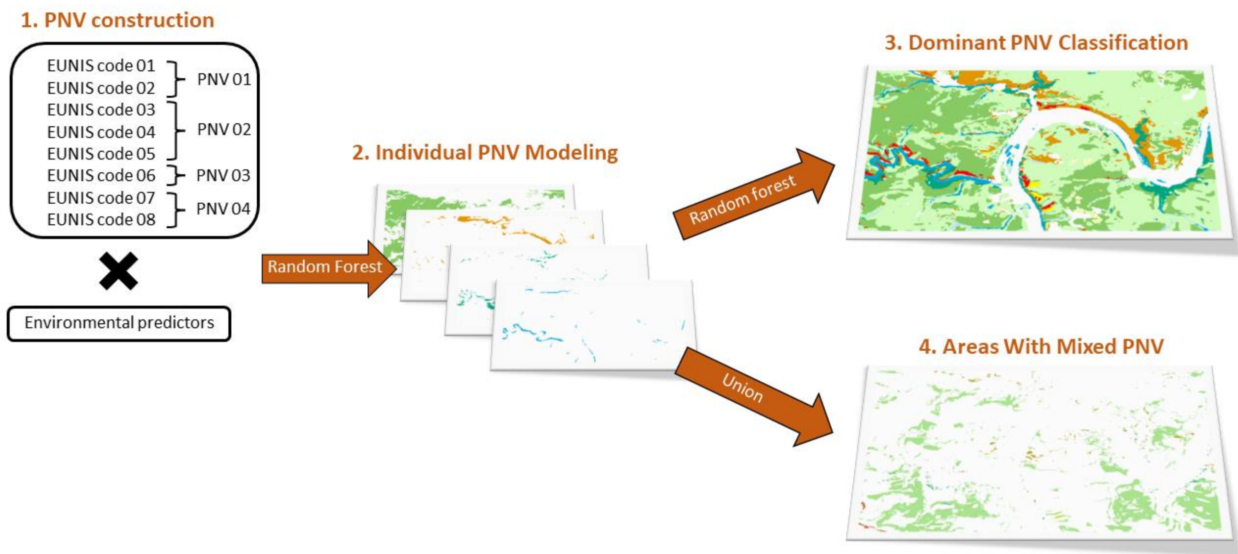

Figure 1. Schematization of the general workflow for the proposed approach with the different results. 1. Biotopes are attributed to defined PNV. 2. Individual PNV are modeled. 3.These individual predictions are used to produce a classification map. 4. Individual prediction maps are overlapped to identify potential areas with multiple PNV predicted.

2.3. PNV construction

The biotope classification and denominations used in this study are based on the EUNIS (European Nature Information System) classification for biotope types (Davies et al., 2004). EUNIS is a hierarchical classification describing every biotope and habitat present in Europe, from natural to artificial, and covering terrestrial, freshwater and marine ecosystems. The EUNIS classification was adapted to Wallonia by the public research agency in order to match regional variations (Dufrêne and Delescaille, 2005). Each biotope is characterized by a unique code.

The PNV modeled in this study were identified to cover almost all ecological contexts of the region using phytosociological literature (Delescaille L.-M. et al., 2021) complemented by expert opinion. The region has historically hosted several famous authors working on phytosociology and vegetation successions, who precisely described the different vegetation stages, their dynamics and links with ecological contexts making expert opinion highly robust (Duvigneaud, 1949; Herbauts and Tanghe, 1987; Noirfalise, 1984; Venseveren, 1969). Common and scarce biotopes were selected if there was sufficient data to model them and if they belong to only one vegetation succession. Some open biotopes such as sub-montane *Vaccinium* and *Calluna* heaths can evolve toward two different climax stages (medio-European acidophilous *Fagus* forests and medio-European thermophile acidophilous *Quercus* forests) and were therefore discarded. Biotopes corresponding to intensive agricultural practices or exotic tree plantations were also eliminated as they are not linked to a particular ecological condition. All biotopes and PNV assignations are compiled in Table 1. For further details, Appendix A presents all biotopes belonging to a PNV even if they are present in multiple PNV.

Table 1. List of biotopes considered in this study and their related PNV. Biotopes are referred to by their EUNIS classification adapted to Wallonia and Habitats Directive classification (asterisks indicate habitat of priority European interest). PNV are sorted by a humidity gradient from marshy to xeric.

Accepted author version

EUNIS Code	Habitats Directive code	Biotopes names	Potential Natural Vegetation	Code	Area mapped in reference dataset (ha)
D1.1	**7110***	Raised bogs	***Sphagnum Betula* woods**	**SB**	**485**
G1.51	**91D0***	*Sphagnum Betula* woods			
D5.21e		Beds of large *Carex spp.*	***Alnus* swamp woods**	**AS**	**229**
G1.4		Broad-leaved swamp woodland not on acid peat			
F4.11b	**4010**	Wet heathland with *Vaccinium* and *Erica tetralix*	***Quercus* and *Betula* forests with *Molinia***	**QBM**	**1 510**
F4.13		*Molinia caerulea* wet heath			
G1.81	**9190***	Atlantic *Quercus robur – Betula* woods			
G1.911a	**9190***	*Betula* facies of *Quercus robur* forests			
F9.12	**91E0***	Lowland and collinear riverine *Salix* scrub	**Riparian and gallery woodland**	**RG**	**1 087**
G1.1	**91E0***	Riparian and gallery woodland, with dominant *Alnus, Betula, Populus* or *Salix*			
G1.2	**91E0*, 91F0**	Mixed riparian floodplain and gallery woodland			
G1.A15a	**9160**	Famennian *Quercus – Carpinus betulus* forests on schist	**Fammenian *Quercus* and *Carpinus* forests**	**FQC**	**9 482**
G1.A1ba		Atlantic neutrophile *Quercus* and *Fraxinus* forests on hydromorphic soils	**Neutrophile *Quercus* and *Fraxinus* forests on wet soils**	**NQF**	**3 231**
G1.A1da	**9160**	Sub-Atlantic neutrophile *Quercus* and *Fraxinus* forests on hydromorphic soils			
G1.A1aa		Atlantic acidocline *Quercus* and *Carpinus* forests on hydromorphic soils	**Acidophilous *Quercus* and *Carpinus* forests on wet soils**	***AQC***	**38**
G1.A1ca	**9160**	Sub-Atlantic acidocline *Quercus* and *Carpinus* forests on hydromorphic soils			
G1.63	**9130**	Medio-European neutrophile *Fagus* forests	**Neutrophile *Fagus* forests**	**NF**	**6 055**
G1.A1bb	**9130**	Atlantic neutrophile *Quercus* and *Fraxinus* forests substitute to *Fagus*			
G1.A1db	**9130**	Sub-Atlantic neutrophile *Quercus* and *Fraxinus* forests substitute to *Fagus*			
G1.A41b	**9180***	*Acer* and *Ulmus* Ardennes forests	**Wet and shady ravine forests**	**WSR**	**640**
G1.A41a	**9180***	*Acer* and *Tilia* forests with *Asplenium scolopendrium*			
G1.61	**9110**	Medio-European acidophilous *Fagus* forests	**Acidophilous *Fagus* forests**	**AF**	**42 199**
G1.62	**9120**	Atlantic acidophilous *Fagus* forests			
G1.82	**9120**	Atlantic acidophilous *Fagus – Quercus* forests			
G1.87a		Medio-European acidophilous non-thermophilic *Quercus* forests			
G1.A1ab	**9120**	Atlantic acidocline *Quercus* and *Fraxinus* forests substitute to *Fagus*			
G1.A1cb		Sub-Atlantic acidocline *Quercus* and *Fraxinus* forests substitute to *Betula*			
E1.26	**6210***	Sub-Atlantic semi-dry calcareous grassland	**Calcareous *Fagus* and *Quercus* forests**	**CFQ**	**2 443**
E1.27	**6210***	Sub-Atlantic very dry calcareous grassland			
G1.66	**9150**	Medio-European limestone *Fagus* forests			
G1.71		Western *Quercus pubescens* woods and related communities			
G1.A17	**9150**	Sub-Atlantic calciphile *Quercus – Carpinus betulus* forests			
G1.87b		Medio-European thermophile acidophilous *Quercus* forests	**Thermophile acidophilous *Quercus* forests**	**TAQ**	**67**

2.4. Environmental dataset

Belgium has the particularity to be covered by a large diversity of environmental datasets at high resolution. For instance, the country is covered by a soil dataset with detailed information on soil structure, composition and humidity at a precise scale (1/10 000) (Bah et al., 2007). More modern tools like digital elevation models based on LIDAR (light detection and ranging) are also available at high resolution (1m). These precise environmental informations make it possible to better capture the relationship between environmental predictor and vegetation assemblages than it could be with generic and larger scale dataset such as Soilgrids (Poggio et al., 2021).

The environmental variables used as predictors to model PNV distributions were split into three categories: climatic variables, soil dependent variables and topographic variables.

Bioclimatic variables were extracted from high-resolution (1 km²) climate data derived from the downscaling of EURO-CORDEX regional climate model (RCM) datasets representing the historical climate between 1971 and 2005 and covering all Europe (De Troch et al., 2020). EURO-CORDEX RCM is considered as more spatially coherent at fine-scale than widely used models such as Worldclim or Climate-EU (Chakraborty et al., 2021). This dataset proposes a subset of best climatic predictor for species distribution modeling and integrate original interesting climatic variable such as annual mean potential evapotranspiration.

Accepted author version

Soil related variables were principally extracted from the digital soil map of Wallonia (Bah et al., 2007; “Carte Numérique des Sols de Wallonie - Série,” 2015). Different soil types (Calcareous, Podzol, Organic, Sandy, Alluvial and Source) were stored as binary variables. Natural soil drainage (Drainage), hydric level (Hydric_lvl) and trophic level (Trophic_lvl) were extracted as ordinal variables. Natural drainage was subdivided into 5 ordinal classes directly derived from the soil –map (dry and very dry soils; moderately dry and wet soils; wet soils with temporary water table; very wet soils with temporary water table; very wet and peaty soils).

Two variables were derived from a dataset created to guide sylvicultural practices (Petit S. et al., 2017). Hydric level (Hydric_lvl) is a variable evaluating the humidity level of a soil based on the combination of natural drainage, soil texture and topographic position. Trophic level (Trophic_lvl) is also a variable quantifying the amount of nutrients available to the plants based on a dichotomous key using soil characteristics. These cartographic data are available online in open access (“Forestimator,” 2021).

Soil texture (Sand and Silt) proportions present in the soil were not available on maps but from another data source by performing a convolution between texture information from the Walloon digital soil map and the kriging of more accurate and latest field observation points (D’Or Dimitri, 2021). These are continuous variables with the proportion of sand and silt measured at a 50 m resolution.

Finally, topographic continuous variables included elevation (Elevation), slope percentage (Slope_prc), topographic position index calculated in a 200 m radius normalized from 0 to 100 (TPI) and the potential incident light energy in W/m² for the first

Accepted author version

day of spring (SunSpring). These predictors are derived from 1m LIDAR acquisition resampled at 10m and produced for the study purpose (“Relief de la Wallonie - Modèle Numérique de Surface (MNS) 2013-2014 – Hillshade,” 2015). The radiative subsector (Radiative_SS) and the potential water intake (Water_int) are categorical variables constructed with dichotomous keys based on relative position, slope and exposure to guide sylvicultural practices (Petit S. et al., 2017). They are also available online in open access (“Forestimator,” 2021).

Predictors used in this study have different resolutions ranging from 1km for climate data to 10 m for topographic ones. We chose to resample all predictors to 10m resolution to match the finest resolution as it is recommended by the literature as the better solution to perform accurate models (Moudrý et al., 2023)..

Autocorrelation was tested to conserve independent variables for modeling. Elevation and all climatic variables except the annual variation of precipitation (AnnualVariationPrecipitation) showed a Pearson correlation coefficient higher than 0.7 or lower than –0.7 and were considered to be strongly correlated (Ratner, 2009). But bioclimatic variables are considered of high biological significance influencing plant and species distribution (Bede-Fazekas and Somodi, 2020). They were summarized by using the two main principal component axes resulting from a principal component analysis (PCA) (Climatic_pca_1 and Climatic_pca_2) to limit correlation but keep the maximum amount of information (88% and 10% of explained variance). This use of multivariate analysis to create a composite predictor has already been used and give convincing results in species and ecosystem distribution modeling (Ejrnæs, 2000; Santos et al., 2020; Simensen et al., 2020).

Accepted author version

Normality of each variable was tested, and values were transformed if necessary. Slope percentage and annual variation of precipitation were transformed with a logarithmic function. In total, 19 environmental predictors were kept for modeling (Table 2).

Table 2. List of environmental predictors used in this study and their descriptions.

Accepted author version

Predictor alias	Description	Unit	Range	Native Resolution	Transformation	Source
Climatic_pca_1	First principal component axis coordinate of climatic predictors and elevation	-	-500 – 709	1000 m (climatic) – 1 m (elevation)	-	(De troch et al., 2020)
Climatic_pca_2	Second principal component axis coordinate of climatic predictors and elevation	-	-358 – 336	1000 m – 1 m (elevation)	-	(De troch et al., 2020)
AnnualVariationPrecipitation	Mean annual variation of precipitation between 1971 and 2005	Millimeter	7-23	1000 m	Log(x+1)	(De troch et al., 2020)
Calcareous	The presence of calcareous soil	-	0-1	75 m	-	(Bah et al., 2007)
Podzol	The presence of podzolic soil	-	0-1	75 m	-	(Bah et al., 2007)
Organic	The presence of organic (peaty) soil	-	0-1	75 m	-	(Bah et al., 2007)
Sandy	The presence of sandy soil (very high proportion of sand)	-	0-1	75 m	-	(Bah et al., 2007)
Alluvial	The presence of alluvial soil	-	0-1	75 m	-	(Bah et al., 2007)
Source	The presence of source related soil	-	0-1	75 m	-	(Bah et al., 2007)
Drainage	Classes of natural soil drainage: Dry and very dry soils; Moderately dry and wet soils; Wet soils with temporary water table; Very wet soils with temporary water table; Very wet and peaty soils	-	1-5	75 m	-	(Bah et al., 2007)
Hydric_lvl	Includes 10 categories of soil humidity from xeric to marshy soils + 3 categories of alternative water regime (moving water table)	-	1-13	10 m	-	(Petit S. et al., 2017)
Trophic_lvl	Categories of trophic level from oligotrophic soils to carbonate soils	-	1-6	10 m	-	(Petit S. et al., 2017)
Sand	Proportion of the sandy texture found in soil sample extrapolated with kriging	-	0-0.88	50 m	-	(D'Or Dimitri, 2021)
Silt	Proportion of the silty texture found in soil sample extrapolated with kriging	-	0-0.82	50 m	-	(D'Or Dimitri, 2021)
Slope_prc	The slope percentage	-	0-50	10 m	Log(x+1)	Created for the study purpose based on LIDAR data
TPI	Topographic position index. The relative position in terms of elevation of the pixel compared to other pixels in a 200 m radius	-	0-100	10 m	-	Created for the study purpose based on LIDAR data
SunSpring	The potential incident light energy on the pixel for the first day of spring	W/m²	16-3434	10 m	-	Created for the study purpose based on LIDAR data

Accepted author version

Radiative_SS	Categories of different radiative sub-sectors identified by slope and exposure: 1 = cold sub-sector (from 285° to 125° oriented slopes and valley bottom); 2 = neutral sub-sector (plains, plateaus and gentle slopes); 3 = hot sub-sector (from 125° to 285° oriented slopes)	-	1-3	10 m	-	(Petit S. et al., 2017)
Water_int	Categories of water intake: 1= areas without lateral water inflows (convex plateaus and slopes); 2 = areas with variable water inflows (bottom of slopes, flats, valleys and concave areas); 3 = areas with permanent water inflows (areas connected to the hydrographic network, essentially alluvial areas)	-	1-3	10 m	-	(Petit S. et al., 2017)

Accepted author version

Areas not described in soil maps, i.e., urban areas, watercourses and parts of military domains, were not considered in the model. They appear as 'no data' on maps presented in this paper.

2.5. Vegetation data

As part of the implementation of the European Union Council Directive 92/43/EEC on the conservation of natural habitats and of wild fauna and flora (Loidi, 1999), detailed mapping of biotopes has been carried out by the different member states. Field biotope inventories realized in the Walloon region between 2005-2020 as part of the implementation of this directive were used as a reference dataset for PNV. Biotope limits were identified as polygons mapped at 1:10.000 and each of them is characterized by a EUNIS code. This field mapping is performed using dichotomous keys based on vegetation structure and species assemblage to identify EUNIS biotopes. Limits of biotopes are drawn using national geographic maps to differentiate open areas from forest and digital soil maps are used to refine limits identified on the field. Precise biotope mapping is not published and was provided directly by the public service in charge (SPW/DGARNE/DEMNA). As natural vegetation rarely follows tangible limits, the change from one biotope to another is not always a sharp edge but rather a continuous local gradient (Kuchler, 1973). Transitions and mosaics between biotopes can also be complicated and difficult to map. Field survey agents address this problem by identifying and mapping complexes of several biotopes which are intertwined (Dufrêne and Delescaille, 2007; Hearn et al., 2011). Those complexes may negatively affect a model's accuracy by decreasing strong differences between ecological conditions of biotopes, and were therefore discarded (Horvath et al., 2019). We therefore only selected

Accepted author version

polygons containing a unique EUNIS code that could be linked to a PNV (following Table 1), covering 65 599 ha.

To limit the imprecision of biotope and ecological variable mapping, we shrank biotope polygons greater than 1 ha corresponding to more widespread PNV with a 20 m inner buffer. Fammenian *Quercus* and *Carpinus* forests, neutrophile *Fagus* forests, acidophilous *Fagus* forests and calcareous *Fagus* and *Quercus* forests are the most widespread biotopes in Wallonia (Delescaille L.-M. et al., 2021). The area covered by theses PNV potentially affected by errors and polygon limit inaccuracies is not negligible compared to scarcer biotope areas which are more typical of extreme ecological conditions. The rarest biotopes have been subject to more precise mapping because nearly all are biotopes of community interest (*sensu* the European Habitat Directive defining aims of Natura 2000 network) (Loidi, 1999). They are also easily delineable on the ground. This step of shrinking the most widespread biotopes improved the quality of scarcer biotope models.

Modified biotope polygons were then used to link each PNV to every cell of the environmental predictor raster stack that falls into them. To reduce the spatial autocorrelation effect, we applied a random spatial sampling to select a maximum of 50 pixels per km² for each PNV. That reduced the dataset to approximately 200 000 rows (5% of the complete dataset). This sampling helped to initially balance the distribution of PNV occurrences by limiting those dominating the landscape. The evolution of PNV areas following the different sampling steps is available in Appendix B.

Accepted author version

2.6. *Individual PNV modeling*

A first modeling step was performed for each PNV with presence/absence data. The presence of a PNV was attributed to each pixel where an EUNIS habitat associated to the PNV have been mapped on the field. All other EUNIS habitat associated to other PNV were considered as absence (Horvath et al., 2019; Maggini et al., 2006). We could have used presence only models such as Maxent (Álvarez-Martínez et al., 2018) but the way PNV were defined with unambiguous biotopes related to specific ecological context made it possible to perform robust presence/absence models.

Models were calibrated through a Random Forest algorithm using the randomForest R package (Breiman, 2001). Random forest algorithm is recognized for its accuracy and its computational efficiency. It can easily handle small sample size and high dimension models which make it perfect to model scarce PNV (Biau and Scornet, 2016). Random forest is also known to be less sensitive to overfitting when calibrated with large sample size which is interesting in our case for common PNV (Belgiu and Drăguţ, 2016). Finally, Random forest performs for both regression and classification that make it versatile enough for our methodological framework achieving individual modeling and global classification without complexifying the method (Biau and Scornet, 2016).

The model was calibrated using 70% of the selected presence/absence dataset. On this calibration dataset, a SMOTE (synthetic minority oversampling technique) algorithm was applied to balance presence/absence frequency without losing too much information (Chawla et al., 2002). The SMOTE algorithm decreases dominant class (generally absences) and artificially increases the minority class (generally presences). Before modeling, a variable selection was performed on the calibration dataset using VSURF

Accepted author version

(Variable Selection Using Random Forest) (Genuer et al., 2015). All variables selected at the prediction step of the VSURF algorithm were kept for modeling. Random Forest model tuning focused on the mtry argument – i.e., the number of variables randomly sampled as candidates at each split – as it is considered to be the most influential parameter (Probst et al., 2019). Multiple tests were performed to choose the best mtry value giving low error and less computational time. These tests resulted in the use of the square root of the number of variables kept after selection, which corresponds to what is found in the literature (Probst et al., 2019).

Model results were first validated using the 30% remaining test set. The area under the curve (AUC) was used to evaluate model robustness using this test set. A second validation was performed on the 95% dataset discarded by spatial sampling to measure the accuracy of predictive maps. Different accuracy metrics were calculated in order to compare predictive maps and assess their adequacy: the overall accuracy (OA), the producer accuracy (PA) and the user accuracy (UA). OA is the proportion of well-classified pixels (presence and absence) compared to the total number of pixels used for validation. PA is the proportion of correctly classified reference pixels compared to the total number of reference pixel that represent the point of view of the cartographer evaluating the model prediction. UA is the proportion of well-classified presence pixels compared to the total number of predicted presences. This evaluates the model performance to predict PNV presence from the point of view of the map user. For these threshold-dependent metrics (OA, PA and UA), the threshold maximizing the F-score was used to transform model probability to presence/absence. The F-score is the harmonic mean of the receiver operating characteristic (ROC) curve. The five most

Accepted author version

important environmental predictors for the Random Forest model and their importance for the model were also identified based on the Gini index (Aldrich, 2020).

2.7. Dominant PNV classification

A second modeling step was performed using results of each PNV prediction as an input for a new Random Forest classification to produce a unique classification map of PNV. Results of specific PNV probabilities cannot be compared to identify what the dominant PNV is for a pixel. Specific PNV probabilities are average results from multiple specific classification trees with specific distributions of values. However, they can be used as variables for a new classification that will identify the most likely PNV for each pixel based on PNV probabilities.

The SMOTE algorithm can only be performed on a two-class dataset. To balance the different PNV frequencies before modeling, we decided to weight PNV classes as this is the recommended approach to balance a dataset for Random Forest classification of more than two classes while keeping all the information (Chen and Breiman, 2004). We weighted PNV classes according to their prevalence by applying Equation1 to each class:

Equation 1.

$$Weight = (1 - freq_{class})^2$$

Where $freq_{class}$ is the proportion of the class – i.e., the PNV.

Accepted author version

This helped to give more weight to less-represented PNV. For instance, the weight of the rare xerophile Fammenian *Quercus* and *Carpinus* forests is 0.99 while the weight of the common acidophilous *Fagus* forests is 0.53. The same OA, PA and UA metrics used for the specific PNV modeling were calculated. We also calculated the Cohen's kappa coefficient that measures the global accuracy of a classification comparing the proportion of well-classified pixels to the total number of pixels compared to a random classifier. To interpret Cohen's kappa coefficient results, we used the classification proposed by Landis and Koch (1977) that consider a kappa value below 0.2 as slight, between 0.21 and 0.40 as fair, between 0.41 and 0.60 as moderate, between 0.61 and 0.80 as substantially good and above 0.81 as an almost perfect classifier.

2.8. Identification of areas with mixed PNV

As previously explained, biotope distributions do not follow tangible and exclusive limits. In the field, gradients and complexes of multiple biotopes are common where biotope environmental niches overlap. Dominant PNV classification attributes to each pixel the dominant PNV even in less ecologically specific areas where multiple PNV could overlap. In order to get closer to reality and identify less certain classified areas, we chose to identify areas where several PNV are in competition. To do so, predictive distributions of each individual PNV were transformed to binary presence/absence maps using the threshold maximizing F-score. These binary maps were therefore all added together and areas where the presence of multiple PNV overlapped were considered as mixed areas. The area of each type of PNV present in mixed areas has been measured and compared to the total area of the concerned PNV. If the mixed area represented more than 10% of the total surface of one PNV in question, it was kept; those below

Accepted author version

10% were discarded. This allowed most PNV associated with important confusion after classification to be kept, discarding small and irrelevant confusion issues. All retained mixed areas were added to the PNV classification map as an overlay. Then we recalculated accuracy metrics for each PNV with validation data by considering the potential presence of the PNV in single and mixed areas. Mixed areas were therefore considered as a presence for the validation of each PNV concerned.

2.9. Complementary analysis

2.9.1. Individual biotope modeling vs. PNV

To identify gains or losses of using the PNV approach compared to individual biotope modeling that consider biotopes independent as it is commonly applied (Horvath et al., 2019; Maggini et al., 2006), we also performed modeling of biotopes composing the PNV for calcareous *Fagus* and *Quercus* forests. We chose this PNV as it includes several biotopes of different succession stages that are exclusive to the PNV. It is also the only one with sufficient data about open and semi-open biotopes to enable accurate and robust individual models.

Biotopes belonging to calcareous *Fagus* and *Quercus* forest succession were modeled considering the targeted biotope as presences and all other biotopes of the dataset – i.e., also other biotopes belonging to the same PNV – as absences, as is done in individual biotope modeling approaches (Horvath et al., 2019). All independent models were used to predict presence/absence maps using the threshold maximizing F-score. To make individual biotope models and the PNV model comparable, we added together all biotopes' predicted ranges, which was compared to the PNV predicted presence range. Producer accuracy was calculated for each individual biotope – i.e., the

Accepted author version

proportion of reference pixels related to each biotope belonging to the PNV or the cumulative biotope distributions. Furthermore, each total area of the PNV predicted presence and the cumulative individual biotope predicted presence were calculated to evaluate the specificity of the model. A presence prediction that is too large can increase producer accuracy, but the model specificity may be low.

2.9.2. Quality assessment with floristic data

Indicator plant species observations were also used to complete the quality assessment of PNV predictions using an independent dataset to the one used for model calibration. A validated plant dataset compiled for the Walloon Flora Atlas (Delescaille and Delaitte, 2011) was filtered to keep only observations of indigenous plants between 2000 and 2020 and location with precision below 100 m. PNV information was then assigned to each observation of a plant species. A Chi² test was performed to measure the dependence between the presence of a plant taxa and each specific PNV, discarding mixed PNV.

The dependence is based on residual Chi² (T) calculated with Equation 2:

Equation 2.

$$T = \sum_{ij} \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}}$$

where O_{ij} is the number of observations of a taxon i in a PNV j and E_{ij} is calculated based on Equation 3:

Equation 3.

Accepted author version

$$E_{ij} = \frac{O_{i+} \times O_{+j}}{N}$$

where O_{i+} is the total number of observations for a taxon, O_{+j} is the total number of observations in a PNV and N is the total number of observations in all PNV.

Chi² test was preferred to similar metric such as Indval (Dufrêne & Legendre, 1997). Indeed, using Indval to compare species occurrence to PNV prediction should results to low values of indicative level as it depends on the proportion of species occurrence on PNV occurrences. PNV occurrences is far wider than as it is a potential prediction and could not be compared to species occurrence that are linked to actual distribution of biotopes.

3. Results

3.1. Dominant PNV classification

PNV were firstly modeled individually using the Random Forest algorithm. Accuracy metrics results and an example of resulting map are available in Appendix C.

Random Forest prediction results were then used as independent predictors to perform a classification to identify the dominant PNV for each pixel. Results of accuracy metrics (Table 3) show that global accuracy metrics for the classification are almost perfect with 0.95 and 0.91 for OA and Cohen's kappa coefficient, respectively.

Table 3. Accuracy metric results of the classification map for each PNV. Producer (PA) and user accuracy (UA) above 0.8 (considered to be highly accurate for our objectives) are highlighted in bold text. To better view trends in accuracy metrics between individual

Accepted author version

PNV modeling and dominant classification, differences of PA and UA were added. Absolute differences higher than 0.1 are highlighted in bold font.

Potential Natural Vegetation	Code	Area used for calibration (Ha)	PA	PA differences	UA	UA differences
Sphagnum Betula **woods**	**SB**	50.8	**0.898**	+0.075	**0.829**	+0.000
Alnus **swamp woods**	**AS**	141.1	**0.936**	**+0.122**	0.678	**−0.148**
Quercus **and** ***Betula*** **forests with** ***Molinia***	**QBM**	144.9	**0.912**	+0.096	**0.827**	−0.034
Riparian and gallery woodland	**RG**	337.8	**0.930**	**+0.114**	0.689	**-0.129**
Fammenian ***Quercus*** **and** ***Carpinus*** **forests**	**FQC**	64.8	**0.968**	+0.082	**0.932**	+0.023
Neutrophile ***Quercus*** **and** ***Fraxinus*** **forests on wet soils**	**NQF**	156.5	**0.936**	+0.059	**0.888**	−0.026
Acidophilous ***Quercus*** **and** ***Carpinus*** **forests on wet soils**	**AQC**	8.1	**0.868**	**+0.244**	0.731	+0.099
Neutrophile ***Fagus*** **forests**	**NF**	338.8	**0.921**	+0.046	**0.914**	−0.005
Wet and shady ravine forests	**WSR**	94.9	**0.820**	**+0.185**	0.688	+0.018
Acidophilous ***Fagus*** **forests**	**AF**	559.0	**0.959**	−0.010	**0.990**	+0.018
Calcareous ***Fagus*** **and** ***Quercus*** **forests**	**CFQ**	131.3	**0.918**	+0.066	**0.884**	+0.021
Thermophile acidophilous ***Quercus*** **forests**	**TAQ**	24.5	**0.836**	**+0.330**	0.580	**+0.137**
Xerophile Fammenian ***Quercus*** **and** ***Carpinus*** **forests**	**XFQC**	8.4	**0.856**	**+0.195**	**0.842**	+0.003

Accepted author version

If we go into further detail, we can see that the classification increases producer accuracy for almost all PNV. The PA increase is particularly important for 5 PNV out of 13 with an increase higher than 10%. The results for UA are more heterogeneous with some PNV gaining in UA and others decreasing. Two PNV particularly lost accuracy according to UA: riparian and gallery woodland and *Alnus* swamp woods. To better understand these accuracy losses, we performed a confusion matrix that highlights an important confusion between the predicted value for some PNV and reference pixels of acidophilous *Fagus* forests (Table 4).

Table 4. Confusion matrix resulting from classification of PNV. Columns concern predicted classes and rows refer to reference classes. Values in the matrix correspond to the number of pixels divided by the total number of pixels predicted in each PNV. The diagonal therefore corresponds to UA values. SB: *Sphagnum Betula* woods; AS: *Alnus* swamp woods; QBM: *Quercus* and *Betula* forests with *Molinia*; RG: riparian and gallery woodland; FQC: Fammenian *Quercus* and *Carpinus* forests; NQF: neutrophile *Quercus* and *Fraxinus* forests on wet soils; AQC: acidophilous *Quercus* and *Carpinus* forests on wet soils; NF: neutrophile *Fagus* forests; WSR: wet and shady ravine forests; AF: acidophilous *Fagus* forests; CFQ: calcareous *Fagus* and *Quercus* forests; TAQ: thermophile acidophilous *Quercus* forests; XFQC: xerophile Fammenian *Quercus* and *Carpinus* forests.

Accepted author version

		Prediction												
		SB	**AS**	**QBM**	**RG**	**FQC**	**NQF**	**AQC**	**NF**	**WSR**	**AF**	**CFQ**	**TAQ**	**XFQC**
	SB	**0.83**	0.02	0.02	0	0	0	0	0	0	0	0	0	0
	AS	0	**0.68**	0	0	0	0	0	0	0	0	0	0	0
	QBM	**0.14**	0.02	**0.83**	0	0	0	0	0	0	0	0	0	0
	RG	0	0.01	0	**0.69**	0	0	0	0	0.01	0	0	0	0
	FQC	0	0.04	0	0.05	**0.93**	0.01	0	0.01	0	0	0.02	0.02	0.09
	NQF	0	0.01	0	0.05	0	**0.89**	0.02	0	0	0	0.01	0	0
Reference	**AQC**	0	0	0	0	0	0	**0.73**	0	0	0	0	0	0
	NF	0	0.02	0	0.04	0	0.02	0.01	**0.91**	0.08	0	0.05	0.03	0.01
	WSR	0	0	0	0.01	0	0	0	0	**0.69**	0	0.02	0.02	0
	AF	**0.03**	**0.21**	**0.15**	**0.15**	0.06	0.08	**0.24**	0.05	**0.11**	**0.99**	0.02	**0.33**	0.03
	CFQ	0	0	0	0	0	0	0	0.01	**0.11**	0	**0.88**	0.02	0.02
	TAQ	0	0	0	0	0	0	0	0	0	0	0	**0.58**	0
	XFQC	0	0	0	0	0	0	0	0	0	0	0	0	**0.84**
	Total	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**

Accepted author version

This is particularly the case – in decreasing order of confusion importance – for thermophile acidophilous *Quercus* forests, acidophilous *Quercus* and *Carpinus* forests on wet soils, *Alnus* swamp woods, riparian and gallery woodland, *Quercus* and *Betula* forests with *Molinia* and wet and shady ravine forests. All these PNV are scarce except for *Quercus* and *Betula* forests with *Molinia* and riparian and gallery woodland. We can also see a notable confusion between wet and shady ravine forests and calcareous *Fagus* and *Quercus* forests. Figure 2 shows the results of classification as a unique map of PNV potential presence.

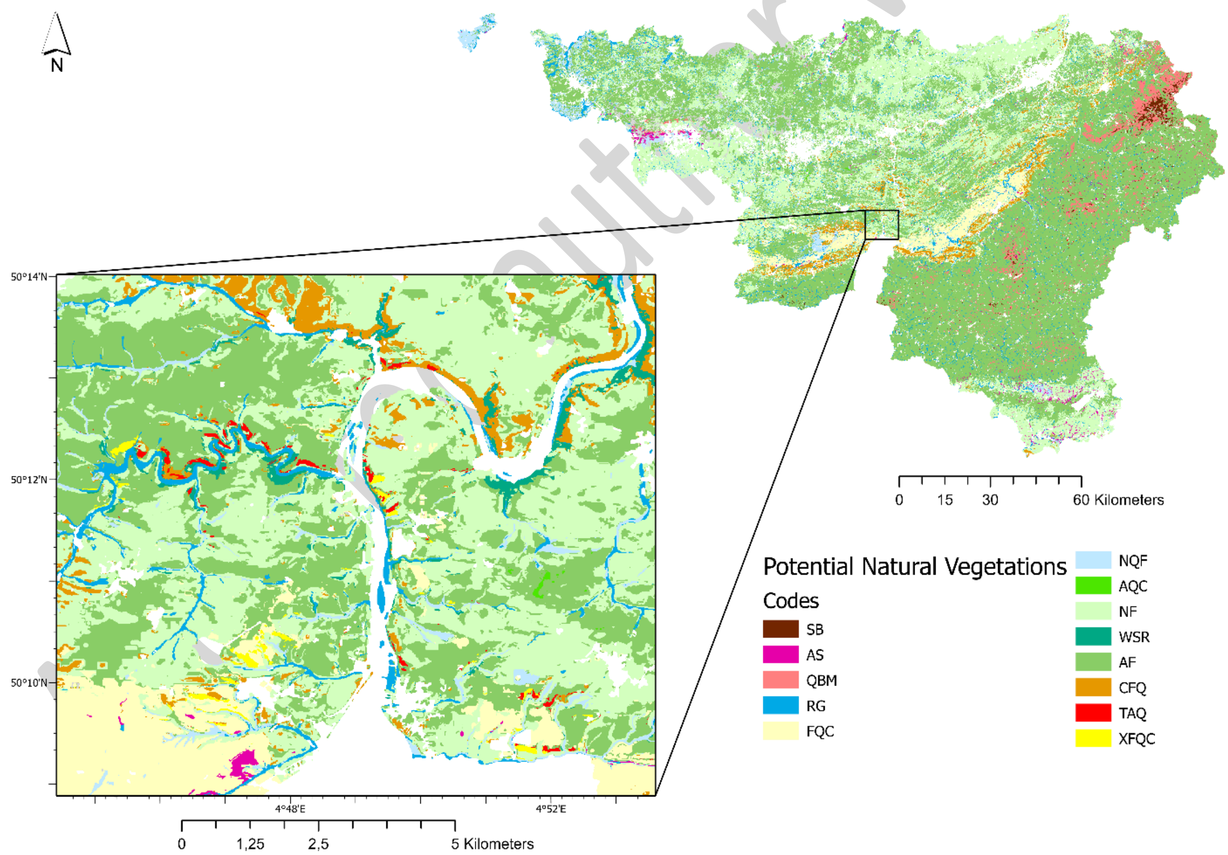

Figure 2. Result of classification map of PNV with a close-up of the Hermetton Valley region where important diversity of PNV can be seen. SB: *Sphagnum Betula* woods; AS:

author version

Alnus swamp woods; QBM: *Quercus* and *Betula* forests with *Molinia*; RG: riparian and gallery woodland; FQC: Fammenian *Quercus* and *Carpinus* forests; NQF: neutrophile *Quercus* and *Fraxinus* forests on wet soils; AQC: acidophilous *Quercus* and *Carpinus* forests on wet soils; NF: neutrophile *Fagus* forests; WSR: wet and shady ravine forests; AF: acidophilous *Fagus* forests; CFQ: calcareous *Fagus* and *Quercus* forests; TAQ: thermophile acidophilous *Quercus* forests; XFQC: xerophile Fammenian *Quercus* and *Carpinus* forests.

Results of PNV individual modeling and classification can be viewed online ("Lifewatch-FWB : UCL - Geomatics," 2022)

3.2. Identification of areas with mixed potential natural vegetation

The identification of potential zones with competition between PNV resulted to ten mixed PNV complexes where the sum of PNV proportions is higher than 10%. It appears that PNV concerned with mixed areas representing a large proportion of their presence range are generally those with a lot of overlay, as shown in the confusion matrix. This is particularly the case for scarcer PNV such as thermophile acidophilous *Quercus* forests, acidophilous *Quercus* and *Carpinus* forests on wet soils and xerophile Fammenian *Quercus* and *Carpinus* forests. There is an important overlap between neutrophile *Fagus* forests and acidophilous *Fagus* forests, involving 30% of neutrophile *Fagus* forests. This overlap is not present in the classification as there is not a notable confusion between the two PNV in the confusion matrix. Accuracy metrics were recalculated considering the potential presence of PNV in mixed areas. Global results were logically better, with a slight decrease of PA in some cases and an appreciable increase of UA for almost all PNV. OA slightly increased by 0.004 while kappa increased

Accepted author version

by 0.019. The list of mixed PNV, an example of maps with both PNV and mixed PNV such as trends in accuracy metrics with and without mixed areas are available in Appendix D.

3.3. Complementary analysis

3.3.1. Individual biotope modeling vs. PNV

Results of the modeling of the calcareous *Fagus* and *Quercus* forests PNV were compared to the remodeling of each of the biotopes that characterize the vegetation succession of this PNV. Results are presented in Table 7. We can see that PA values were low for individual biotope predictions, except for sub-Atlantic calciphile *Quercus* and *Carpinus betulus* forests. The PA values increased when biotope predictions were grouped, but the PNV approach performs better for all biotopes except for sub-Atlantic calciphile *Quercus - Carpinus betulus* forests where the difference is only 0.4%. In other words, more biotope reference pixels were encompassed in the PNV range than in the addition of individual biotope predicted presences. This represents a mean PA increase of 29% compared to individual biotope predictions and 8% compared to when they are grouped. In addition, areas of grouped biotope predictions and PNV are similar, with 30 250 ha for the PNV and 25 250 ha for the addition of individual biotopes model. The large PA for the PNV approach is therefore not due to an excessively large potential distribution.

Table 5. Producer accuracy (PA) of each biotope used to define calcareous *Fagus* and *Quercus* forests considering PNV range or addition of individual biotope models. Highest values of PA are represented in bold text.

Accepted author version

Biotopes	Biotopes (EUNIS code)	PA of individual biotope predictions	PA of individual biotope predictions inside grouped area	PA individual biotope predictions inside PNV
Sub-Atlantic semi-dry calcareous grassland	**E1.26**	0.609	**0.843**	**0.905**
Calcareous thermophilic thickets and scrub	**F3.1b**	0.488	0.710	**0.861**
Medio-European limestone *Fagus* forests	**G1.66**	0.615	0.720	**0.817**
Western *Quercus pubescens* woods and related communities	**G1.71**	0.304	0.781	**0.871**
Sub-Atlantic calciphile *Quercus* and *Carpinus betulus* forests	**G1.A17**	**0.857**	**0.863**	**0.859**

3.3.2. Quality assessment with floristic data

An additional validation of the PNV models assessed to what extent most dependent species of each PNV based on Chi² are known to be indicators of the different related biotopes. Chi² analyses were performed on plant occurrence data using a PNV classification map to identify most dependent species without considering aquatic species and very rare species (Appendix E summarized in Table 8).

Table 6. This table presents the five species with the highest values resulting from the Chi² test for each PNV, excluding aquatic and too scarce species. Species names in bold with bigger font are those typical of the associated biotopes (but not necessarily exclusive). In bold but with smaller font: companion species of the associated biotopes. Small normal font is used for generalist species. Finally, underlined species names are

Accepted author version

those not corresponding to the biotope according to their ecological requirements or corresponding to ruderal species. Exponents above the names denote if species are typical of open environments (O), develop in open and forest environments (M) or strict forest species (F).

Potential Natural Vegetation	**Five most dependent species**
Sphagnum Betula **woods**	***Vaccinium oxycoccos^{M}, Eriophorum vaginatum^{M}, Narthecium ossifragum^{O}, Eriophorum angustifolium^{O}, Drosera rotundifolia^{O}***
Alnus **swamp woods**	***Carex appropinquata^{O}, Filipendula ulmaria^{M}, Cirsium oleraceum^{M}, Caltha palustris^{M}, Triglochin palustris^{O}***
Quercus **and** ***Betula*** **forests with** ***Molinia***	***Erica tetralix^{O}, Molinia caerulea^{M}, Vaccinium uliginosum^{M}, Juncus squarrosus^{O}, Viola palustris^{M}***
Riparian and gallery woodland	***Filipendula ulmaria^{M}, Phalaris arundinacea^{M}, Caltha palustris^{M}, Iris pseudacorus^{M}, Scirpus sylvaticus^{M}***
Fammenian ***Quercus*** **and** ***Carpinus*** **forests**	***Orchis morio^{O}, Silaum silaus^{O}***, *Dianthus armeria^{O}*, ***Colchicum autumnale^{O}, Selinum carvifolia^{O}***
Neutrophile ***Quercus*** **and** ***Fraxinus*** **forests on wet soils**	*Lythrum salicaria^{M}*, ***Alnus glutinosa^{M}, Filipendula ulmaria^{M}, Glechoma hederacea^{M}***, *Phragmites australis^{M}*
Acidophilous ***Quercus*** **and** ***Carpinus*** **forests on wet soils**	*Dactylorhiza praetermissa subsp.praetermissa var.j,^{O}* *Oxybasis rubra^{O}*, ~~*Dipsacus pilosus^{F}*~~, *Lathyrus nissolia^{O}* *Senecio vulgaris^{O}*
Neutrophile ***Fagus*** **forests**	***Galium odoratum^{F}, Paris quadrifolia^{F}***, *Saxifraga granulata^{O}, Viscum album^{M}, Equisetum arvense^{O}*
Wet and shady ravine forests	***Asplenium scolopendrium^{F}, Polystichum aculeatum^{F}, Asplenium trichomanes^{M}, Mercurialis perennis^{F}***, *Biscutella laevigata^{O}*
Acidophilous ***Fagus*** **forests**	***Pteridium aquilinum^{M}, Teucrium scorodonia^{F}, Cytisus scoparius^{O}, Vaccinium myrtillus^{M}, Luzula luzuloides^{F}***
Calcareous ***Fagus*** **and** ***Quercus*** **forests**	***Helianthemum nummularium^{O}, Vincetoxicum hirundinaria^{M}, Teucrium chamaedrys^{O}, Globularia bisnagarica^{O}, Gymnadenia conopsea^{O}***

Accepted author version

Thermophile acidophilous *Quercus* forests	***Asplenium septentrionale***^{O}**, *Asplenium adiantum-nigrum***^{M}, *Filago minima*^{O}**, *Silene nutans***^{M}**, *Galeopsis segetum***^{O}
Xerophile Fammenian *Quercus* and *Carpinus* forests	***Silene nutans***^{M}, ***Dianthus carthusianorum***^{O}, ***Asplenium adiantum-nigrum***^{M}**, *Potentilla argentea***^{O}**,** *Ulmus laevisF*

Most dependent species resulting from Chi² analyses are generally indicative or related to biotopes concerned by the modeled PNV. Species characteristic of both open and forest stages of the PNV are present in the five most dependent species. For open habitats species, this is true for *Helianthemum nummularium* that is indicative of open stages of calcareous *Fagus* and *Quercus* forests or *Erica tetralix*, typical of open stages of *Quercus* and *Betula* forests with Molinia; for forest species, *Asplenium scolopendrium* that is typical of forest stages of wet and shady ravine forests or *Luzula luzuloides* that is present in forest stages of acidophilous *Fagus* forests. Two PNV had poorer results: neutrophile *Quercus* and *Fraxinus* forests on wet soils and acidophilous *Quercus* and *Carpinus* forests on wet soils. Some species of these two PNV highlighted by the analysis are not related to the expected environmental conditions.

4. Discussion

4.1. PNV modeling performance

The modeling of individual and dominant PNV showed very convincing results. The results of accuracy metrics were good to excellent with OA and kappa values were extremely high. These results can be explained by the contrasted definition of PNV related to the main ecological context of the study area. The diversity of specific ecological conditions covered by the different PNV would favor the prediction of true absences increasing OA. Other studies modeling biotope distributions have also found better accuracy metric values for the rarest biotopes (Horvath et al., 2019; Mi et al.,

Accepted author version

2017; Zhang et al., 2020). However, these studies only used the AUC, which is also influenced by true presence and absences, as do many studies dealing with single biotope models (Horvath et al., 2019; Maggini et al., 2006) or PNV modeling (Hemsing and Bryn, 2012). However, using true absence rate may lead to a false assumption that model results are good and especially concerning scarce biotope/PNV. Indeed, for scarce biotope/PNV depending on specific ecological context, class imbalance between presence and absence data will favor true absence rate. PA and UA metrics based principally on true presence showed that results are more nuanced, and that attention must be paid with some predicted PNV.

To model PNV, we assumed a deterministic evolution of vegetation succession toward a climax stage, which is not agreed upon by all the scientific community (Chiarucci et al., 2010). We therefore paid attention to PNV definition and identifying biotopes belonging to each PNV. Our results showed a good distinction between the different PNV, which proves they were well ecologically defined and that our hypothesis about a deterministic evolution of vegetation succession could be validated. Furthermore, these results underline the benefits and sturdiness of a prior definition of PNV compared to more complex methods where PNV are constructed based on vegetation occurrence datasets (De Keersmaeker et al., 2013). However, the existence of mixed areas shows some important overlaps between independent PNV predictions. Nevertheless, modeling PNV has many advantages presented in this paper that brings new light to this old concept.

Classification results to identify dominant PNV had an important global PA increase at the expense of UA, which shows that the modeled distribution of PNV had been widened after classification. This is especially the case for riparian and gallery

Accepted author version

woodlands and *Alnus* swamp woods. A detailed analysis of the confusion matrix performed on classification results revealed a significant confusion between acidophilous *Fagus* forests and scarcer PNV. Acidophilous *Fagus* forests is the largest of all PNV representing 64% of reference data and its high prevalence in the validation dataset could also explain why prediction errors are more widely detected in widespread biotopes. Even if areas of confusion between acidophilous *Fagus* forests and other PNV are small and rare compared to acidophilous *Fagus* forests' total area, the impact on accuracy metrics may still be significant for the rarest PNV for which less validation data is available. This confusion can be explained by different factors.

Both predictor maps (such as soil maps) and biotope data used for calibration are derived from human interpretation, which means that the representation and simplification on a cartographic medium contributes to confusion between PNV.

Secondly, *Fagus* forests are the natural dominant vegetation on the Walloon territory and can sometimes extend at the border of their range. This effect is accentuated by *Fagus sylvativca* competitiveness toward other species, homogenization of stands by silvicultural management and high game pressure on less common tree species.

Thirdly, biotopes present in scarce PNV depend to their specific ecological conditions that have a spatial pattern defined by topography and hydrography generating many edges. Therefore, the extraction of rasterized environmental predictors of 10 m wide may have created border effects where predictor values of specific ecological conditions may be attributed to the scarce PNV or to the surrounding common PNV. This issue has

Accepted author version

been considered to define the reference data set and its impact limited by removing a 20 m inside buffer from common PNV polygons.

Fourthly, more common biotope maps are susceptible to mask scarcer ones. As the mapping of common biotopes may be coarser, they can include smaller areas with ecological conditions suitable to other PNV. A detailed analysis of the acidophilous *Fagus* forests reference data showed that a bigger proportion of riparian and gallery woodlands reference polygons are on wet alluvial soils (43%) than that of acidophilous *Fagus* forests (1%) but the latter has a greater absolute area of wet alluvial soil (83 398 ha) than riparian and gallery woodlands (47 972 ha). This issue was anticipated by balancing calibration datasets; firstly with spatial sampling and secondly by weighting PNV before classification. By giving more importance to scarce PNV we also gave more importance to biotope polygons for which we can have better confidence as more attention was given to them during their mapping.

4.2. Individual biotope versus PNV modeling

The analysis of the independent biotope models of calcareous *Fagus* and *Quercus* forests showed that PNV approach better covers the existing biotope distributions compared to single biotope modeling. Furthermore, PNV modeling using only climax stage vegetation as reference would have used Medio-European limestone *Fagus* forests (Delescaille L.-M. et al., 2021). However, we can see that using this vegetation alone to model the PNV leads to missing 20% more cells of this biotope than our approach. For this example, using several biotopes of the succession stage to calibrate the PNV better encompass ecological niche and leads to better predictive results.

Accepted author version

The better results of our approach can be explained by the fact that open biotopes are maintained by human management. In our case, only a part of them have resumed their succession and another part had their natural succession blocked or brought back to the starting point by human activities. Our hypothesis is that human activity restricts the distribution of open biotopes to more extreme local ecological contexts that could be only a limited part of the actual ecological niche of the biotope. By modeling calcareous *Fagus* and *Quercus* forest PNV, we therefore better consider the ecological niche of sub-Atlantic semi-dry calcareous grassland, as evidenced by our results.

4.3. Quality assessment with floristic data

Quality control of the potential distribution of PNV using floristic data demonstrates that the most dependent PNV species are those that are found in corresponding biotope floristic assemblages (Delescaille L.-M. et al., 2021).

Most important issues revealed by this analysis concern neutrophile *Quercus* and *Fraxinus* forests on wet soils and acidophilous *Quercus* and *Carpinus* forests on wet soils. However, this can be explained by too low accuracy of available observations, poorer sampling inside the biotopes concerned and inherent PNV scarcity – especially acidophilous *Quercus* and *Carpinus* forests on wet soils – that impact Chi² analysis.

4.4. Strength and limitations of the approach

Our approach helped to tackle different issues related to biotope mapping. In our case, biotopes that are part of other vegetation series were therefore considered as true absences which improved the model's capacity to distinguish them. The use of independent models for each PNV to produce a classification as well as the distinction

Accepted author version

of the mixed zones allowed potential PNV distributions to be refined. Important PA values for classification and the small change in UA showed that this model was very powerful, avoiding omitting the presence of different PNV while keeping a good detection of true presences. Other approaches have allowed a map to be built with the most dominant biotope from multiple predictions (Álvarez-Martínez et al., 2018), but the probabilities between two concurrent biotopes were compared and the dominant biotope was assigned to the pixel if the difference of probabilities was higher than a subjective threshold. In our approach, any subjectivity was avoided through the classification based on individual PNV predictions.

Other methods of biotope modeling may also be followed. Initiated by Margules and Stein(1989), vegetation communities could also be considered based on the combined potential distribution of species belonging to the community. This approach could be adapted to aim modeling vegetation communities as we did. It has the advantage that it could now be performed on vegetation observation from naturalist platforms. These data are numerous, regularly completed and a large majority are validated. However, species observations suffer from many problems related to the sampling, quality and nature of biological data.

The data collected rarely follow a standardized methodology. They are often spatially biased in favor of the most accessible or attractive areas of a territory such as protected areas. Biotope mapping is in principle more standardized and systematic, and one can trust the absence of a biotope where another has been mapped, which is not the case for the presence of a species. Species location is also not exempt of errors of different

Accepted author version

types while biotopes delineation corresponds to a combination of regularly observed species and a particular combination of ecological factors.

But one major problem is the nature of the distribution data of species that cannot in principle share the same ecological niche that characterizes a biotope. Rare species that are pore specific of environmental conditions are very difficult to model because their absence can be explained by local ecological factors not measurable at the size of the selected grid (10x10m), biotic factors, particular adaptation strategies, factors associated with the dispersal filter, human activities, ... While most frequent species are less indicative and are therefore also generally observed in other combinations of ecological conditions.

The identification of mixed PNV areas revealed the difficult distinction and blurred delimitation between different PNV where ecological conditions are very close. These mixed areas are mainly distributed in the areas most disturbed by human activities and have fewer biotopes of reference for model calibration. These transitional areas between different biotopes reflect those typically encountered in the field. Other methods to tackle overlapping of multiple PNV predictions are complex and do not provide convenient maps but give a more complete information with different comparable probabilities for all overlapping PNV (Somodi et al., 2017). Our approach is simpler and we discard some complexes but all information is available in a single convenient map. Moreover, marginal complexes of PNV that were discarded are mainly unrealistic in term of ecological condition that should not overlap.

.

Accepted author version

4.5. Potential use of PNV map for nature protection and conservation

PNV modeling can identify possible priority areas for biotope restoration. The new European Biodiversity Strategy for 2030 and the new European regulation on nature restoration defined very ambitious objectives of surface to protect and conservation status to improve. As they concern the implementation of the Natura 2000 network and many biotopes of European interest (see Table 1), PNV modeling make it possible to assess the extent of restoration possibilities and to identify the highest priority areas for increasing existing surfaces or for improving connectivity between them.

Within a PNV, the distinction between forest and open biotopes can then be identified using available land cover data, which is becoming increasingly accurate and easily available through remote sensing approaches (Bourdouxhe et al., 2020; Radoux et al., 2019). The use of land-use and land-cover data in the PNV allows for targeted field sampling of the most potentially interesting areas to complete the mapping of the core areas of biodiversity.

The PNV concept also allows for the development of much more dynamic approaches to nature conservation to account for natural successions of vegetation rather than trying to fix the different stages in specific locations. It is also an opportunity to verify the extent to which all stages are present in the landscape.

5. Conclusion

This study developed a modeling approach that offers new possibilities for large-scale biotope modeling with higher predictive accuracy. The different datasets and methods used to evaluate the analyze strategy highlight its robustness. However, PNV modeling

Accepted author version

is not without its limitations. Care must be taken with initial definitions and significant phytosociological knowledge is recommended to ensure consistency with field methodology. As with many modeling approaches, care must also be taken in the selection of baseline data and environmental predictors to ensure that the ecological niches being modeled are distinguishable.

Funding

This work was supported by the Fédération Wallonie-Bruxelles within the Lifewatch-WB project.

Acknowledgments

The authors would like to thank the DEMNA (Département de l'Étude du Milieu Naturel et Agricole) for sharing their data and expertise.

References

Aldrich, C., 2020. Process Variable Importance Analysis by Use of Random Forests in a Shapley Regression Framework. Minerals 10, 420. https://doi.org/10.3390/min10050420

Álvarez-Martínez, J.M., Jiménez-Alfaro, B., Barquín, J., Ondiviela, B., Recio, M., Silió-Calzada, A., Juanes, J.A., 2018. Modelling the area of occupancy of habitat types with remote sensing. Methods Ecol. Evol. 9, 580–593. https://doi.org/10.1111/2041-210X.12925

Bah, B.B., Legrain, X., Engels, P., Colinet, G., Bock, L., 2007. Légende de la Carte Numérique des Sols de Wallonie - version 2. Legend of the Digital Soil Map of Wallonia - version 2.

Bede-Fazekas, Á., Somodi, I., 2020. The way bioclimatic variables are calculated has impact on potential distribution models. Methods Ecol. Evol. 11, 1559–1570. https://doi.org/10.1111/2041-210X.13488

Belgiu, M., Drăguţ, L., 2016. Random forest in remote sensing: A review of applications and future directions. ISPRS J. Photogramm. Remote Sens. 114, 24–31. https://doi.org/10.1016/j.isprsjprs.2016.01.011

Biau, G., Scornet, E., 2016. A random forest guided tour. TEST 25, 197–227. https://doi.org/10.1007/s11749-016-0481-7

Blasi, C., Filibeck, G., Frondoni, R., Rosati, L., Smiraglia, D., 2004. The map of the vegetation series of Italy. Fitosociologia 41, 21–26.

Bohn, U., Gollub, G., 2006. THE USE AND APPLICATION OF THE MAP OF THE NATURAL VEGETATION OF EUROPE WITH PARTICULAR REFERENCE TO GERMANY. Biol. Environ. Proc. R. Ir. Acad. 106B, 199–213.

Accepted author version

Bourdouxhe, A., Duflot, R., Radoux, J., Dufrêne, M., 2020. Comparison of methods to model species habitat networks for decision-making in nature conservation: The case of the wildcat in southern Belgium. J. Nat. Conserv.

Breiman, L., 2001. Random Forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324

Carte Numérique des Sols de Wallonie - Série [WWW Document], 2015. URL http://geoportail.wallonie.be/catalogue/c5bedf2b-1cac-4231-9d9a-854e0ef2c9ce.html (accessed 2.13.23).

Chakraborty, D., Dobor, L., Zolles, A., Hlásny, T., Schueler, S., 2021. High-resolution gridded climate data for Europe based on bias-corrected EURO-CORDEX: The ECLIPS dataset. Geosci. Data J. 8, 121–131. https://doi.org/10.1002/gdj3.110

Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P., 2002. SMOTE: synthetic minority over-sampling technique. J. Artif. Intell. Res. 16, 321–357.

Chen, C., Breiman, L., 2004. Using Random Forest to Learn Imbalanced Data. Univ. Calif. Berkeley.

Chiarucci, A., Araújo, M.B., Decocq, G., Beierkuhnlein, C., Fernández-Palacios, J.M., 2010. The concept of potential natural vegetation: an epitaph? J. Veg. Sci. 21, 1172–1178. https://doi.org/10.1111/j.1654-1103.2010.01218.x

Davies, C.E., Moss, D., Hill, M.O., 2004. EUNIS habitat classification revised 2004. Rep. Eur. Environ. Agency-Eur. Top. Cent. Nat. Prot. Biodivers. 127–143.

De Keersmaeker, L., Rogiers, N., Vandekerkhove, K., De Vos, B., Roelandt, B., Cornelis, J., De Schrijver, A., Onkelinx, T., Thomaes, A., Hermy, M., Verheyen, K., 2013. Application of the Ancient Forest Concept to Potential Natural Vegetation Mapping in Flanders, A Strongly Altered Landscape in Northern Belgium. Folia Geobot. 48, 137–162. https://doi.org/10.1007/s12224-012-9135-z

De Troch, R., Termonia, P., Van Schaeybroeck, B., 2020. High-resolution future climate data for species distribution models in Europe. https://doi.org/10.5281/zenodo.3694065

Delescaille, L.M., Delaitte, S., 2011. L'Atlas floristique de Wallonie : où en est-on? Adoxa 68 17–19.

Delescaille L.-M., Wibail L., Claessens H., Dufrêne M., Mahy G., Peeters A., Sérusiaux E., 2021. Les Habitats d'Intérêt Communautaire de Wallonie, Publication du Département de l'Étude du Milieu Naturel et Agricole (SPW ARNE), Série « Faune – Flore – Habitat ». Gembloux.

D'Or Dimitri, 2021. Soil texture – Ephesia Consult. URL https://www.ephesia-consult.com/portfolio/soil-texture/ (accessed 3.28.22).

Dufrêne, M., Delescaille, L.-M., 2007. Guide méthodologique pour l'inventaire et la cartographie des habitats et des habitats d'espèces dans le cadre de la réalisation des arrêtés de désignation en Région wallonne.

Dufrêne, M., Delescaille, L.-M., 2005. La typologie WalEUNIS des biotopes wallons.

Dufrêne, M., Legendre, P., 1991. Geographic structure and potential ecological factors in Belgium. J. Biogeogr. 18, 257–266.

Duvigneaud, P., 1949. Classification phytosociologique des tourbières de l'Europe. Bull. Société R. Bot. Belg. Van K. Belg. Bot. Ver. 81, 58–129.

Ejrnæs, R., 2000. Can we trust gradients extracted by Detrended Correspondence Analysis? J. Veg. Sci. 11, 565–572. https://doi.org/10.2307/3246586

Forestimator [WWW Document], 2021. URL https://forestimator.gembloux.ulg.ac.be/ (accessed 2.13.23).

Forman, R.T.T., 1995. Land Mosaics: The Ecology of Landscapes and Regions, Cambridge University Press. ed. Cambridge, UK.

Gallizia Vuerich, L., Poldini, L., Feoli, E., 2001. Model for the potential natural vegetation mapping of Friuli-Venezia Giulia (Ne Italy) and its application for a biogeographic classification of the region. Plant Biosyst. 135, 319–336.

Accepted author version

Genuer, R., Poggi, J.-M., Tuleau-Malot, C., 2015. VSURF: An R Package for Variable Selection Using Random Forests. R J. 7.

Hall, C.M., 1988. The “worthless lands hypothesis” and Australia’s national parks and reserves. Aust. Ever Chang. For. Aust. Def. Force Acad. Canberra Aust. 441–459.

Hearn, S.M., Healey, J.R., McDonald, M.A., Turner, A.J., Wong, J.L.G., Stewart, G.B., 2011. The repeatability of vegetation classification and mapping. J. Environ. Manage. 92, 1174–1184. https://doi.org/10.1016/j.jenvman.2010.11.021

Hemsing, L.Ø., Bryn, A., 2012. Three methods for modelling potential natural vegetation (PNV) compared: A methodological case study from south-central Norway. Nor. Geogr. Tidsskr.-Nor. J. Geogr. 66, 11–29.

Herbauts, J., Tanghe, M., 1987. RELATIONS ENTRE FORMATIONS SUPERFICIELLES, SOLS ET ASSOCIATIONS FORESTIÈRES SUR LA CUESTA BAJOCIENNE DE LORRAINE BELGE. L’EXEMPLE DU BOIS DE LA CÔTE À VIRTON-SAINT-MARD. Bull. Société R. Bot. Belg. Bull. Van K. Belg. Bot. Ver. 120, 93–105.

Hilty, J., Worboys, G.L., Keeley, A., Woodley, S., Lausche, B.J., Locke, H., Carr, M., Pulsford, I., Pittock, J., White, J.W., Theobald, D.M., Levine, J., Reuling, M., Watson, J.E.M., Ament, R., Groves, C., Tabor, G.M., 2020. Guidelines for conserving connectivity through ecological networks and corridors. IUCN.

Horvath, P., Halvorsen, R., Stordal, F., Tallaksen, L.M., Tang, H., Bryn, A., 2019. Distribution modelling of vegetation types based on area frame survey data. Appl. Veg. Sci. 22, 547–560. https://doi.org/10.1111/avsc.12451

IPBES, 2019. Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. Bonn, Germany.

Jones, M.E., Bain, G.C., Hamer, R.P., Proft, K.M., Gardiner, R.Z., Dixon, K.J., Kittipalawattanapol, K., Zepeda de Alba, A.L., Ranyard, C.E., Munks, S.A., Barmuta, L.A., Burridge, C.P., Johnson, C.N., Davidson, N.J., 2021. Research supporting restoration aiming to make a fragmented landscape ‘functional’ for native wildlife. Ecol. Manag. Restor. 22, 65–74. https://doi.org/10.1111/emr.12504

Kuchler, A.W., 1973. Problems in Classifying and Mapping Vegetation for Ecological Regionalization. Ecology 54, 512–523. https://doi.org/10.2307/1935336

Landis, J.R., Koch, G.G., 1977. An Application of Hierarchical Kappa-type Statistics in the Assessment of Majority Agreement among Multiple Observers. Biometrics 33, 363–374. https://doi.org/10.2307/2529786

Leguédois, S., Party, J.-P., Dupouey, J.-L., Gauquelin, T., Gégout, J.-C., Lecareux, C., Badeau, V., Probst, A., 2011. La carte de végétation du CNRS à l’ère du numérique. Cybergeo Eur. J. Geogr. https://doi.org/10.4000/cybergeo.24688

Lifewatch-FWB : UCL - Geomatics [WWW Document], 2022. URL https://maps.elie.ucl.ac.be/lifewatch/habitat.html?lang=en (accessed 2.13.23).

Lillesand, T., Kiefer, R.W., Chipman, J., 2008. Remote Sensing and Image Interpretation. Wiley.

Liu, H., Wang, L., Yang, J., Nakagoshi, N., Liang, C., Wang, W., Lv, Y., 2009. Predictive modeling of the potential natural vegetation pattern in northeast China. Ecol. Res. 24, 1313–1321. https://doi.org/10.1007/s11284-009-0616-3

Löfvenhaft, K., Björn, C., Ihse, M., 2002. Biotope patterns in urban areas: a conceptual model integrating biodiversity issues in spatial planning. Landsc. Urban Plan., Fragmentation and Land Use Planning: Analysis and beyond? 58, 223–240. https://doi.org/10.1016/S0169-2046(01)00223-7

Loidi, J., 1999. Preserving biodiversity in the European Union: the Habitats Directive and its application in Spain. Plant Biosyst. 133, 99–106.

Loidi, J., Fernández-González, F., 2012. Potential natural vegetation: reburying or reboring? J. Veg. Sci. 23, 596–604. https://doi.org/10.1111/j.1654-1103.2012.01387.x

Accepted author version

Longcore, T., Noujdina, N., Dixon, P.J., 2018. Landscape Modeling of the Potential Natural Vegetation of Santa Catalina Island, California. West. North Am. Nat. 78, 617–632. https://doi.org/10.3398/064.078.0406
Maggini, R., Lehmann, A., Zimmermann, N.E., Guisan, A., 2006. Improving generalized regression analysis for the spatial prediction of forest communities. J. Biogeogr. 33, 1729–1749. https://doi.org/10.1111/j.1365-2699.2006.01465.x
Margules, C.R., Stein, J.L., 1989. Patterns in the distributions of species and the selection of nature reserves: An example from Eucalyptus forests in South-eastern New South Wales. Biol. Conserv., Australian Developments in Conservation Evaluation 50, 219–238. https://doi.org/10.1016/0006-3207(89)90011-6
Moudrý, V., Keil, P., Cord, A.F., Gábor, L., Lecours, V., Zarzo-Arias, A., Barták, V., Malavasi, M., Rocchini, D., Torresani, M., Gdulová, K., Grattarola, F., Leroy, F., Marchetto, E., Thouverai, E., Prošek, J., Wild, J., Šímová, P., 2023. Scale mismatches between predictor and response variables in species distribution modelling: A review of practices for appropriate grain selection. Prog. Phys. Geogr. Earth Environ. 03091333231156362. https://doi.org/10.1177/03091333231156362
Niering, W.A., 1987. Vegetation Dynamics (Succession and Climax) in Relation to Plant Community Management. Conserv. Biol. 1, 287–295. https://doi.org/10.1111/j.1523-1739.1987.tb00049.x
Noirfalise, A., 1984. Forêts et stations forestières en Belgique, Les Presses Agronomiques. ed. Persée - Portail des revues scientifiques en SHS, Gembloux.
Petit S., Cordier S., Claessens H., Ponette Q., Vincke C., Marchal D., Weissen F., 2017. Fichier écologique des essences.
Poggio, L., de Sousa, L.M., Batjes, N.H., Heuvelink, G.B.M., Kempen, B., Ribeiro, E., Rossiter, D., 2021. SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty. SOIL 7, 217–240. https://doi.org/10.5194/soil-7-217-2021
Prach, K., Tichý, L., Lencová, K., Adámek, M., Koutecký, T., Sádlo, J., Bartošová, A., Novák, J., Kovář, P., Jírová, A., Šmilauer, P., Řehounková, K., 2016. Does succession run towards potential natural vegetation? An analysis across seres. J. Veg. Sci. 27, 515–523. https://doi.org/10.1111/jvs.12383
Probst, P., Wright, M.N., Boulesteix, A.-L., 2019. Hyperparameters and tuning strategies for random forest. WIREs Data Min. Knowl. Discov. 9, e1301. https://doi.org/10.1002/widm.1301
Radoux, J., Bourdouxhe, A., Coos, W., Dufrêne, M., Defourny, P., 2019. Improving Ecotope Segmentation by Combining Topographic and Spectral Data. Remote Sens. 11, 354. https://doi.org/10.3390/rs11030354
Ratner, B., 2009. The correlation coefficient: Its values range between +1/−1, or do they? J. Target. Meas. Anal. Mark. 17. https://doi.org/10.1057/jt.2009.5
Rees, M., Condit, R., Crawley, M., Pacala, S., Tilman, D., 2001. Long-Term Studies of Vegetation Dynamics. Science 293, 650–655. https://doi.org/10.1126/science.1062586
Reger, B., Häring, T., Ewald, J., 2014. The TRM Model of Potential Natural Vegetation in Mountain Forests. Folia Geobot. 49, 337–359. https://doi.org/10.1007/s12224-013-9158-0
Relief de la Wallonie - Modèle Numérique de Surface (MNS) 2013-2014 – Hillshade [WWW Document], 2015. URL http://geoportail.wallonie.be/catalogue/4ff6714d-61ae-485f-94e5-e6a2bebdf7ef.html (accessed 2.13.23).
Resasco, J., 2019. Meta-analysis on a Decade of Testing Corridor Efficacy: What New Have we Learned? Curr. Landsc. Ecol. Rep. 4, 61–69. https://doi.org/10.1007/s40823-019-00041-9
Santos, J.P., Sobral-Souza, T., Brown Jr, K.S., Vancine, M.H., Ribeiro, M.C., Freitas, A.V.L., 2020. Effects of landscape modification on species richness patterns of fruit-feeding butterflies in Brazilian Atlantic Forest. Divers. Distrib. 26, 196–208. https://doi.org/10.1111/ddi.13007

Accepted author version

Simensen, T., Horvath, P., Vollering, J., Erikstad, L., Halvorsen, R., Bryn, A., 2020. Composite landscape predictors improve distribution models of ecosystem types. Divers. Distrib. 26, 928–943. https://doi.org/10.1111/ddi.13060
Somodi, I., Molnár, Z., Czúcz, B., Bede-Fazekas, Á., Bölöni, J., Pásztor, L., Laborczi, A., Zimmermann, N.E., 2017. Implementation and application of multiple potential natural vegetation models – a case study of Hungary. J. Veg. Sci. 28, 1260–1269. https://doi.org/10.1111/jvs.12564
Stevens, J.P., Blackstock, T.H., Howe, E.A., Stevens, D.P., 2004. Repeatability of Phase 1 habitat survey. J. Environ. Manage. 73, 53–59. https://doi.org/10.1016/j.jenvman.2004.05.009
Stumpel, A.H.P., Kalkhoven, J.T.R., 1978. A Vegetation Map of the Netherlands, Based on the Relationship between Ecotopes and Types of Potential Natural Vegetation, in: van der Maarel, E., Werger, M.J.A. (Eds.), Plant Species and Plant Communities. Springer Netherlands, Dordrecht, pp. 137–147. https://doi.org/10.1007/978-94-009-9987-9_18
Tuxen, R., 1956. Die huetige potentielle naturliche Vegetation als Gegestand der Vegetationskarierung. Angew. Pflanzensozioligie 13, 5–42.
Venseveren, J.P., 1969. Étude phytosociologique de deux transects de la vallée de la Lesse (Belgique). Bull. Société R. Bot. Belg. Bull. Van K. Belg. Bot. Ver. 102, 149–164.
Woodward, J., 2009. The Physical Geography of the Mediterranean. OUP Oxford.

Accepted author version