A low-cost and repeatable procedure for modelling the regional distribution of Natura 2000 terrestrial habitats

Michele Dalle Fratte, Guido Brusa and Bruno Enrico Leone Cerabolini

Department of Theoretical and Applied Sciences, Università degli studi dell’Insubria, Varese, Italy

ABSTRACT
The present paper describes a procedure for mapping the distribution of Natura 2000 terrestrial habitats (Habitats Directive 92/43/EEC) at the regional scale (Lombardy, Northern Italy) by means of open-source software (QGIS and R). The habitat map within Natura 2000 sites was used for modelling the regional distribution of three selected habitats, by applying classification trees on freely available and fine-scale resolution environmental layers. Land use and forest type maps were combined to refine the regional distribution of selected habitats. The statistical validation showed a fairly substantial overall accuracy of predicted habitat distribution, which was used to determine the regional extent of the habitats and to evaluate the regional effectiveness of Natura 2000 network. We provide an easy and inexpensive procedure, replicable in other contexts in which just basic information on Natura 2000 terrestrial habitats are available, and usable for habitats monitoring according to the Habitats Directive.

1. Introduction

The Habitats Directive (92/43/EEC) provides the legal basis of Natura 2000 (N2K), a network of protected areas established across all Member States (MS) of the European Union (EU) for the conservation of selected targets, i.e. natural habitat types (to which we will refer from here on as ‘EU habitats’) and species listed in the Annexes of the Directive. The monitoring and the 6-yearly reporting of the conservation status of EU habitats are among the requests made by the European Commission to all the MS. These requirements involve comprehensive, reliable and updated distribution maps of EU habitats, not just inside the N2K network (Evans & Arvela, 2011). However, these kind of thematic maps are usually lacking in several EU territories, especially outside N2K network; hence, data concerning the spatial distribution of EU habitats are extremely heterogeneous (Gruber et al., 2012; Maiorano et al., 2015; Rosati, Marignani, & Blasi, 2008; Viciani, Lastrucci, Geri, & Foggi, 2016). The most complete available information on EU habitat distribution refers to maps of 10 x 10 km grid resolution or coarser (http://data.copernicus.eu/data-access.html; EEA, 2014; Münch, Hennekens, Bunce, Schaminée, & Schaepman, 2009), while more detailed information are often available only for some regions, or moreover only for N2K sites (e.g. Bertacchi, 2017; Viciani, Dell’Olmo, et al., 2016; Viciani, Dell’Olmo, Vicenti, & Lastrucci, 2017). To make EU habitat monitoring feasible, mapmaking should be started from fine-resolution spatial data (i.e. at the scale of N2K sites) and involve consistent and inexpensive procedures, respectively based on existing free available data and statistical methods.

Many machine-learning procedures have been used for habitat modelling in the frame of different research topics (e.g. Baselga & Araújo, 2009; Ferrier & Guisan, 2006; Guisan & Zimmermann, 2000; Thuiller, Araújo, & Lavorel, 2003). Among these, classification trees have been identified as one of the most effective techniques in GIS modelling (Muñoz & Felicísimo, 2004). For example, they have been successfully applied to map EU habitat spatial distributions at a coarse scale (10 x 10 km grid scale) across Europe (Mücher et al., 2009).

The aim of the present work was to define a straightforward procedure for implementing a fine scale regional distribution of EU habitats, knowledge of which is necessary and urgent to satisfy requests related to the 6-yearly reporting under articles 11 and 17 of the Habitats Directive. As a case study we selected the administrative region of Lombardy (Northern Italy), where an integrated map of EU habitat distribution is currently available only within the N2K network (Brusa, Cerabolini, Corti, & De Molli, 2016), while
outside, which is a large part of the regional territory (~85%), information is completely lacking. The procedure was specifically developed to map the three main structural types of terrestrial EU habitats (grassland, scrub and forest formations) by using freely available maps concerning environmental factors and land use. As an application of our procedure, we estimated the regional distribution of three selected target EU habitats, one for each structural type, widely spread in the study area but also occurring in many EU MS (http://eunis.eea.europa.eu/habitats): Mountain hay meadows (EU habitat 6520) for grasslands, Alpine and Boreal heaths (EU habitat 4060) for scrubs, and Asperulo-Fagetum beech forests (EU habitat 9130) for forest formations.

2. Methods

2.1. Study area

The study area corresponds to the whole territory of Lombardy, an administrative region of Northern Italy, which spans a surface area of 23,870 km² between 44°40’-46°37’N in latitude and 8°29’-11°25’E in longitude (Figure 1).

Lombardy spreads across the Alpine and Continental biogeographic regions (ETC/BD, 2006), and its bioclimatic types range from continental to oceanic (Pesaresi, Galdenzi, Biondi, & Casavecchia, 2014), with a large variability of mesoclimates due to its complex orography. Geological substrates consist of a wide range of litho-types, including silicate or carbonate rocks, morainic and alluvial deposits. The high variability of environmental factors is consistent with the occurrence of a variety of plant communities, ranging from evergreen Mediterranean oak forests by Lake Garda to Alpine tundras of the highest mountains, and justify the large number of EU habitats present in the study area (Brusa, Cerabolini, Dalle Fratte, & De Molli, 2017): 58 on the whole, 49 terrestrial (code = 2, 4, 6, 7, 8 and 9) and 9 aquatic (code = 3).

The Lombardy N2K network occupies approximately 15% of the whole regional surface (Figure 1) and consists of 193 Sites of Community Importance, or Special Areas of Conservation, and 67 Zones of Special Protection.

2.2. Selection of data sources

A schematic diagram of the development of the whole procedure is shown in Figure 2. We selected 13 layers concerning regional maps of EU habitats, environmental factors, land use types and geographic boundaries, all freely available from online data sources (Table 1): the

![Figure 1](https://example.com/image1.jpg)
EU habitats map of Lombardy from the regional observatory for biodiversity (‘Osservatorio Regionale per la Biodiversità’, http://www.biodiversita.lombardia.it) and all the others from the Geoportal of Lombardy (GOL; http://www.cartografiaregione.lombardia.it/geoportale).

We selected layers concerning environmental factors according to their supposed strength in explaining EU habitat regional distributions (Biondi et al., 2010; Brusa et al., 2017). This latter was obtained from the regional map of land use types (DUSAF, ‘Destinazione Figure 2. Methodological scheme of the adopted procedure for modelling the regional distribution of N2K terrestrial habitats. Legend: EU = European Union; N2K = Natura 2000 network.

Table 1. Layers used for modelling the regional distribution of EU habitats.

| Layers                  | Description                                                                 | Source                                      | Resolution/scale of detail | Format          |
|-------------------------|-----------------------------------------------------------------------------|---------------------------------------------|-----------------------------|-----------------|
| EU habitats map         | Integrated map of EU habitat distribution within the N2K network of the study area. | Brusa et al., 2016                         | 1:10,000                    | Polygon shapefile |
| Environmental factors:  |                                                                              |                                             |                             |                 |
| Elevation               | Elevation (m a.s.l.) from Digital Elevation Model (DEM).                    | GOL                                         | 20 m                        | Raster          |
| Slope                   | Derived from DEM by GIS application.                                        | GOL                                         | 20 m                        | Raster          |
| Substrate               | Geolithological substrates (6 classes: silicate rocks; carbonate rocks; alluvial deposits; recent western moraines; recent eastern moraines; other moraine deposits). | GOL                                         | 1:250,000                   | Polygon shapefile |
| Soil                    | Soils classification (14 classes: cambisols, leptosols, regosols, fluvisols, podzol, umbirsols, phaeozem, luvisols, calcisols, alisols, histosols, gleysols, vertisols, arenosols). | GOL                                         | 1:250,000                   | Polygon shapefile |
| pH                      | Soil reaction (38 classes; minimum value = 4.6, maximum value = 8.4)        | GOL                                         | 1:250,000                   | Polygon shapefile |
| Precipitation           | Mean annual precipitation (mm y⁻¹) from GIS interpolation of free data (data among 1951–1982 form technical reports from ISTAT, Rome and ‘Ufficio idrografico e mareografico per il Bacino del Po’). | Unpublished data                           | 20 m                        | Raster          |
| (Solar) Radiation       | Mean annual total incoming radiation (Wh m⁻² d⁻¹) computed from DEM by means of Solar Analyst (Fu & Rich, 2000). | Unpublished data                           | 20 m                        | Raster          |
| Land uses types:        |                                                                              |                                             |                             |                 |
| DUSAF                   | Regional land use, v.5.0 (78 classes).                                       | GOL                                         | 1:10,000                    | Polygon shapefile |
| PIF                     | Forestland, v.2017, inclusive of a classification in forest types (231 types). | GOL                                         | 1:10,000                    | Polygon shapefile |
| Geographic boundaries:  |                                                                              |                                             |                             |                 |
| Regions                 | Geobotanical districts redrawn from forest regions (Del Favero, 2002).       | Unpublished data                           | 1:250,000                   | Polygon shapefile |
| Biogeographical regions | Biogeographical regions, sensu European Environmental Agency.               | ETC/BD (2006)                              | #ND                         | Polygon shapefile |
| Province                | Administrative province of Lombardy (v.2015).                               | GOL                                         | 1:10,000                    | Polygon shapefile |

Note: If not specified, the selected data were freely available online: the EU habitat map from the regional observatory for biodiversity (http://www.biodiversita.lombardia.it), while environmental factors, land use types and geographic boundaries from the Geoportal of Lombardy (GOL; http://www.cartografiaregione.lombardia.it/geoportale). Legend: #ND = not defined.
d’Uso dei Suoli Agricoli e Forestali’ = ‘Intended use of agricultural and forest land’) and/or from the map of forest types (PIF, ‘Piano d’Indirizzo Forestale’ = ‘Forest management plans’). We used geographic boundaries in the last phase of procedures to impose biogeographic constraints based on regional distribution of plant communities corresponding to EU biogeographic constraints based on regional distributions in the last phase of procedures to impose management plans forest types (PIF, agricultural and forest land).

Original layers in vector format were transformed into raster (adjusted to the DEM 20 m resolution, which was the lowest resolution of the input variables), and georeferenced to the regional coordinate system (UTM32N-WGS84, EPSG 32632), in order to perform computing operations.

2.3. Relationship between EU habitats and land use types

To assess the correspondence between one or a group of EU habitats and land use types, we firstly calculated how much of each EU habitat was represented in each land use type within the N2K regional network, by means of GIS intersections. Afterwards, for each land use class we selected the EU habitats most represented and congruent in terms of vegetation and ecology for that land use type (Biondi et al., 2014; Brusa et al., 2017; Del Favero, 2002). Forest EU habitats were selected from woodland land use types and further investigated using the forest types map (PIF). We uniquely assigned some forest types in PIF to a specific forest EU habitats, while for forest types in PIF presenting a wide physionomic and ecological range we modelled EU habitats as detailed in the next paragraph.

2.4. Dataset construction and EU habitat modelling

When a specific land use (or forest) type corresponded to a group of EU habitats (the most common case), we computed a model to split EU habitats from one another according to their ecology.

For each group we selected environmental factors on the base of scientific knowledge about ecological requirements of each EU habitats (AA.VV., 2014; Angelini, Casella, Grignetti, & Genovesi, 2016; Biondi et al., 2010; Brusa et al., 2017; Del Favero, 2002). For sampling the values of environmental factors for each EU habitat, we randomly generated a set of points (N = 1000, at a minimum distance of 25 m from each other) within each EU habitat within the N2K regional network. The sets of points were further randomly divided in training (2/3*N) and validation subsets (1/3*N) (Araújo, Pearson, Thuiller, & Erhard, 2005; Guisan & Zimmermann, 2000; Muñoz & Felicísimo, 2004).

We used the training subsets for modelling EU habitats in the same group (response variables) by means of conditional inference trees (Hothorn & Lausen, 2003; Lauen & Schumacher, 1992), considering environmental factors as predictive variables. This technique estimates a classification relationship by binary recursive partitioning of predictive variables in a conditional inference framework (Hothorn, Hornik, & Zeileis, 2006); conditional nodes were evaluated through Monte Carlo testing (p-value < 0.01). The validation subset was used to check the model prediction efficiency, comparing observed and predicted distribution by means of confusion matrices. For each model, we computed different commonly applied indexes: (i) on the whole model, the proportion of correctly classified (accuracy, and its 95% confidence interval), no-information rate (NIR) and Cohen’s Kappa (Cohen, 1960), sensitivity (% of positive effects) and specificity (% of negative effects), and (ii) for each EU habitat, sensitivity and specificity.

We fixed an acceptance threshold for a high overall accuracy at 85% (Foody, 2002). The NIR was used to evaluate whether the accuracy was higher than the proportion in the observed values of the largest class by means of a one-side binomial test, which performs a simple null hypothesis test about the probability of success in a Bernoulli experiment (i.e. a binomial trial) (Hollander, Wolfe, & Chicken, 2015).

We computed also the true skill statistic (TSS; Allouche, Tsioar, & Kadmon, 2006) on sensitivity and specificity values. We interpreted TSS and Cohen’s Kappa as reported by Landis and Koch (1977): 0% poor; 20% slight; 40% fair; 60% moderate; 80% substantial; and 100% perfect agreement between predicted and observed classes. After the validation assessment, we rendered classification models into easily interpretable and applicable algorithms (i.e. dichotomous keys, Supplementary materials).

2.5. Map production

For each EU habitat, we firstly mapped the potential distribution in GIS using the algorithms computed by the classification trees, and subsequently the estimated distribution through intersecting its potential distribution and the corresponding land use class (i.e. DUSAF, and eventually PIF for forest EU habitats). When a single EU habitat was assigned to two or more different land use types, we computed its regional distribution by merging all its estimated distribution in GIS. As an example, the estimated distribution of EU habitat 9130 derived by the sum of (a) direct assignment of PIF classes to the EU habitat, and (b) output from the model applied on the PIF class ‘not-classified beech forests’.

The final output was the regional distribution of a given EU habitat, obtained filtering the estimated distribution by its geographic limits, based on well-
established technical and scientific knowledge (e.g. data on presence/absence in administrative areas or biogeographic regions) (AA.VV., 2014; Biondi et al., 2010; Brusa et al., 2017). In our example, we applied filters to the estimated distribution of two target EU habitats: habitat 6520 was excluded from administrative province of Varese and from the Apennine belt; habitat 4060 from the low Po plain.

For the validation of regional distributions, we used the cartography of N2K network, as this is currently the only existing source concerning the real distribution of EU habitats in the study area. For each EU habitat, we divided the N2K network into two parts: the area occupied from a target EU habitat and the remaining N2K area (i.e. the area of non-target habitats). To overcome this unbalanced design, due to extensive differences between areas of a target EU habitat (smaller) and that of non-target habitats (wider), we selected random points (N = 1000) stratified in each of these two areas (Hirzel & Guisan, 2002). For each point we assessed the presence/absence of target EU habitat, both observed (N2K map) and predicted (regional distribution in N2K). Finally, we computed a 2 × 2 confusion matrix and the corresponding statistics as reported for EU habitat modelling (paragraph 2.4).

3. Results

Within each of the three structural types of terrestrial EU habitats, a group of EU habitats was assigned to one or two land use (DUSAF, and eventually PIF) classes (Table 2), and then a classification model was calculated for each of these groups (Supplementary material).

One of the advantages of using conditional inference trees (Hothorn et al., 2006) is that they automatically remove variables without a statistically significant association with the response (p > 0.01); in our study, they removed radiation for grasslands, slope for scrubs, pH and soil for forests (Supplementary material). Accordingly, substrate and precipitation were the most effective environmental factors, determining a statistically significant effect in all of the three models. Elevation and radiation were also significant, the former for grasslands and scrubs, the latter for scrubs.

The accuracy values were statistically higher than the NIR values in all models (Table 3). High values of overall accuracy were computed for scrubs and forests. However, the model of forest EU habitats was the most accurate, with a substantial agreement between observed and predicted classes according to the Kappa value. The three models showed high rate of specificity (> 84%) for each EU habitat, while in contrast, sensitivities showed a high variability, in particular for grasslands (between 64% and 89%) and scrubs (between 70% and 96%). TSS values ranged from fairly-moderate (only two cases) to substantial agreement between observed and predicted EU habitats: between 53% and 88% for grasslands, 58% and 95% for scrubs, 75% and 92% for forests.

The accuracy values of each regional distribution were statistically higher than NIR values (Table 4). A moderate or substantial accuracy was generally detected in each estimated distribution according to Kappa values. The specificity was high for all target EU habitats (>94%), but the sensitivity resulted in low values (the lowest for 9130 and the highest for 6520). The values of TSS ranged between 51% (EU habitat 9130) and 60% (EU habitat 6520).

The regional distribution of the three selected EU habitats cover overall 524 Km² (ca. 2% of Lombardy; Main Map). The largest area is covered by EU habitat 6520 (254 km²), while the other two target EU habitats are of approximately the same size (147 km² for 4060 and 123 km² for 9130). EU habitat 4060 is mainly within N2K (93% of its regional distribution), while smaller proportion of EU habitats 9130 and especially 6520 are covered by N2K (respectively 57% and 13%).

Table 2. The three groups of habitats that were modelled by land use and environmental factors, and eventual filter layers (targets EU habitats are in bold, while environmental factors automatically removed by classification trees are in italic).

| Habitat | EU habitats (code: description) | Land use | Environmental factors | Filter layers |
|---------|--------------------------------|----------|-----------------------|--------------|
| grasslands | 6210: Semi-natural dry grasslands and scrubland facies on calcareous substrates (Festuco-Brometalia) | Grass | Meadow | Elevation | Province |
| | 6230: Species-rich Nardus grasslands, on siliceous substrates in mountain areas (and submountain areas, in Continental Europe) | | | Substrate | Regions |
| | 6510: Lowland hay meadows (Alopecurus pratensis, Sanguisorba officinalis) | | | Precipitation | |
| | 6520: Mountain hay meadows | | | Radiation | |
| scrubs | 4030: European dry heaths | Scrub | | Elevation | Province |
| | 4060: Alpine and Boreal heaths | | Substrate | Substrate | |
| | 4070: Bushes with Pinus mugo and Rhododendron hirsutum (Mugo-Rhododendretum hirsutii) | | Radiation | |
| | 4080: Sub-Arctic Salix spp. scrub | | Precipitation | Slope | |
| forests | 9110: Luzulo-Fagetum beech forests | Wood | Beech | Substrate | none |
| | 9130: Asperulo-Fagetum beech forests | | | Pre | |
| | 9140: Illyrian Fagus sylvatica forests (Armenio-Fagion) | | | cipitation | |

Legend: Grass = permanent grassland without trees and scrubs (from DUSAF); Meadow = permanent meadows with sparse trees and scrubs (from DUSAF); Scrub = scrublands (from DUSAF); Wood = forest areas (from DUSAF); Beech = not classifiable beech forests (from PIF).
4. Discussions

Our study provided for the first time an accurate map of the distribution of three EU habitats (6520, mountain hay meadows; 4060, alpine and boreal heaths; 9130, Asperulo-Fagetum beech forests) in the entire territory of the Lombardy administrative region (resolution 20 m; Main Map). Maps such as these are essential for monitoring and reporting the conservation status of EU habitats as required by art. 17 of the Habitats Directive (Evans & Arvela, 2011).

A detailed knowledge of EU habitat spatial distribution is necessary to implement standardized procedures for their monitoring (Bunce et al., 2008; Hochkirch et al., 2013; Velázquez, Tejera, Hernando, & Núñez, 2010), as well as to build more effective monitoring designs (Chiarucci, 2007), and to implement management plans and instruments for decision-making in land-use planning (Falucci, Maiorano, & Boitani, 2007; Hochkirch et al., 2013; Louette, Adriaens, Paelinckx, & Hoffmann, 2015).

Remote sensing has been widely used for mapping EU habitat spatial distributions (e.g. Mührer et al., 2013; Schmidt et al., 2017; Stenzel, Feilhauer, Mack, Metz, & Schmidtlein, 2014; Zlinszky, Deák, Kania, Schroiff, & Pfeifer, 2015) or in general for conservation monitoring (e.g. Nagendra et al., 2013). However, they require considerable ground-truthing and field-base monitoring (Zlinszky, Heilmeyer, Balzer, Czúcz, & Pfeifer, 2015), which instead is not mandatory for machine learning techniques (Guisan & Zimmermann, 2000; Thuiller et al., 2003; Vayssières, Plant, & Allen-Diaz, 2000).

Among these, classification models are easy to interpret, they have excellent predictive capabilities (Laborczi, Szatmári, Takács, & Pásztor, 2016), and solve many problems compared to other modelling approaches, for instance overfitting and a selection bias towards covariates with many possible splits or missing values (Hothorn et al., 2006). In our study, we successfully validated the modelled regional distribution of target EU habitats, confirming the goodness of this technique in the context of habitat modelling (Mührer et al., 2009) and its efficiency at fine resolution scale.

In Italy, the increasing availability of environmental datasets in recent years (Capotorti, Guida, Siervo, Smiraglia, & Blasi, 2012; Pesaresi et al., 2014; Smiraglia et al., 2013) together with the national vegetation and flora database (Gigante et al., 2012) provide a reliable groundwork for EU habitat modelling. In our study, we defined a standard procedure that could be applied to such datasets in developing fine scale resolution maps of EU habitats also occurring outside protected areas. Detailed maps are highly recommended for EU habitat monitoring, since the coarse scale of environmental layers can hide the high local variability of EU habitats (Feilhauer et al., 2014).

Mapping EU habitat distributions is essential for evaluating the efficiency of N2K for protecting species and habitats. N2K sites provide a central role in recovery of process of natural ecosystems (Prisco, Carboni, Jucker, & Acosta, 2016), as well as an important additional value for biodiversity conservation (Gruber et al., 2012; Maiorano et al., 2015; Trochet & Schmeller, 2013; van der Sluis et al., 2016; Viciani, Lastrucci, et al., 2012; Maiorano et al., 2015; Trochet & Schmeller, 2013; van der Sluis et al., 2016; but see Araújo, Alagador, Cabeza, Nogués-Bravo, & Thuiller, 2011). Despite this, many studies comparing non-protected and protected areas under N2K, highlighted an uneven distribution of EU habitats (Angiolini, Viciani, Bonari, & Lastrucci, 2017; Joppa & Pfaff, 2009; Maiorano, Falucci, Garton, & Boitani, 2007; Rosati et al., 2008) and protected species populations (Araújo, Lobo, & Moreno, 2007; Bagella, Caria, & Filigheddu, 2013; Rubio-Salcedo, Martínez, Carreno, & Escudero, 2013). Also our analysis revealed
that large parts of EU habitats 9130 and 6520 are not included within the regional N2K network. The percentage of EU habitat 9130 distribution not included in the N2K network is comparable to those identified by gap-analysis for Italy (Rosati et al., 2008), but at the same time, EU habitat 6520 was demonstrated to be underrepresented in the regional N2K network.

The Habitats Directive considers that the monitoring of EU habitat conservation status requires full comprehension of the variables driving habitat spatial distribution. Our results underlined the role of precipitation among the most effective environmental factors shaping EU habitat distributions at the regional scale. This confirms and highlights current threats for terrestrial habitats under predicted future climate changes (Barredo et al., 2016; Bittner, Jaeschke, Reineking, & Griebeler, 2015; EEA, 2017; Stocker et al., 2013).

Predictive models indicate that habitats will dramatically lose range in the future, also within N2K (Barredo et al., 2016; Bittner, Jaeschke, Reineking, & Beierkuhnlein, 2011), with drastic consequences for the provision of services (Essl, Dullinger, Moser, Rabitsch, & Kleinbauer, 2012; Jantke, Müller, Trapp, & Blanz, 2016; Marchetti et al., 2012). Eastwood et al. (2016) reported that protected sites deliver higher levels of ecosystem services compared to non-protected sites. Hence, mapping EU habitat distribution at a fine scale over wide EU territories could help to implement monitoring and management of protected areas, maximizing the benefits of the services they provide.

5. Conclusions

In our study, we provided a straightforward, but robust, procedure for mapping EU habitat distribution over the entire EU territory. Higher resolution environmental data and classification of land use types will be increasingly available in future; this protocol could provide an easy and low cost procedure to define EU habitat distributions with increasing accuracy. Furthermore, our procedure is suitable for predicting EU habitat distributions at the regional scale, when scenarios of future climate and/or land use changes are introduced in the models. Additionally, when old maps of land use types and corresponding climate data are available, it also can be useful for modelling historical trends of past EU habitat distributions.

Software

For our analysis, we used open source software. Processing and derivation of ancillary environmental variables, spatial analysis of EU habitats and land use maps, random points sampling, as well as the application of models outside N2K, were carried out in QGIS (v.2.14.12; QGIS Development Team, 2009). The classification trees were computed with R software (R Core Team, 2017) through the function ‘ctree’ available in the package ‘party’ (Hothorn et al., 2006), and statistics were calculated through the function ‘confusionMatrix’ (Kuhn, 2008) available in the package ‘caret’ (Kuhn et al., 2017).

Acknowledgments

We thanks Simon Pierce for his precious help in the revision of the manuscript.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research was part of the activities of the Lombardy region biodiversity observatory and funded by Fondazione Lombardia per l’Ambiente (Environmental Foundation of Lombardy, Italy).

References

AA.VV. (2014). Formulazione del programma di monitoraggio scientifico della rete Azione D1. Progetto LIFE GESTIRE. ERSAF e Università degli Studi dell’Insubria-Dipartimento di Scienze Teoriche e Applicate.

Allouche, O., Tsoar, A., & Kadmon, R. (2006). Assessing the accuracy of species distribution models: Prevalence, kappa and the true skill statistic (TSS). Journal of Applied Ecology, 43(6), 1223–1232. doi:10.1111/j.1365-2664.2006.01214.x

Angelini, P., Casella, L., Grignetti, A., & Genovesi, P. (2016). Manuali per il monitoraggio di specie e habitat di interesse comunitario (Direttiva 92/43/CEE) in Italia: habitat. ISPRA, Serie Manuali e linee guida, 142/2016.

Angioli, C., Viciani, D., Bonari, G., & Lastrucci, L. (2017). Habitat conservation prioritization: A floristic approach applied to a Mediterranean wetland network. Plant Biosystems-An International Journal Dealing with all Aspects of Plant Biology, 151(4), 598–612. doi:10.1080/11263504.2016.1187678

Araújo, M. B., Alagador, D., Cabeza, M., Nogués-Bravo, D., & Thuiller, W. (2011). Climate change threatens European conservation areas. Ecology Letters, 14(5), 484–492. doi:10.1111/j.1461-0248.2011.01610.x

Araújo, M. B., Lobo, J. M., & Moreno, J. C. (2007). The effectiveness of iiberian protected areas in conserving terrestrial biodiversity. Conservation Biology, 21(6), 1423–1432. doi:10.1111/j.1523-1739.2007.00827.x

Araújo, M. B., Pearson, R. G., Thuiller, W., & Erhard, M. (2005). Validation of species–climate impact models under climate change. Global Change Biology, 11(9), 1504–1513. doi:10.1111/j.1365-2486.2005.01000.x

Bagella, S., Caria, M. C., & Filigheddu, R. (2013). Gap analysis revealed a low efficiency of Natura 2000 network for the conservation of endemic species in Mediterranean temporary freshwater habitats. Plant Biosystems-An International Journal Dealing with all Aspects of Plant Biology, 147(4), 1092–1094. doi:10.1080/11263504.2013.860055
Hothorn, T., Hornik, K., & Zeileis, A. (2006). Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics, 15*(3), 651–674. doi:10.1198/106186006X139933

Hothorn, T., & Lausen, B. (2003). On the exact distribution of maximally selected rank statistics. *Computational Statistics & Data Analysis, 43*(2), 121–137. doi:10.1016/S0167-9473(02)00225-6

Jantke, K., Müller, J., Trapp, N., & Blanz, B. (2016). Is climate-smart conservation feasible in Europe? Spatial relations of protected areas, soil carbon, and land values. *Environmental Science & Policy, 57*, 40–49. doi:10.1016/j.envsci.2015.11.013

Joppa, L. N., & Pfaff, A. (2009). High and far: Biases in the location of protected areas. *PloS one, 4*(12), e8273. doi:10.1371/journal.pone.0008273

Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software, 28*(5), 1–26. doi:10.18637/jss.v028.i05

Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., ... Kenkel, B., & the R Core Team (2017). caret: Classification and Regression Training. https://CRAN.R-project.org/package=caret

Laborzci, A., Szatmári, G., Takács, K., & Pásztor, L. (2016). Mapping of topsoil texture in Hungary using classification trees. *Journal of Maps, 12*(5), 999–1009. doi:10.1080/17445647.2015.1113896

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159–174. doi:10.2307/2529310

Lausen, B., & Schumacher, M. (1992). Maximally selected rank statistics. *Biometrics, 48*, 73–85. doi:10.2307/2533740

Louette, G., Adriaens, D., Paelinckx, D., & Hofmann, M. (2015). Implementing the habits directive: How science can support decision making. *Journal for Nature Conservation, 23*, 27–34. doi:10.1016/j.jnc.2014.12.002

Maiorano, L., Amori, G., Montemaggiord, A., Rondinini, C., Santini, L., Saura, S., & Boitani, L. (2015). On how much biodiversity is covered in Europe by national protected areas and by the Natura 2000 network: Insights from terrestrial vertebrates. *Conservation Biology, 29*(4), 986–995. doi:10.1111/cobi.12535

Maiorano, L., Falconi, A., Garton, E. O., & Boitani, L. (2007). Contribution of the Natura 2000 network to biodiversity conservation in Italy. *Conservation Biology, 21*(6), 1433–1444. doi:10.1111/j.1523-1739.2007.00831.x

Marchetti, M., Sallustio, L., Ottaviano, M., Barbati, A., Corona, P., Tognetti, R., ... Capotorti, G. (2012). Carbon sequestration by forests in the national parks of Italy. *Plant Biosystems-An International Journal Dealing with all Aspects of Plant Biology, 146*(4), 1001–1011. doi:10.1080/11263504.2012.738715

Muñoz, J., & Felicísimo, ÁM. (2004). Comparison of statistical methods commonly used in predictive modelling. *Journal of Vegetation Science, 15*(2), 285–292. 10.1111/j.1654-1103.2004.tb02263.x

Mücher, C. A., Hennekens, S. M., Bunce, R. G., Schaminée, J. H., & Schaepman, M. E. (2009). Modelling the spatial distribution of Natura 2000 habitats across Europe. *Landscape and Urban Planning, 92*(2), 148–159. doi:10.1016/j.landurbplan.2009.04.003

Mücher, C. A., Kooistra, L., Vermeulen, M., Borre, J. V., Haest, B., & Haveman, R. (2013). Quantifying structure of Natura 2000 heathland habitats using spectral mixture analysis and segmentation techniques on hyperspectral imagery. *Ecological Indicators, 33*, 71–81. doi:10.1016/j.ecolind.2012.09.013

Nagendra, H., Lucas, R., Honrado, J. P., Jongman, R. H., Tarantino, C., Adamo, M., & Mairotta, P. (2013). Remote sensing for conservation monitoring: Assessing protected areas, habitat extent, habitat condition, species diversity, and threats. *Ecological Indicators, 33*, 45–59. doi:10.1016/j.ecolind.2012.09.014

Paresi, S., Galdenzi, D., Biondi, E., & Casavecchia, S. (2014). Bioclimate of Italy: Application of the worldwide bioclimatic classification system. *Journal of Maps, 10*(4), 538–553. doi:10.1016/j.jag.2014.09.017

Prisc, I., Carboni, M., Jucker, T., & Acosta, A. T. (2016). Temporal changes in the vegetation of Italian coastal dunes: Identifying winners and losers through the lens of functional traits. *Journal of Applied Ecology, 53*(5), 1533–1542. doi:10.1111/1365-2664.12684

QGIS Development Team. (2009). QGIS Geographic Information System. Open Source Geospatial Foundation. URL: http://qgis.osgeo.org

R Core Team. (2017). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. URL: https://www.R-project.org/

Rosati, L., Marignani, M., & Blasi, C. (2008). A gap analysis comparing Natura 2000 vs national protected area network with potential natural vegetation. *Community Ecology, 9*(2), 147–154. doi:10.1556/ComEc.9.2008.2.3

Rubio-Salcedo, M., Martínez, I., Carreno, F., & Escudero, A. (2013). Poor effectiveness of the Natura 2000 network protecting Mediterranean lichen species. *Journal for Nature Conservation, 21*(1), 1–9. doi:10.1016/j.jnc.2012.06.001

Schmidt, J., Fassnacht, F. E., Neff, C., Lausch, A., Kleinschmit, B., Förster, M., & Schmidtlein, S. (2017). Adapting a natura 2000 field guideline for a remote sensing-based assessment of heathland conservation status. *International Journal of Applied Earth Observation and Geoinformation, 60*, 61–71. doi:10.1016/j.jag.2017.04.005

Smiraglia, D., Capotorti, G., Guida, D., Mollo, B., Siervo, V., & Blasi, C. (2013). Land units map of Italy. *Journal of Maps, 9*(2), 239–244. doi:10.1016/j.jag.2013.07.71290

Stenzel, S., Feihauer, H., Mack, B., Metz, A., & Schmidtlein, S. (2014). Remote sensing of scattered Natura 2000 habitats using a one-class classifier. *International Journal of Applied Earth Observation and Geoinformation, 33*, 211–217. doi:10.1016/j.jag.2014.05.012

Stocker, T. F., Qin, D., Plattner, G.-K., Alexander, L. V., Allen, S. K., Bindoff, N. L., ... Xie, S.-P. (2013). Technical summary. In T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, ... P. M. Midgley (Eds.), *Climate change 2013: The physical science basis. Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change*. Cambridge: Cambridge University Press.

Thullier, W., Araújo, M. B., & Lavorel, S. (2003). Generalized models vs. Classification tree analysis: Predicting spatial distributions of plant species at different scales. *Journal of Vegetation Science, 14*(5), 669–680. doi:10.1111/j.1654-1103.2003.tb02199.x

Trotchet, A., & Schmeller, D. (2013). Effectiveness of the Natura 2000 network to cover threatened species. *Nature Conservation, 4*(35), doi:10.3979/natureconservation

van der Sluis, T., Foppen, R., Gillings, S., Groen, T. A., Henkens, R., Hennekens, S., ... Siersema, H. (2016). How much biodiversity is in Natura 2000?: the “umbrella effect” of the European Natura 2000 protected area network. Wageningen, Alterra Wageningen UR (University & Research centre), *Alterra report 2730R*. 148 pp.
Vayssières, M. P., Plant, R. E., & Allen-Diaz, B. H. (2000). Classification trees: An alternative non-parametric approach for predicting species distributions. *Journal of Vegetation Science, 11*(5), 679–694. doi:10.2307/3236575

Velázquez, J., Tejera, R., Hernando, A., & Núñez, M. V. (2010). Environmental diagnosis: Integrating biodiversity conservation in management of Natura 2000 forest spaces. *Journal for Nature Conservation, 18*(4), 309–317. doi:10.1016/j.jnc.2010.01.004

Viciani, D., Dell’Olmo, L., Ferretti, G., Lazzaro, L., Lastrucci, L., & Foggi, B. (2016). Detailed Natura 2000 and CORINE biotopes habitat maps of the island of Elba (Tuscan Archipelago, Italy). *Journal of Maps, 12*(3), 492–502. doi:10.1080/17445647.2015.1044040

Viciani, D., Dell’Olmo, L., Vicenti, C., & Lastrucci, L. (2017). Natura 2000 protected habitats, massaciuccoli lake (northern tuscany, Italy). *Journal of Maps, 13*(2), 219–226. doi:10.1080/17445647.2017.1290557

Viciani, D., Lastrucci, L., Geri, F., & Foggi, B. (2016). Gap analysis comparing protected areas with potential natural vegetation in Tuscany (Italy) and a GIS procedure to bridge the gaps. *Plant Biosystems-An International Journal Dealing with all Aspects of Plant Biology, 150*(1), 62–72. doi:10.1080/11263504.2014.950623

Zlinszky, A., Deák, B., Kania, A., Schroiiff, A., & Pfeifer, N. (2015). Mapping Natura 2000 habitat conservation status in a pannonic salt steppe with airborne laser scanning. *Remote Sensing, 7*(3), 2991–3019. doi:10.3390/rs70302991

Zlinszky, A., Heilmeier, H., Balzter, H., Czúc, B., & Pfeifer, N. (2015). Remote sensing and GIS for habitat quality monitoring: New approaches and future research. *Remote Sensing, 7*, 7987–7994. doi:10.3390/rs70607987