Large-scale and fine-grained mapping of heathland habitats using open-source remote sensing data

Laurence Hubert-Moy, Clémence Rozo, Gwenhael Perrin, Frédéric Bioret, Sébastien Rapinel

Remote Sensing in Ecology and Conservation
Open Access
ZSL LET'S WORK FOR WILDLIFE

ORIGINAL RESEARCH

Large-scale and fine-grained mapping of heathland habitats using open-source remote sensing data

Laurence Hubert-Moy[1], Clémence Rozo[1], Gwenhael Perrin[2], Frédéric Bioret[2] & Sébastien Rapinel[1]

[1]University of Rennes, LETG UMR 65554 CNRS, place du recteur Henri Le Moal, Rennes 35000, France
[2]University of Brest, Laboratoire Géoarchitecture, Brest 29200, France

Keywords
Conservation status, heathland, MaxEnt, natural vegetation, satellite imagery, vegetation mapping

Correspondence
Laurence Hubert-Moy, University of Rennes, LETG UMR 65554 CNRS, place du recteur Henri Le Moal, 35000 Rennes, France. Tel: +33 2 99141847; Fax: +33 2 99141895; E-mail: laurence.hubert@univ-rennes2.fr

Editor: Mat Disney
Associate Editor: Abdulhakim Abdi

Received: 25 June 2021; Revised: 10 December 2021; Accepted: 17 December 2021

doi: 10.1002/rse2.253

Remote Sensing in Ecology and Conservation 2022;**8** (4):448–463

Abstract

Mapping natural habitats remains challenging, especially at a national scale. Although new open-access variables for vegetation and its environment and increased spatial resolution derived from satellite remote sensing data are available at the global scale, the relevance of these new variables for fine-grained mapping of natural habitats at a national scale remains underexplored. This study aimed to map the fine-grained pattern of four heathland habitats throughout France (550 000 km^2). Environmental (bioclimatic, soil and topographic) and spectral (vegetation) variables derived from MODerate resolution Imaging Spectroradiometer, Advanced Spaceborne Thermal Emission and Reflection Radiometer, and Sentinel-2 satellite data were analyzed using the MaxEnt classifier. Open-access field databases were used to calibrate and validate the classification, based on the threshold-independent area under the curve (AUC) index and the conventional F1-score. For each heathland habitat, potential and actual areas were mapped using environmental and spectral variables, respectively. The results showed high classification accuracy for potential (AUC 0.92–0.99) and actual (AUC 0.88–0.99) suitability maps of the four heathland habitats. Visual interpretation of maps of the probability of occurrence indicated that the fine-grained distribution of heathland habitat was detected satisfactorily. However, although the accuracy of the crisp map of combined classifications of actual heathland habitats was high (overall accuracy 0.72), estimated producer's accuracies in terms of proportion of area were low (<0.25). This study provides the first fine-grained pattern maps of heathland habitats at a national scale, thus highlighting the value of combining environmental and spectral variables derived from open-remote sensing data and open-source field databases. These suitability maps could support the identification of heathland habitats in the framework of national conservation policies.

Introduction

Heathland natural and semi-natural habitats, which are dominated by chamaephytic plant communities, provide many ecosystem services such as biodiversity maintenance and carbon storage (Alonso & Härdtle, 2015). However, heathland conservation monitoring is complicated due to the lack of information on large, fine-grained and accurate habitat mapping. The fine-grained pattern of heathlands is well known for sites of heritage interest of a few square kilometers, especially using very high spatial resolution remote sensing (RS) time series (Fenske et al., 2020; Neumann et al., 2015; Schmidt et al., 2017, 2018). In comparison, mapping heathland habitats at a national or continental scale is hindered by several limitations that depend on the data sources, classification methods or data-processing techniques currently used. These techniques include (i) processing satellite data imagery at high spatial resolution (10–30 m) to develop maps of heathlands considered as one single land use and land cover (LULC) class (Inglada et al., 2017; Malinowski et al., 2020; Venter & Sydenham, 2021) or as general habitat categories (Lucas et al., 2011), which results in low classification accuracy (Maxwell et al., 2018) or in

one-to-many correspondence issues with recognized natural habitat classification schemes (Adamo et al., 2014; Lucas et al., 2015), respectively, and (ii) spatial interpolating field vegetation surveys to develop broad-scale maps of heathlands considered as natural habitats according to the European Union Nature Information System (EUNIS) classification (Chytrý et al., 2020).

Recently, the open access of ecological variables derived from satellite RS data and field data provides the opportunity to map the fine-grained pattern of habitats at a national scale. Moreover, several types of products derived from RS data that characterize climate, soil, topography or vegetation phenology are available in open access at global coverage (Rapinel & Hubert-Moy, 2021). Specifically, the 10-m Sentinel-2 multispectral time series can discriminate among physiognomically similar natural habitats based on their phenological profile (Tarantino et al., 2021). In parallel, the EU Directive INSPIRE (Bartha & Kocsis, 2011) led to the diffusion of open-access field databases that support calibration and validation of RS-based classifications (Rocchini et al., 2017), especially at a national scale, for which the cost of collecting field data would be too high. Although the quality of open data is considered heterogeneous, whether derived from RS (Moudrý et al., 2018) or field data (Suarez-Seoane et al., 2020), open data have recently been used to map grasslands (Rapinel et al., 2021) and forest habitats (Agrillo et al., 2021) at a national scale. However, these two studies used a multi-classifier, which requires first masking the area of interest (e.g. forest, grassland) from an existing LULC layer, whose quality (e.g. spatial scale, completeness) inevitably biased the classification accuracy.

Compared to multiclassifiers, one-class classifiers (OCCs) avoid using masks because they can classify a natural habitat regardless of other types of vegetation present in the landscape (Álvarez-Martínez et al., 2018). OCCs have been widely used to map plant species or LULC types, but rarely to map natural habitats (Rapinel & Hubert-Moy, 2021). Álvarez-Martínez et al. (2018) successfully mapped potential and actual areas of 24 habitats at 30 m spatial resolution over a Spanish Mediterranean region of 1200 km^2 (with 70% overall accuracy for the actual area) using environmental (bioclimatic and topographic) and vegetation (derived from Landsat and LiDAR data) variables, respectively. Although this approach is interesting since it highlights advantages of the complementarity between environmental and vegetation variables, three issues remain: (i) its reproducibility at a national scale; (ii) the spatial resolution of the maps produced, which remains too coarse to detect the fine-grain pattern of certain natural habitats such as heathlands; and (iii) the lack of using annual time series of satellite images for soil variables, which would better characterize the environment and discriminate among heathland habitats according to their phenology. Moreover, robust calibration of OCCs at a high spatial resolution and a national scale remains complex and notably requires hyperparameterization (Fernandez & Morales, 2019) based on spatially independent samples (Karasiak et al., 2021).

The present study aimed to assess the advantage of using new open-access environmental and vegetation variables derived from RS and field data to classify the fine-grained pattern of heathland habitats at a national scale. Four research questions were addressed: (i) Is the quality of the open data suitable for mapping heathland habitats? (ii) Are open-access environmental and vegetation variables sufficient for classifying potential and actual heathland areas? (iii) Does applying OCCs at a national scale accurately map heathland habitats? and (iv) What is the conservation value of these classifications?

Materials and Methods

Study site

The study site encompassed mainland France (c. 550 000 km^2) (Fig. 1) and covered Atlantic, Continental, Alpine and Mediterranean biogeographic regions (European Environment Agency, 2016). Six heathland habitats in France are of interest to the EU, including two of priority interest due to their unfavorable conservation status (Bensettiti et al., 2002; Bensettiti et al., 2005). This study focused on four heathland habitats characterized by chamaephytic plant communities and ericoid species:

- 4010 – Northern Atlantic wet heaths with *Erica tetralix*: humid, peaty or semi-peaty heaths, other than blanket bogs, of the Atlantic and sub-Atlantic domains
- 4020 – Temperate Atlantic wet heaths with *Erica ciliaris* and *Erica tetralix*: hygrophilous heaths of areas with a temperate oceanic climate, on semi-peaty or dried-out soils (habitat of priority interest)
- 4030 – European dry heaths: mesophile or xerophile heaths with on siliceous, podsolic soils in moist Atlantic and sub-Atlantic climates of plains and low mountains of western, central and northern Europe, including *Calluna vulgaris*
- 4060 – Alpine and Boreal heaths: small, dwarf or prostrate shrub formations of the alpine and subalpine zones of the mountains of Eurasia dominated by ericaceous species, *Dryas octopetala*, dwarf junipers, brooms or greenweedsHabitats 4040 (dry Atlantic coastal heaths with *Erica vagans*) was not considered due to few available reference data linked to their limited geographic distribution. The habitat 4090 (Endemic oro-Mediterranean heaths with gorse) was also ignored, because it is

composed of heterogeneous plant formations (i.e. garrigues, swards, etc.).

Remote sensing-based variables

All variables were derived from two main RS data types: environmental, that include bioclimatic, soil, topographic; and spectral that include vegetation (Table 1). Environmental variables, selected based on expert knowledge, were used to predict the potential distribution of habitats, while spectral variables were used to predict their actual distribution. The bioclimate variables included seven maps especially adapted for French ecological systems (Perrin et al., 2020) that were derived from MODerate

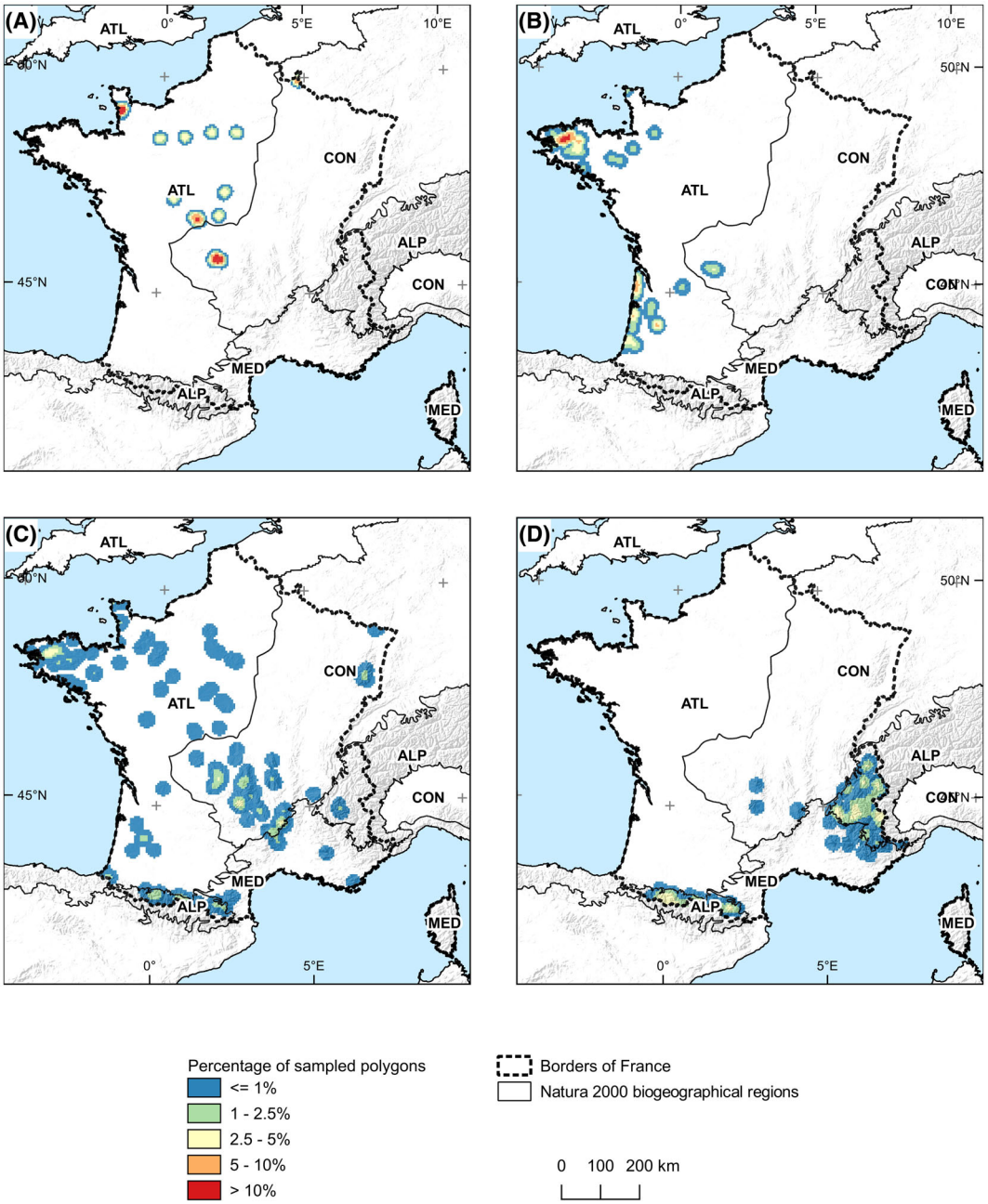

Figure 1. Study site and reference point locations: (A) 4010 – Northern Atlantic wet heaths with Erica, tetralix, (B) 4020 – Temperate Atlantic wet heaths with Erica ciliaris and Erica tetralix, (C) 4030 – European, dry heaths and (D) 4060 – Alpine and Boreal heaths (codes from the European Union Habitats Directive).

Resolution Imaging Spectroradiometer (MODIS) satellite data (Fick & Hijmans, 2017). The soil variables consisted of six SoilGrids featuring physical and chemical properties of the topsoil horizon, which were also derived from MODIS satellite data (Hengl et al., 2017). Four topographic variables were derived from the pan-European EU DEM (García González et al., 2015), which was based on Shuttle Radar Topography Mission and Advanced Spaceborne Thermal Emission and Reflection Radiometer satellite data. A multiscale topographic position index was calculated at three spatial scales: 250, 1000 and 2000 m (Guisan et al., 1999). A topographic wetness index was calculated using a multidirectional flow algorithm (Kopecký & Čížková, 2010). Topographic curvature reflects the convexity (positive values) and concavity (negative values) of the relief. All environmental variables were then resampled using a bicubic interpolation to 25 m to preserve the most resolved variable quality before the following modeling steps.

The vegetation variables were derived from a cloud-free Sentinel-2 satellite annual time series from July 2018 to July 2019. The images were downloaded as a level 3A product that is a synthesis of cloud-free observations averaged over a 45-day period (Hagolle et al., 2018), previously orthorectified and transformed into surface reflectance using the multisensor atmospheric correction and cloud screening algorithm (Hagolle et al., 2015). These Sentinel-2 L3A images were retrieved monthly from the Theia platform (www.theia-land.fr/en) for the whole of mainland France (88 tiles). For each tile, the 20-m spatial resolution bands in the red-edge (b5: 704 nm, b6: 740 nm, b7: 782 nm), near-infrared (b8a: 865 nm) and shortwave infrared (SWIR) (b11: 1614 nm, b12: 2202 nm) spectra were resampled to 10 m using a bilinear interpolation. These bands were then stacked with other 10-m native resolution bands (i.e. blue (b2: 497 nm), green (b3: 560 nm), red (b4: 664 nm) and infrared (b8: 835 nm) spectra). To avoid artifacts related to snow cover, images acquired from November 2018 to March 2019 were removed from the analysis. In addition, the Pearson correlation coefficient was calculated between bands to limit collinearity, and bands with $r > 0.80$ were also removed, which left 13 variables for analysis (Table 1).

Reference data

The reference data consisted of two datasets. The first dataset featured heathland habitats obtained from field maps produced at a scale of 1:5000 from 2000 to 2020. This open-access archive was compiled regionally by botanical conservatories and was provided by each region's environmental directorate. When appropriate, correspondence between habitat classifications of the EU Habitats Directive (Annex I) and biotope classifications of the EUNIS or CORordination of INformation on the Environment was performed according to the HABREF reference system (Clair et al., 2017). A 10-m negative buffer was applied to each heathland polygon to smooth out potential delineation bias. Polygons that represented a mosaic of habitats were discarded, as were those with an area less than 0.36 ha (i.e. the theoretical area of 9 Sentinel-2 pixels). Because some field maps had been created 20 years ago, each polygon was visually verified using recent Google Earth images. Each habitat was first subsampled with a maximum of 5 polygons per 5 × 5 km grid to minimize the heterogeneity in spatial distribution. Habitats 4030 and 4060 (the most frequent) were then subsampled again with a maximum of 2 polygons each per 5 × 5 km grid (i.e. 40% of the polygons) to avoid class unbalancing in the reference dataset. The reference points, which consisted of each Sentinel-2 pixel strictly inside the remaining polygons, were then extracted (Table 2).

The second reference dataset featured LULC classes to quantify classifier over-detection. To this end, c. 100 reference points were extracted for each of the six LULC classes (i.e. bare rock, sward, forest, grassland, buildings, crop) from the 2018 French national LULC map (overall accuracy 0.89) derived from Sentinel-2 annual time series (Inglada et al., 2017). Values of each variable were assigned to each reference point.

Discriminant analysis

This step investigated whether bioclimatic, soil, topographic and vegetation variables discriminated among heathland habitats and between them and other LULC classes. To this end, linear discriminant analysis (LDA) (Balakrishnama & Ganapathiraju, 1998) was applied to all reference points. This dimension-reduction method maximizes interclass variance to represent the partitioning and dispersion of the habitats in two-dimensional space by ordinating the reference points.

MaxEnt classification

This step involved predicting the potential distribution area of each heathland habitat using the 17 environmental variables and then the actual area of each heathland habitat using the 13 spectral variables (Fig. 2). To this end, we selected the widely used MaxEnt OCC (Phillips & Dudík, 2008), which is based only on presence points, so no absence points were needed (Guillera-Arroita et al., 2014). Instead, background points that describe the 'landscape' were used (Phillips & Dudík, 2008). To

Table 1. Description of the environmental (light gray) and spectral (dark gray) variables used to model heathland habitats.

Type	Name	Acronym	Unit	Resolution (m)	Source	References
Climate	Simple continentality index	Ic	°C	~1000	WordlClim v2	Perrin et al. (2020)
	Annual ombrothermic index	Io	–			
	Annual ombro-evaporation index	Ioe	–			
	Bimonthly estival ombrothermic index	Ios2	–			
	Fourmonthly estival ombrothermic index	Ios4	–			
	Compensated thermicity index	Itc	°C × 10			
	Annual positive temperature	Tp	°C × 10			
Topographic	Slope	Slope	degree	25	EU-DEM v1.1	García González et al. (2015)
	Topographic Position Index	TPI	–			
	Topographic Wetness Index	TWI	–			
	Topographic curvature	Curve	–			
Soil	Bulk density	BLDFIE	kg/m^3	250	SoilGrid v1	Hengl et al. (2017)
	Coarse fragment volume	CRFVOL	–			
	Soil organic carbon content	ORCDRC	g/kg			
	Soil pH	PHIHOX	pH			
	Soil texture fraction of silt	SLTPPT	%			
	Soil texture fraction of sand	SNDPPT	%			
Vegetation	July 2018, band 2 (blue)	july.2018.B02	Radians	10	Sentinel-2 L3A	Hagolle et al. (2018)
	July 2018, band 12 (SWIR)	july.2018.B12		20		
	August 2018, band 11 (SWIR)	aug.2018.B11		20		
	September 2018, band 3 (green)	sep.2018.B03		10		
	September 2018, band 8 (NIR)	sep.2018.B08		10		
	October 2018, band 2 (blue)	oct.2018.B02		10		
	October 2018, band 6 (red-edge)	oct.2018.B06		20		
	April 2019, band 3 (green)	april.2019.B03		10		
	April 2019, band 11 (SWIR)	april.2019.B11		20		
	May 2019, band 2 (blue)	may.2019.B02		10		
	May 2019, band 12 (SWIR)	may.2019.B12		20		
	July 2019, band 4 (red)	july.2019.B04		10		
	July 2019, band 8 (NIR)	july.2019.B08		10		

NIR, near infrared spectrum, SWIR, short wave infrared spectrum.

Table 2. The heathland habitats studied (codes from the European Union Habitats Directive) and the number of reference and background point sizes used for modeling.

Habitat	Habitat name	Reference points	Background points
4010	Northern Atlantic wet heaths with *Erica tetralix*	125	4771
4020*	Temperate Atlantic wet heaths with *Erica ciliaris* and *Erica tetralix*	340	12 977
4030	European dry heaths	524	20 000
4060	Alpine and Boreal heaths	506	19 313

The asterisk (*) indicates a habitat of priority interest.

combine the predicted maps and compare their accuracy scores, the same ratio of presence/background points and the same variables were used to classify each habitat (Álvarez-Martínez et al., 2018; Merow et al., 2013; Stenzel et al., 2014).

Potential area classification

Background points were randomly selected from the entire study site. Due to its large size (114 479 columns × 106 646 rows), the default number of background points (10 000) was increased to 20 000. To ensure that the ratio of presence/background points was the same, background points for the less frequent classes were subsampled (Table 2). For each class, the presence and background points were divided into a calibration dataset (80%) and a validation dataset (20%). To avoid autocorrelation bias, this split was run under a 60 × 60 km spatial blocks.

To improve classification accuracy, a hyperparameterization procedure was applied to each MaxEnt model, which has two parameters that require calibration (Elith

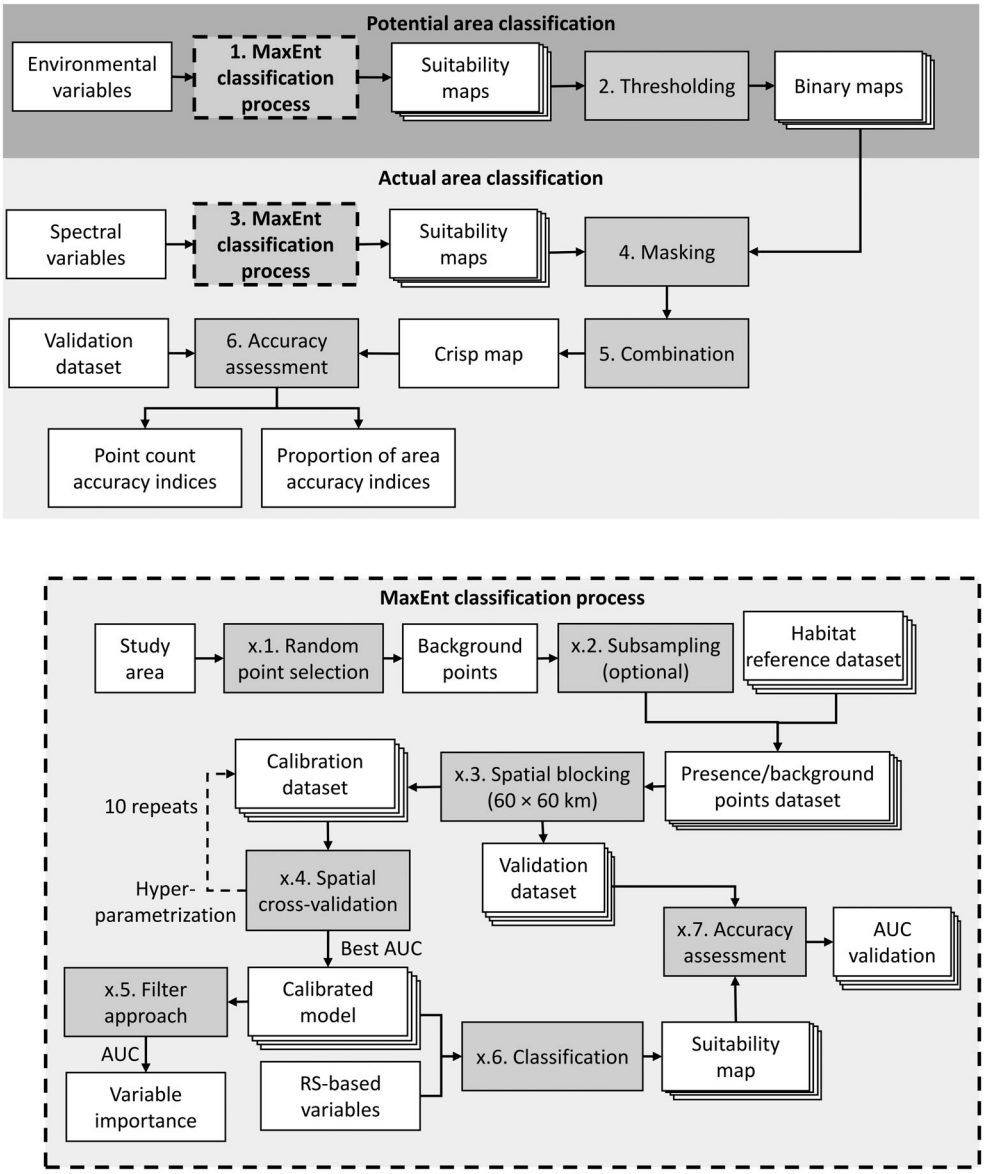

Figure 2. Methodological flowchart: the upper part describes the general approach while the lower part, details the MaxEnt classification process.

et al., 2011): (i) the feature class, which controls the transformation of the variables, and (ii) regularization (beta), which regulates the balance between model complexity and fit (Merow et al., 2013). For each class, the MaxEnt model was hyperparametrized using six feature classes (i.e. linear (L), combination of linear and quadratic (LQ), hinge (H), combination of hinge and quadratic (HQ), threshold (T) and combination of hinge, quadratic, product and threshold (HQPT)), and all regularization values from 1 to 10, with a step size of 0.25. Models were parameterized using threefold cross-validation repeated 10 times (Kuhn & Johnson, 2013) to prevent overfitting. To decrease the influence of spatial autocorrelation between nearby reference points, a 'spatial block' constraint was used to assign each reference point to a fold (Valavi et al., 2019). Specifically, reference points were assigned to a fold based on their location within 60 × 60 km spatial blocks. The best feature class and the best regularization parameter were selected based on the highest area under the curve (AUC) value, which is a threshold-free index of accuracy. AUC is interpreted as the probability that a randomly chosen presence site is ranked above a random background site (Phillips et al., 2006).

The accuracy of each classification was assessed using the AUC value calculated from the validation dataset. Since the AUC is commonly criticized (Merow et al., 2013), we compared, for each class, the validation AUC value to that of a null model using the approach of Bohl et al. (2019). Specifically, 100 modeling iterations were performed for each habitat, with dummy presence points randomly selected in each iteration. The mean AUC of the iterations was then calculated.

Since MaxEnt does not calculate a variable's importance score, it was calculated for each model using a 'filter' approach (Kuhn & Johnson, 2013). Specifically, a sequence of thresholds was run for each variable to predict the heathland habitat. Sensitivity and specificity were then calculated for each threshold, and the receiver operating characteristic curve was derived. See the supplementary materials for the resulting AUC that was used as an index of variable importance (Figs. S1 and S2). Each habitat was predicted using the calibrated classifier at a national scale, and then a binary map (presence/absence) was generated using the maximum sensitivity specificity thresholding method (Liu et al., 2013).

Actual area classification

For each habitat, the model was calibrated using the same procedure as that for the potential area, but using the 13 spectral variables instead. The predicted actual area was applied within the potential area and then the output maps of each habitat were combined. To this end, logistic values below the maximum sensitivity specificity threshold were set to zero. Each pixel was then assigned to the heathland class that had the highest probability among the four heathland classes. Pixels with a zero value in all heathland classes were assigned to the 'others' class (i.e. non-heathland). The accuracy of the combined map of actual heathland habitat was derived from the validation set and from the six LULC points merged into the 'others' class using the conventional point count approach (Congalton & Green, 2019). While the number of validation points differed between heathland classes, the estimation of map accuracy could be weighted by the proportion of the study site represented by each map class. For this reason, an additional accuracy assessment was calculated using the estimated proportion of area per class (Olofsson et al., 2014).

All analyses were performed using R software version 3.6 (R Core Team, 2019) and its blockCV (Valavi et al., 2019), caret (Kuhn, 2008), mapaccuracy (Costa, 2021), oneClass (Mack, 2014), raster (Hijmans, 2015), RSAGA (Brenning et al., 2018) and rgdal (Bivand et al., 2015) packages.

Results

Two-dimensional space ordination

Globally, the first two axes of the LDA explained 48% and 26% of the between-class variance, respectively (Fig. 3). These results also highlight the importance of environmental and spectral variables in discriminating heathland habitats. Specifically, axis 1 was positively correlated with a warm climate, wet soils and high reflectance values in the SWIR spectrum in early spring, while it was negatively correlated with cool and wet climates, steep topography, large volume of coarse fragments in the soil, organic-rich soils and high reflectance values in the visible spectrum in early spring. The first axis clearly discriminated among alpine heaths (4060) (negative values), dry heaths (4030) and, to a smaller extent, Northern Atlantic wet heaths (4010) and temperate Atlantic wet heaths (4020) (positive values). Axis 2 was positively correlated with sandy soils and negatively correlated with basic and silty soils, and high reflectance values in the summer. The second axis clearly discriminated heathland habitats (positive values) from the other LULC classes (negative values), although overlaps occurred between alpine heaths (4060) and bare rock, but also between dry heaths (4030) or Northern Atlantic wet heaths (4010) and forest and sward.

Classification calibration and accuracy

MaxEnt model parameter and threshold values for each potential and actual heathland habitat varied (Table 3). The feature class parameter HQPT was the most frequent ($n = 3$), although this parameter varied greatly among models. Conversely, values of the regularization parameter were similar and small (<5), as were those for thresholding (<0.25).

Overall, the statistical accuracy of the MaxEnt model for each potential and actual area of heathland habitat was high (AUC > 0.880) (Table 4). These results were supported by low null-model accuracies (AUC 0.36–0.56). Moreover, AUC values for calibration and validation were similar (absolute difference < 0.05), which highlighted low model over-fitting, except for the actual area model of habitat 4020 (absolute difference 0.10).

Accuracy measures of the crisp map of combined classifications of heathland habitats based on the conventional point count approach are presented Table 5. Generally, the combined actual area of the four heathland habitats was accurate (overall F1 score 0.68) (Table 5). Specifically, alpine heaths (4060) were combined with very good accuracy (F1 score 0.80), dry heaths (4030) were combined with moderate accuracy (F1 score 0.58),

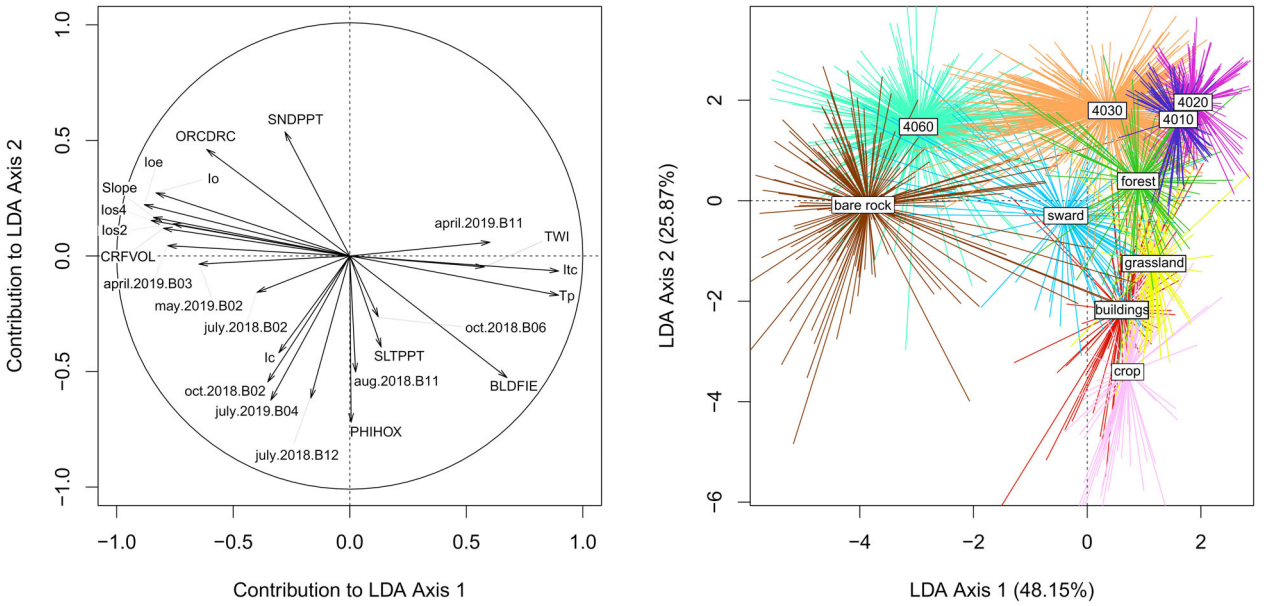

Figure 3. Linear discriminant analysis (LDA) of heathland habitats (see Table 2 for codes) based on environmental and spectral variables (see Table 1 for acronyms). For clarity, only the most variables that, contributed the most (cos2 > 0.2) are shown.

Table 3. MaxEnt model parameter and threshold values set for potential and actual areas of each heathland habitat (codes from the European Union Habitats Directive). (L: Linear, Q: Quadratic, H Hinge, P: Product, T: Threshold).

Area	Habitat	Feature class	Regularization	Threshold
Potential	4010	HQ	1.00	0.243
	4020	LQ	3.50	0.068
	4030	HQPT	3.00	0.199
	4060	HQPT	5.00	0.137
Actual	4010	HQPT	1.75	0.127
	4020	H	3.00	0.208
	4030	T	1.00	0.182
	4060	T	2.50	0.138

Table 4. MaxEnt model accuracy expressed as the area under the curve (AUC) for potential and actual areas of each heathland habitat (codes from the European Union Habitats Directive).

Area	Habitat	AUC calibration	AUC validation	AUC null
Potential	4010	0.958	0.984	0.357
	4020	0.983	0.969	0.444
	4030	0.934	0.919	0.564
	4060	0.984	0.994	0.342
Actual	4010	0.959	0.973	0.527
	4020	0.981	0.879	0.507
	4030	0.941	0.935	0.500
	4060	0.949	0.992	0.518

whereas temperate Atlantic wet heaths (4020) and Northern Atlantic wet heaths (4010) were combined with the lowest accuracies (F1 score 0.46 and 0.29 respectively). Analysis of the confusion matrix highlighted that the main confusion occurred between dry heathlands (4030) and all other classes (Table 5). Interestingly, except for alpine heaths (4060), the heathland habitats were over-detected little among the other LULC classes. Instead, the heathland habitats were under-detected among the others LULC classes. Accuracy measures estimated in terms of proportion of area are shown Table 6. Estimated overall accuracy (0.74) et user's accuracies (0.54–0.79) are similar to those of the point count approach, but estimated producer's accuracies decrease substantially (0.01–0.23).

National maps of heathland habitats

The potential area of the four heathland habitats predicted by MaxEnt varied (Fig. 4). The potential area of Northern Atlantic wet heaths (4010) was fragmented in northern and central France, while that of temperate Atlantic wet heaths (4020) extended logically along the Atlantic coast. The potential area of dry heaths (4030) was the most extensive, while that of alpine heaths (4060) clustered logically within the Alpine biogeographic region. These potential area distributions were consistent with the distribution areas reported under Article 17 of the EU Habitat Directive.

Table 5. Confusion matrix between crisp map of combined classifications of actual heathland habitats (rows) and independent field vegetation points (columns) (codes from the European Union Habitats Directive).

Classification	Reference					User's accuracy
	4010	4020	4030	4060	Others	
4010	7	0	6	0	1	0.50
4020	1	18	6	0	0	0.72
4030	18	19	60	3	8	0.56
4060	0	0	5	71	18	0.76
Others	9	16	21	9	205	0.79
Producer's accuracy	0.20	0.34	0.61	0.85	0.88	
F1 score	0.29	0.46	0.58	0.80	0.83	

Overall F1 score 0.68, overall accuracy 0.72. The shaded regions highlight the diagonale of matrix, which facilitates the comprehension of the results.

Table 6. Accuracy measures of the crisp map of combined classifications of actual heathland habitats estimated using conventional point count and proportion of area approaches (*sensu* Olofsson et al., 2014), respectively.

Class	Point count		Proportion of area	
	PA	UA	PA	UA
4010	0.20	0.50	0.02 ± 0.01	0.54 ± 0.28
4020	0.34	0.72	0.05 ± 0.02	0.72 ± 0.18
4030	0.61	0.56	0.10 ± 0.04	0.59 ± 0.10
4060	0.85	0.76	0.23 ± 0.11	0.79 ± 0.08
Others	0.88	0.79	0.99 ± 0.01	0.74 ± 0.06
OA	0.72		0.74 ± 0.06	

PA, producer's accuracy; UA, user's accuracy; OA, overall accuracy, expressed with a 95% confidence interval for the latter approach.

Subsets of predicted actual area of heathland habitats emphasized the fine-grained pattern of vegetation (e.g. distinguished heathland patches from small patches of sward, forest or bare rock) in the biogeographic regions (Fig. 4). One subset in the Atlantic biogeographic region targeted an alluvial bottom valley site where temperate Atlantic wet heaths (4020) occur (Fig. 5A). Probability of occurrence maps indicated high values for classes 4020 and 4030 but null values for 4010 and 4060. The combined map indicated a dominant pattern for temperate Atlantic wet heaths (4020), with many patches of dry heaths (4030). A second subset focused on an upland depression (elevation ~800 m) in the Continental biogeographic region where Northern Atlantic wet heaths (4010) occur (Fig. 5B). Probability of occurrence maps indicated high values for classes 4010, 4020 and 4030. The combined map thus included a mixture of these three classes, with a strong overestimate of dry heaths (4030). A third subset featured a high elevation site (>2000 m) in the Alpine biogeographic region (Fig. 5C). Probability of occurrence maps indicated positive values only for alpine heaths (4060), while the combined map indicated a large patch of alpine heaths. A fourth subset focused on a site in the Mediterranean biogeographic region dominated by forest (Fig. 5D). Probability of occurrence maps indicated positive values only for dry heathland (4030). Interestingly, the combined map showed patches of dry heath only in areas where forest was not present.

Discussion

Is the quality of open data suitable for mapping heath habitats?

LDA and classification based on open data provided results consistent with the ecological knowledge of heathland habitats. However, besides the usual local biases in the quality of bioclimatic (Cord et al., 2014) and DEM (Moudrý et al., 2018) variables in the mountains, the quality of the Sentinel-2 products used should be discussed more thoroughly. We used monthly cloud-free composite images (L3A processing level) to reduce the total amount of data collected ('only' 1.7 TB for France), but this choice influenced the results. Haze or shadows decreased the quality of certain tiles, which decreased classification accuracy (Lopatin et al., 2019). In addition, the acquisition dates sometimes differed among composite images, which caused a tile effect in the classification. Gap filling and time-series smoothing of Sentinel-2 L2A products could solve these problems, although doing so requires cloud-computing due to the huge amount of data to process at a national scale (Mahdianpari et al., 2020).

The quality of the open-access field data also likely influenced the classification accuracy. Three problems were encountered. The first related to the classification scheme: the definition of each habitat in Annex 1 of the EU Habitat Directive is sometimes confusing and can lead to misinterpretation (Bioret et al., 2017) (e.g. between habitats 4010 and 4020 which share common plant species assemblage). In addition, the definition of habitat 4030 encompasses a wide ecological range and varied physiognomic forms (Bensettiti et al., 2005), which could explain some classification errors with other heathland habitats. Analysis that uses a more detailed nomenclature, such as EUNIS, could address these issues (Agrillo et al., 2021). The second problem related to the geometric quality of field polygons. The outlines of some polygons were approximate and/or included a mosaic of habitats. The third problem related to the 20-year period of field

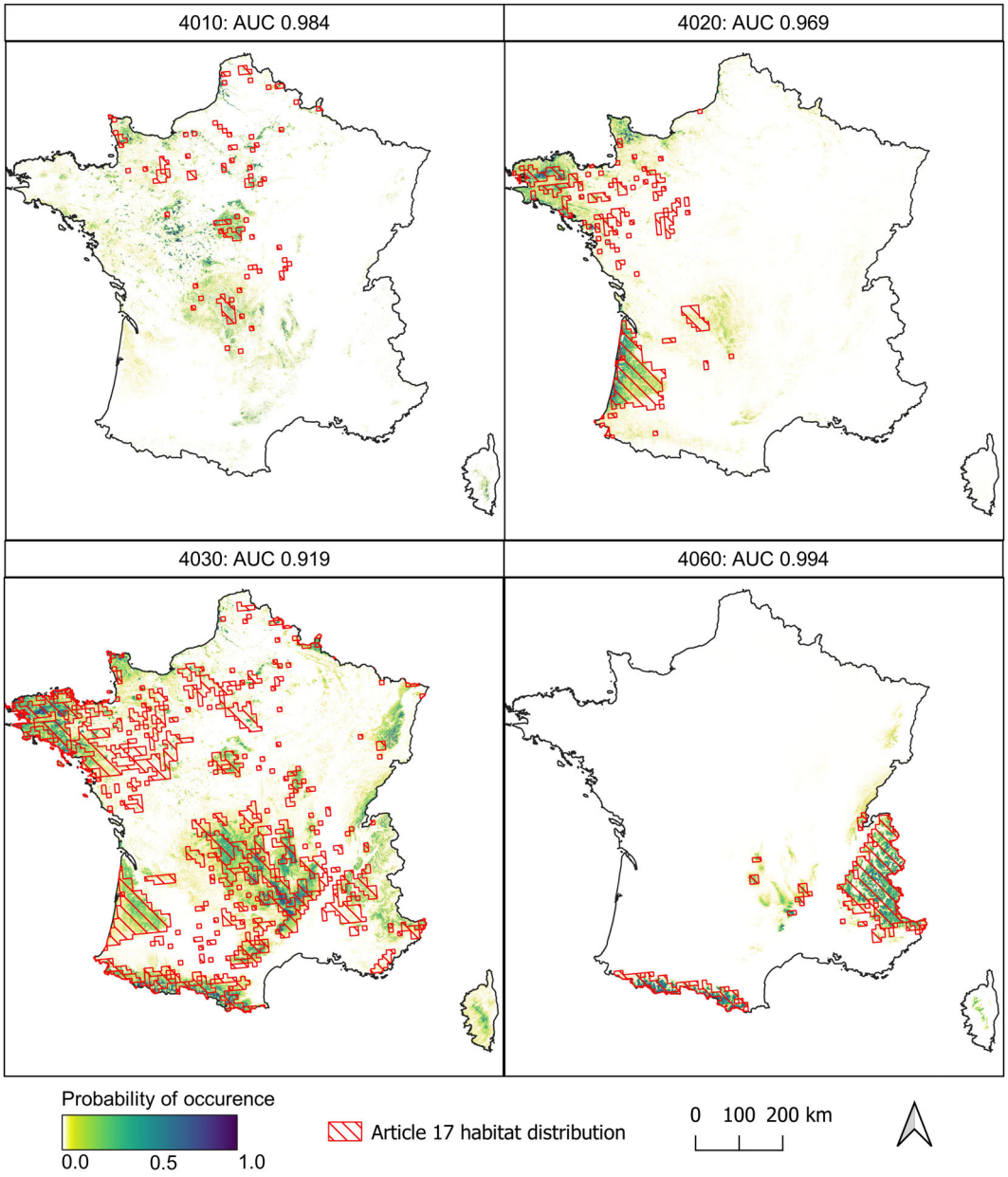

Figure 4. Predicted potential areas of Northern Atlantic wet heaths (4010), Temperate Atlantic wet heaths, (4020), European dry heaths (4030) and Alpine and Boreal heaths (4060) throughout France. For comparison, the distribution of each habitat reported under Article 17 of the European Union Habitat, Directive is highlighted in red hatches. AUC, area under the curve.

maps, as the habitats mapped may have changed between 2000 and the date of Sentinel-2 image acquisition. However, we retained these field maps to ensure enough presence points, especially for the habitat 4010. We addressed the two latter problems by visually interpreting Google Earth images, which was time consuming and somewhat subjective. Regarding these problems, it is unclear whether the false positive pixels identified in this study (Figs. 4 and 5) are false for the following reasons: (i) field surveys outside Natura 2000 sites were not exhaustive and (ii) within Natura 2000 sites, small patches of heathland detected from Sentinel-2 images were not always mapped on the field. As suggested by Stehman and Foody (2019), a new campaign of field data collection stratified using the heathland maps derived from the Sentinel-2 images would provide a more meaningful accuracy assessment. In future research studies, the use of automatic outlier detection (Halladin-Dąbrowska et al., 2020; Liu

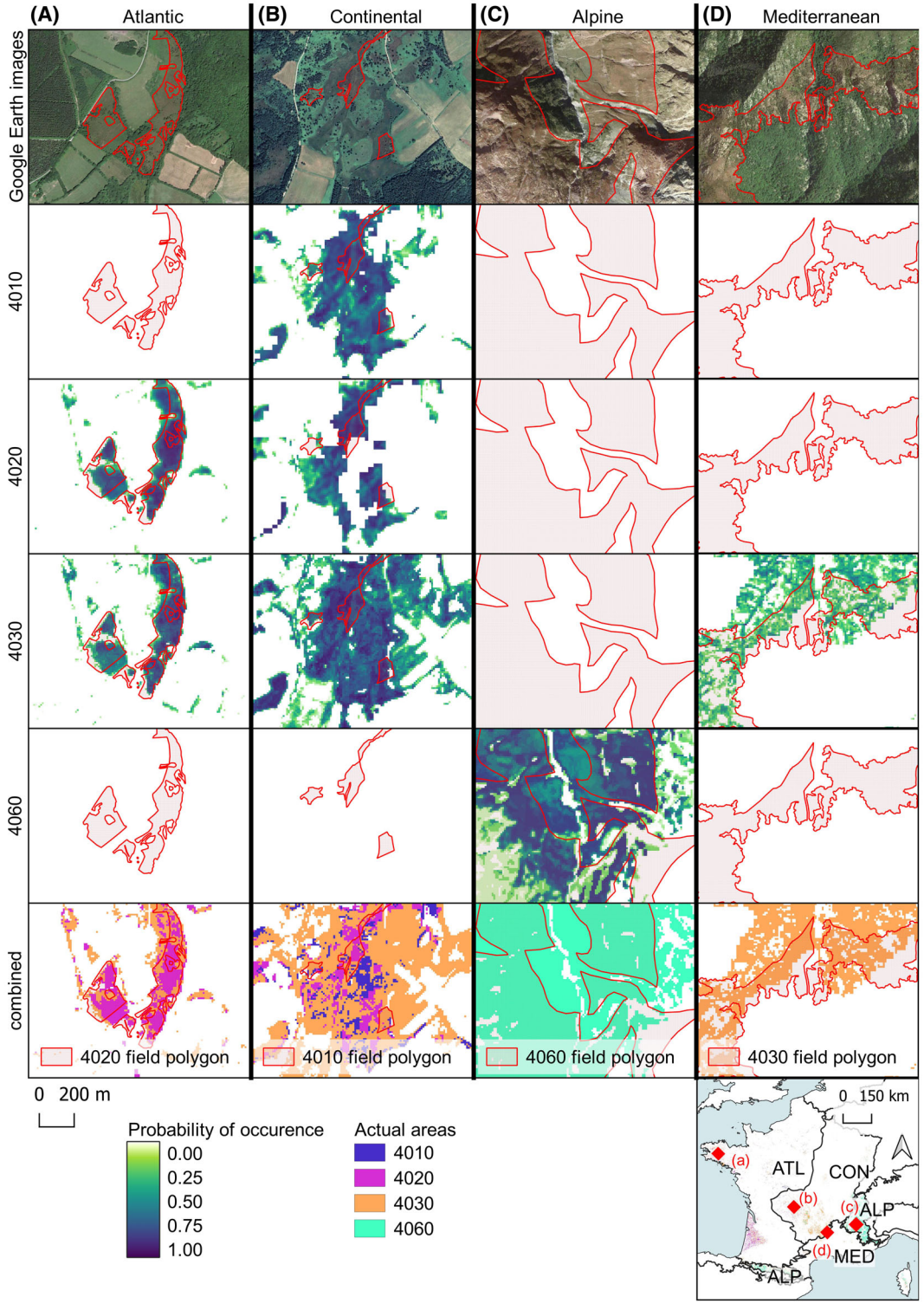

Figure 5. Subsets of the MaxEnt model in the four biogeographical regions (A, B, C, D): probability of occurrence and, combined maps of four actual heathland areas (4010, 4020, 4030 and 4060). The validation points came, from field observations but were not comprehensive in their coverage.

et al., 2018) and active-learning approaches (Neumann, 2020) could increase the quality of the field sample dataset.

Are open-access environmental and vegetation variables sufficient for classifying potential and actual heathland areas?

The LDA emphasized the complementarity of the environmental variables, which discriminated the potential area of each heathland habitat, and the vegetation variables, which discriminated the actual area of heathland habitats from the other LULC classes, as supported by Álvarez-Martínez et al. (2018). Recently, the value of using SoilGrids data to map grassland (Rapinel et al., 2021) and forest (Agrillo et al., 2021) habitats was highlighted. Our study shows that these data can be relevant for identifying low-carbon (ORCDRC) and acidic (PHIHOX) soils favorable to heathland habitats. The LDA highlighted the advantage of the high temporal and spectral resolution of Sentinel-2 data, since many dates (April, May, July, August, October) and spectral domains (blue, green, red, red-edge, SWIR) can be used to discriminate heathland habitats. However, our results also indicate remaining confusion between the actual areas of each habitat and LULC classes with similar physiognomy (e.g. forest, sward). Some of this confusion could be addressed by using a multi-annual Sentinel-2 time series, hyperspectral satellite data (e.g. PRISMA, DESIS, HyspIRI) (Zhang et al., 2021), variables that characterize vegetation structure from SAR (Schmidt et al., 2018) or LiDAR (Álvarez-Martínez et al., 2018) data, or perturbation variables such as snow-cover duration (Gascoin et al., 2019). Regarding the specific spectral response of heathland plant species in late fall (Bayle et al., 2019), we considered that Sentinel-2 images acquired from November 2018 to March 2019 could be useful for increasing heathland habitats discrimination. However, our preliminary analyses that included these images pointed out not only that the generation of a national cloud-free composite was challenging in winter, especially in mountains areas (Schindler et al., 2021), but also that snow cover considerably biased the extent of classified alpine heathlands over swards.

Does applying OCCs at a national scale accurately map heathland habitats?

Despite the large size (114 479 columns × 106 646 rows) and ecological diversity (four biogeographic regions) of the study site, our results demonstrate the value of OCCs for mapping heathland habitat suitability at a national scale. Calibrating the OCCs based on good practice (i.e. considering spatial autocorrelation and hyperparametrization) likely reduced overfitting. The overall accuracy of the crisp combined map based on the point counts approach is satisfactory (0.73) and similar to that obtained by Álvarez-Martínez et al. (2018) (0.67) for a smaller study site that had more classes. However, accuracy measures estimated in terms of proportion of area reveal low producer's accuracies, that is consistent given the strong influence (~95% of the map area) of the other LULC classes in the accuracy estimation. Furthermore, the models were validated using spatially independent samples (60 × 60 km spatial blocks) which decreased the classification accuracy (Karasiak et al., 2021) but provided more reliable results and raised new research questions (Braun, 2021). For example, the misclassification errors occurring in this study between dry (4030) and wet (4010/4020) heathland habitats (Fig. 4) were also reported by Hufkens et al. (2010), who mapped a Natura 2000 site using an airborne hyperspectral image and a linear discriminant least squares classification. These errors could be explained by the spatial coexistence that naturally exists between wet and dry heathland habitats (Hufkens et al., 2010), and enhanced by the ecological continuum occurring between dry and wet heathland (Trodd, 1996).

Regarding the modeling of actual areas, we considered 10 Sentinel-2 spectral bands, including the red edge and SWIR spectra, to exploit the full capability of Sentinel-2 images to characterize the phenology of heathland vegetation (Misra et al., 2020). However, we removed the highly correlated spectral bands for two reasons: (i) removing highly correlated variables does not influence the Maxent modeling accuracy (Feng et al., 2019), and *a contrario* (ii) including a large number of variables in the Maxent model would have resulted in impractical computation times (Schnase et al., 2021).

We used an OCC instead of a multiclass classifier to map heathland habitats without knowing the other habitats present in the landscape (Rapinel & Hubert-Moy, 2021). Indeed, mapping of some habitats of interest using a multiclass classifier would have required masking areas where those habitats are expected to occur (Agrillo et al., 2021; Rapinel et al., 2021; Schindler et al., 2021). However, providing an accurate and comprehensive 'heathland' mask would have been challenging given (i) the one-to-many correspondences between heathland habitats and LULC classes (shrub, sward, forest, grassland, etc.) and (ii) the moderate classification accuracy of some of these LULC classes (Inglada et al., 2017; Venter & Sydenham, 2021). As a result, small patches of heathland habitats (e.g. along roads) that were successfully detected in this study using an OCC, would probably have been omitted using a 'heathland' mask in a multiclass classifier

procedure. However, using an OCC raises three issues. First, using the highest logistic value to combine the classifications is questionable since logistic values of the class with the most samples (4030) were the highest, which may explain why this class was over-detected. While the logistic values of MaxEnt are sensitive to the number of presence points used for the OCC (Merow et al., 2013), the combination of several OCC where the number of points is different from one OCC to another is not straightforward. To overcome this issue, we applied a workaround solution using the same ratio of presence/background points for each OCC (Álvarez-Martínez et al., 2018; Stenzel et al., 2014). However, our results emphasize that this solution does not fully address this issue. Second, the quality of the binary map (presence/absence) is influenced significantly by the thresholding method used (Rapinel & Hubert-Moy, 2021), and the under-detection of heathland habitats among LULC classes in our study could be explained by a higher threshold. Finally, the calibration and prediction time needed (ca. 2 days for each OCC) is proportional to the number of habitats to be surveyed.

What is the conservation value of these classifications?

This study provides the first fine-grained maps of heathland habitats at a national scale. The spatial resolution of the Sentinel-2 images (10 m) revealed small patches of heathland that probably would not have been detected in images with lower spatial resolution (e.g. Landsat-8). While the combined crisp map may be unreliable for estimating the area and range of heathland habitats under Article 17 of the EU Habitats Directive (Röschel et al., 2020), suitability maps are of great value for supporting field surveys but also identifying suitable sites for restoration, especially outside protected areas where knowledge of habitat distribution is still insufficient. Although the spatial resolution of Sentinel-2 images is adequate to identify heathland habitats, the use of airborne remote sensing data with submetric spatial resolution is still needed to finely characterize their inner structure (Schmidt et al., 2017) and dynamics (Mobaied et al., 2015).

Acknowledgments

This study was funded by the French Ministry of Ecology in the framework of the program on the assessment of the conservation status of natural habitats (Article 17 of the Habitat Directive). WorldClim data were collected from https://www.worldclim.org/, SoilGrids data were obtained from https://soilgrids.org/, the EU DEM was released by the Copernicus program (https://land.copernicus.eu/) and the Sentinel-2 data were obtained from CNES (https://theia.cnes.fr/). Field data were standardized at the regional level by National Botanical Conservatories and released by the DREALs. We sincerely thank Farid Bensettiti (Museum National d'Histoires Naturelles), Arnault Lalanne (University of Occidental Brittany), Jan-Bernard Bouzillé and Bernard Clément (University of Rennes) for their valuable expertise.

Funding information

This study was funded by the French Ministry of Ecology in the framework of the program on the assessment of the conservation status of natural habitats (Article 17 of the Habitat Directive).

References

Adamo, M., Tarantino, C., Tomaselli, V., Kosmidou, V., Petrou, Z., Manakos, I. et al. (2014) Expert knowledge for translating land cover/use maps to General Habitat Categories (GHC). *Landscape Ecology*, **29**, 1045–1067. https://doi.org/10.1007/s10980-014-0028-9

Agrillo, E., Filipponi, F., Pezzarossa, A., Casella, L., Smiraglia, D., Orasi, A. et al. (2021) Earth observation and biodiversity big data for forest habitat types classification and mapping. *Remote Sensing*, **13**, 1231.

Alonso, I. & Härdtle, W. (2015) Resolving potential conflicts between different heathland ecosystem services through adaptive management. *Ecological Questions*, **21**, 101–103.

Álvarez-Martínez, J.M., Jiménez-Alfaro, B., Barquín, J., Ondiviela, B., Recio, M., Silió-Calzada, A. et al. (2018) Modelling the area of occupancy of habitat types with remote sensing. *Methods in Ecology and Evolution*, **9**, 580–593.

Balakrishnama, S. & Ganapathiraju, A. (1998) Linear discriminant analysis-a brief tutorial. *Institute for Signal and Information Processing*, **18**, 1–8.

Bartha, G. & Kocsis, S. (2011) Standardization of geographic data: the European inspire directive. *European Journal of Geography*, **2**, 79–89.

Bayle, A., Carlson, B.Z., Thierion, V., Isenmann, M. & Choler, P. (2019) Improved mapping of mountain shrublands using the sentinel-2 red-edge band. *Remote Sensing*, **11**, 2807.

Bensettiti, F., Boulet, V., Chavaudret-Laborie, C. (2005) Connaissance et gestion des habitats et des espèces d'intérêt communautaire. Tome 4 - Habitats agropastoraux (Cahiers d'habitats » Natura 2000). MATE/MAP/MNHN/Ed. La Documentation française, Paris.

Bensettiti, F., Gaudillat, V., Haury, J. (2002) Connaissance et gestion des habitats et des espèces d'intérêt communautaire. Tome 3 - Habitats Humides. (Cahiers d'habitats NATURA 2000). MATE/MAP/MNHN/Ed. La Documentation française, Paris.

Bioret, F., Capelo, J. & Pedrotti, F. (2017) À propos de la cartographie des habitats d'intérêt communautaire de la Directive européenne Habitats FauneFlore 92/43/CE. *Documents Phytosociologiques*, **6**, 447–451.

Bivand, R., Keitt, T., Rowlingson, B. (2015) rgdal: Bindings for the Geospatial Data Abstraction Library.

Bohl, C.L., Kass, J.M. & Anderson, R.P. (2019) A new null model approach to quantify performance and significance for ecological niche models of species distributions. *Journal of Biogeography*, **46**, 1101–1111. https://doi.org/10.1111/jbi.13573

Braun, A.C. (2021) More accurate less meaningful? A critical physical geographer's reflection on interpreting remote sensing land-use analyses. *Progress in Physical Geography: Earth and Environment*, **45**(5), 706–735. https://doi.org/10.1177/0309133321991814

Brenning, A., Bangs, D., Becker, M., Schratz, P., Polakowski, F. (2018) Package 'RSAGA.' The Comprehensive R Archive Network https://CRAN.R-project.org/package=RSAGA

Chytrý, M., Tichý, L., Hennekens, S.M., Knollová, I., Janssen, J.A., Rodwell, J.S. et al. (2020) EUNIS Habitat Classification: expert system, characteristic species combinations and distribution maps of European habitats. *Applied Vegetation Science*, **23**(4), 648–675.

Clair, M., Gaudillat, V., Michez, N., Poncet, L., Poncet, L. (2017) HABREF v3. 1, référentiel des typologies d'habitats et de végétation pour la France. Guide méthodologique. Service du patrimoine naturel, Muséum national d'histoire naturelle, Paris.

Congalton, R.G. & Green, K. (2019) *Assessing the accuracy of remotely sensed data: principles and practices*. Boca Raton, FL: CRC Press.

Cord, A.F., Klein, D., Gernandt, D.S., Perez de la Rosa, J.A. & Dech, S. (2014) Remote sensing data can improve predictions of species richness by stacked species distribution models: a case study for Mexican pines. *Journal of Biogeography*, **41**, 736–748. https://doi.org/10.1111/jbi.12225

Costa, H. (2021) mapaccuracy.

Elith, J., Phillips, S.J., Hastie, T., Dudík, M., Chee, Y.E. & Yates, C.J. (2011) A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions*, **17**, 43–57.

European Environment Agency. (2016). Biogeographical regions.

Feng, X., Park, D.S., Liang, Y., Pandey, R. & Papeş, M. (2019) Collinearity in ecological niche modeling: confusions and challenges. *Ecology and Evolution*, **9**, 10365–10376.

Fenske, K., Feilhauer, H., Förster, M., Stellmes, M. & Waske, B. (2020) Hierarchical classification with subsequent aggregation of heathland habitats using an intra-annual RapidEye time-series. *International Journal of Applied Earth Observation and Geoinformation*, **87**, 102036. https://doi.org/10.1016/j.jag.2019.102036

Fernandez, I.C. & Morales, N.S. (2019) One-class land-cover classification using MaxEnt: the effect of modelling parameterization on classification accuracy. *PeerJ*, **7**, e7016. https://doi.org/10.7717/peerj.7016

Fick, S.E. & Hijmans, R.J. (2017) WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, **37**, 4302–4315.

García González, J.C., Redondo, J.A., Garzón, A. (2015) eu-dem-v1-1-user-guide, Documentation EEA User Manual. Indra Sistemas.

Gascoin, S., Grizonnet, M., Bouchet, M., Salgues, G. & Hagolle, O. (2019) Theia Snow collection: high-resolution operational snow cover maps from Sentinel-2 and Landsat-8 data. *Earth System Science Data*, **11**, 493–514.

Guillera-Arroita, G., Lahoz-Monfort, J.J. & Elith, J. (2014) Maxent is not a presence–absence method: a comment on Thibaud et al. *Methods in Ecology and Evolution*, **5**, 1192–1197. https://doi.org/10.1111/2041-210X.12252

Guisan, A., Weiss, S.B. & Weiss, A.D. (1999) GLM versus CCA spatial modeling of plant species distribution. *Plant Ecology*, **143**, 107–122.

Hagolle, O., Huc, M., Villa Pascual, D. & Dedieu, G. (2015) A multi-temporal and multi-spectral method to estimate aerosol optical thickness over land, for the atmospheric correction of FormoSat-2, LandSat, VENμS and Sentinel-2 images. *Remote Sensing*, **7**, 2668–2691.

Hagolle, O., Kadiri, M., Morin D. & CESBIO (2018). Detailed processing model for the weighted average synthesis processor (WASP) for sentinel-2. https://doi.org/10.5281/zenodo.1401360

Halladin-Dąbrowska, A., Kania, A. & Kopeć, D. (2020) The t-SNE algorithm as a tool to improve the quality of reference data used in accurate mapping of heterogeneous non-forest vegetation. *Remote Sensing*, **12**, 39.

Hengl, T., de Jesus, J.M., Heuvelink, G.B.M., Gonzalez, M.R., Kilibarda, M., Blagotić, A. et al. (2017) SoilGrids250m: global gridded soil information based on machine learning. *PLoS One*, **12**, e0169748. https://doi.org/10.1371/journal.pone.0169748

Hijmans, R.J. (2015) raster: Geographic Data Analysis and Modeling.

Hufkens, K., Thoonen, G., Vanden Borre, J., Scheunders, P. & Ceulemans, R. (2010) Habitat reporting of a heathland site: classification probabilities as additional information, a case study. *Ecological Informatics*, **5**, 248–255. https://doi.org/10.1016/j.ecoinf.2009.09.002

Inglada, J., Vincent, A., Arias, M., Tardy, B., Morin, D. & Rodes, I. (2017) Operational high resolution land cover map production at the country scale using satellite image time series. *Remote Sensing*, **9**, 95.

Karasiak, N., Dejoux, J.-F., Monteil, C. & Sheeren, D. (2021) Spatial dependence between training and test sets: another pitfall of classification accuracy assessment in remote

sensing. *Machine Learning*. https://doi.org/10.1007/s10994-021-05972-1

Kopecký, M. & Čížková, Š. (2010) Using topographic wetness index in vegetation ecology: does the algorithm matter? *Applied Vegetation Science*, **13**, 450–459.

Kuhn, M. (2008) Caret package. *Journal of Statistical Software*, **28**, 1–26.

Kuhn, M. & Johnson, K. (2013) *Applied predictive modeling*. New York, NY, USA: Springer-Verlag.

Liu, C., White, M. & Newell, G. (2013) Selecting thresholds for the prediction of species occurrence with presence-only data. *Journal of Biogeography*, **40**, 778–789. https://doi.org/10.1111/jbi.12058

Liu, C., White, M. & Newell, G. (2018) Detecting outliers in species distribution data. *Journal of Biogeography*, **45**, 164–176. https://doi.org/10.1111/jbi.13122

Lopatin, J., Dolos, K., Kattenborn, T. & Fassnacht, F.E. (2019) How canopy shadow affects invasive plant species classification in high spatial resolution remote sensing. *Remote Sensing in Ecology and Conservation*, **5**, 302–317. https://doi.org/10.1002/rse2.109

Lucas, R., Blonda, P., Bunting, P., Jones, G., Inglada, J., Arias, M. et al. (2015) The Earth Observation Data for Habitat Monitoring (EODHaM) system. *International Journal of Applied Earth Observation and Geoinformation, Special Issue on Earth observation for habitat mapping and biodiversity monitoring*, **37**, 17–28. https://doi.org/10.1016/j.jag.2014.10.011

Lucas, R., Medcalf, K., Brown, A., Bunting, P., Breyer, J., Clewley, D. et al. (2011). Updating the Phase 1 habitat map of Wales, UK, using satellite sensor data. *ISPRS Journal of Photogrammetry and Remote Sensing*, **66**, 81–102. https://doi.org/10.1016/j.isprsjprs.2010.09.004

Mack, B. (2014) oneClass [WWW Document]. oneClass: an R Package for One-Class Classification. https://github.com/benmack/oneClass

Mahdianpari, M., Brisco, B., Granger, J.E., Mohammadimanesh, F., Salehi, B., Banks, S. et al. (2020). The second generation Canadian wetland inventory map at 10 meters resolution using google earth engine. *Canadian Journal of Remote Sensing*, **46**, 360–375.

Malinowski, R., Lewiński, S., Rybicki, M., Gromny, E., Jenerowicz, M., Krupiński, M. et al. (2020) Automated production of a land cover/use map of Europe based on sentinel-2 imagery. *Remote Sensing*, **12**, 3523. https://doi.org/10.3390/rs12213523

Maxwell, A.E., Warner, T.A. & Fang, F. (2018) Implementation of machine-learning classification in remote sensing: an applied review. *International Journal of Remote Sensing*, **39**, 2784–2817. https://doi.org/10.1080/01431161.2018.1433343

Merow, C., Smith, M.J. & Silander, J.A., Jr. (2013) A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography*, **36**, 1058–1069.

Misra, G., Cawkwell, F. & Wingler, A. (2020) Status of phenological research using sentinel-2 data: a review. *Remote Sensing*, **12**, 2760. https://doi.org/10.3390/rs12172760

Mobaied, S., Machon, N., Lalanne, A. & Riera, B. (2015) The spatiotemporal dynamics of forest–heathland communities over 60 years in Fontainebleau, France. *ISPRS International Journal of Geo-Information*, **4**, 957–973.

Moudrý, V., Lecours, V., Gdulová, K., Gábor, L., Moudrá, L., Kropáček, J. et al. (2018) On the use of global DEMs in ecological modelling and the accuracy of new bare-earth DEMs. *Ecological Modelling*, **383**, 3–9. https://doi.org/10.1016/j.ecolmodel.2018.05.006

Neumann, C. (2020) Habitat sampler—a sampling algorithm for habitat type delineation in remote sensing imagery. *Diversity and Distributions*, **26**(12), 1752–1766.

Neumann, C., Weiss, G., Schmidtlein, S., Itzerott, S., Lausch, A., Doktor, D. et al. (2015) Gradient-based assessment of habitat quality for spectral ecosystem monitoring. *Remote Sensing*, **7**, 2871–2898.

Olofsson, P., Foody, G.M., Herold, M., Stehman, S.V., Woodcock, C.E. & Wulder, M.A. (2014) Good practices for estimating area and assessing accuracy of land change. *Remote Sensing of Environment*, **148**, 42–57.

Perrin, G., Rapinel, S., Hubert-Moy, L. & Bioret, F. (2020) Bioclimatic dataset of Metropolitan France under current conditions derived from the WorldClim model. *Data in Brief*, **31**, 105815.

Phillips, S.J., Anderson, R.P. & Schapire, R.E. (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, **190**, 231–259.

Phillips, S.J. & Dudík, M. (2008) Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography*, **31**, 161–175.

R Core Team. (2019) *A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing; 2012. https://www.R-project.org

Rapinel, S. & Hubert-Moy, L. (2021) One-class classification of natural vegetation using remote sensing: a review. *Remote Sensing*, **13**, 1892. https://doi.org/10.3390/rs13101892

Rapinel, S., Panhelleux, L., Lalanne, A. & Hubert-Moy, L. (2021) Combined use of environmental and spectral variables with vegetation archives for large-scale modeling of grassland habitats. *Progress in Physical Geography: Earth and Environment*, 03091333211023689. https://doi.org/10.1177/03091333211023689

Rocchini, D., Petras, V., Petrasova, A., Horning, N., Furtkevicova, L., Neteler, M. et al. (2017) Open data and open source for remote sensing training in ecology. *Ecological Informatics*, **40**, 57–61. https://doi.org/10.1016/j.ecoinf.2017.05.004

Röschel, L., Noebel, R., Stein, U., Naumann, S., Romão, C., Tryfon, E. et al. (2020) State of Nature in the EU-Methodological paper Methodologies under the Nature

Directives reporting 2013–2018 and analysis for the State of Nature 2000.

Schindler, J., Dymond, J.R., Wiser, S.K. & Shepherd, J.D. (2021) Method for national mapping spatial extent of southern beech forest using temporal spectral signatures. *International Journal of Applied Earth Observation and Geoinformation*, **102**, 102408. https://doi.org/10.1016/j.jag.2021.102408

Schmidt, J., Fassnacht, F.E., Förster, M. & Schmidtlein, S. (2018) Synergetic use of Sentinel-1 and Sentinel-2 for assessments of heathland conservation status. *Remote Sensing in Ecology and Conservation*, **4**, 225–239.

Schmidt, J., Fassnacht, F.E., Neff, C., Lausch, A., Kleinschmit, B., Förster, M. et al. (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.

Schnase, J.L., Carroll, M.L., Gill, R.L., Tamkin, G.S., Li, J., Strong, S.L. et al. (2021) Toward a Monte Carlo approach to selecting climate variables in MaxEnt. *PLoS One*, **16**, e0237208.

Stehman, S.V. & Foody, G.M. (2019) Key issues in rigorous accuracy assessment of land cover products. *Remote Sensing of Environment*, **231**, 111199. https://doi.org/10.1016/j.rse.2019.05.018

Stenzel, S., Feilhauer, 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.

Suarez-Seoane, S., Jimenez-Alfaro, B. & Obeso, J.R. (2020) Habitat-partitioning improves regional distribution models in multi-habitat species: a case study with the European bilberry. *Biodiversity and Conservation*, **29**, 987–1008. https://doi.org/10.1007/s10531-019-01922-5

Tarantino, C., Forte, L., Blonda, P., Vicario, S., Tomaselli, V., Beierkuhnlein, C. et al. (2021) Intra-annual sentinel-2 time-series supporting grassland habitat discrimination. *Remote Sensing*, **13**, 277.

Trodd, N.M. (1996) Analysis and representation of heathland vegetation from near-ground level remotely-sensed data. *Global Ecology and Biogeography Letters*, **5**(5), 206–216.

Valavi, R., Elith, J., Lahoz-Monfort, J.J. & Guillera-Arroita, G. (2019) blockCV: an R package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. *Methods in Ecology and Evolution*, **10**, 225–232. https://doi.org/10.1111/2041-210X.13107

Venter, Z.S. & Sydenham, M.A.K. (2021) Continental-scale land cover mapping at 10 m resolution over Europe (ELC10). *Remote Sensing*, **13**, 2301. https://doi.org/10.3390/rs13122301

Zhang, Y. Migliavacca, M. Penuelas, J. & Ju, W. (2021) Advances in hyperspectral remote sensing of vegetation traits and functions. *Remote Sensing of Environment*, **252**, 112121. https://doi.org/10.1016/j.rse.2020.112121

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Figure S1. Variable importance of the four heathland habitats used to identify the potential area (see Table in the article for variable codes).

Figure S2. Variable importance of the four heathland habitats used to identify the actual area (see Table in the article for variable codes).