

This is the final peer-reviewed accepted manuscript of:

Pafumi E.; Petruzzellis F.; Castello M.; Altobelli A.; Maccherini S.; Rocchini D.; Bacaro G.: *Using spectral diversity and heterogeneity measures to map habitat mosaics: An example from the Classical Karst*

APPLIED VEGETATION SCIENCE VOL. 26 ISSN 1654-109X

DOI: 10.1111/avsc.12762

The final published version is available online at:

<https://dx.doi.org/10.1111/avsc.12762>

Terms of use:

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (<https://cris.unibo.it/>)

**When citing, please refer to the published version.**

1 **Using remote sensing to map natural habitats: an integrated approach**  
2 **applied to the Classical Karst eco-mosaic**

3

4 Emilia Pafumi<sup>1,\*</sup>, Francesco Petruzzellis<sup>1</sup>, Miris Castello<sup>1</sup>, Alfredo Altobelli<sup>1</sup>, Simona Maccherini<sup>2</sup>,  
5 Duccio Rocchini<sup>3</sup>, Giovanni Bacaro<sup>1</sup>

6

7 <sup>1</sup>Department of Life Sciences, University of Trieste, Via L. Giorgieri 10, 34127 Trieste, Italy

8 <sup>2</sup>Department of Life Sciences, University of Siena, 53100 Siena, Italy

9 <sup>3</sup>BIOME Lab, Department of Biological, Geological and Environmental Sciences, Alma Mater  
10 Studiorum University of Bologna, via Irnerio 42, 40126 Bologna, Italy

11

12 \*email: emilia.pafumi@studenti.units.it

13

## 14 **Abstract**

15 Remote sensing is a well-established tool for habitat mapping, but its use is still challenging in heterogeneous landscape  
16 mosaics. Novel approaches to improve classification performance include multi-temporal data and multiple remotely sensed  
17 variables. In this study, an integrated approach was developed to map the natural habitats in Classical Karst eco-mosaic (NE  
18 Italy), by quantifying the importance of Spectral Heterogeneity (SH) measures and providing a robust framework to include  
19 multi-temporal remotely sensed data.

20 A collection of 12 monthly Sentinel-2 images was retrieved using the Google Earth Engine platform. Vegetation and SH indices  
21 were computed and aggregated in four temporal configurations: (1) monthly layers of vegetation and SH indices; (2) seasonal  
22 layers of vegetation and SH indices; yearly layers of multi-temporal SH indices computed (3) across the months, and (4) across  
23 the seasons. For each temporal configuration, a Random Forest classification was performed, first with the complete set of  
24 input layers and then with a subset obtained by Recursive Feature Elimination. Training and validation points were  
25 independently extracted from field data.

26 The maximum overall accuracy (OA = 0.72) was achieved with the seasonal temporal configuration, after the number of  
27 habitats was reduced from 26 to 11. SH measures allowed to improve the accuracy of the classification and the spectral  $\beta$ -  
28 diversity was the most important variable in most cases. Spectral  $\alpha$ -diversity and Rao's Q, on the other side, had a low relative  
29 importance, possibly due to the small spatial extent of the habitats. Regarding the inclusion of multi-temporal data, the  
30 aggregation of monthly data in seasonal median composites proved to be the best approach, since it allowed to reduce the  
31 number of input layers without losing accuracy. The approach developed in this study allows to improve habitat mapping in  
32 complex landscapes in a cost- and time-effective way, suitable for monitoring applications.

33

34

## 35 **Keywords:**

36 Multi-temporal classification; Random Forest; Sentinel-2; Spectral diversity; Spectral heterogeneity; Vegetation indices

37

38

## 39 **Highlights**

- 40 • Spectral heterogeneity measures increase the accuracy of image classifications.
- 41 • The most important variable in most classifications is spectral  $\beta$ -diversity.
- 42 • Spectral  $\beta$ -diversity mainly distinguishes woodlands, grasslands and pine forests.
- 43 • Spectral  $\alpha$ -diversity and Rao's Q index have a lower importance.
- 44 • Aggregating data in seasonal composites is a reliable way to reduce dimensionality.

45

## 46 **Abbreviations:**

47	BC	Bray-Curtis dissimilarity
48	DT	Decision Tree
49	GNDVI	Green Normalized Difference Vegetation Index
50	I-0	Grassland encroachment level 0 (pure grassland)
51	I-1	Grassland encroachment level 1
52	I-2	Grassland encroachment level 2
53	IRECI	Inverted Red Edge Chlorophyll Index
54	LAI	Leaf Area Index
55	LiDAR	Light Detection And Ranging
56	NDVI	Normalized Difference Vegetation Index
57	NDWI	Normalized Difference Water Index
58	NIR	Near Infra-Red
59	OA	Overall Accuracy
60	OOB	Out Of Bag
61	PA	Producer's accuracy
62	PCA	Principal Component Analysis
63	PCoA	Principal Coordinate Analysis
64	RF	Random Forest
65	RFE	Recursive Feature Elimination
66	SH	Spectral heterogeneity
67	SWIR	Short Wave Infra-Red
68	UA	User's accuracy

69

# 1. Introduction

Mapping natural habitats is a fundamental step for the conservation of biodiversity. The Habitats Directive, for example, requires EU member states to conserve habitats and species “of community interest” and assess their conservation status every six years, by reporting on parameters such as habitat area, range, indicators of habitat quality and future provisions for habitat survival (European Commission, 2005). These reports require habitat mapping. However, habitat maps have traditionally been produced through time-consuming and costly field surveys, that make them unsuitable to regular updates. Thus, more cost- and time-effective monitoring strategies are required, and remote sensing has become an essential tool for this objective (Corbane et al., 2015).

Habitat mapping by remote sensing is generally carried out through the process of automatic image classification, in which all pixels in a remotely sensed image are categorized into classes of ground cover (Borra et al., 2019). Over time, many remote sensing data have become available, including multispectral and hyperspectral satellite images, and data from active sensors such as radar (Richards, 2013), while image processing tools have been improved, allowing to map a broad range of habitats, such as forests, grasslands, heathlands and wetlands (Corbane et al., 2015).

Despite the advances in this field, mapping some types of habitats remains a difficult task, especially in heterogeneous areas. Mosaics of natural and semi-natural grasslands, for example, are particularly challenging to map, due to the typical small spatial extent of the habitat patches, their spectral similarity, and the high spatial, structural and temporal variability of the vegetation (Corbane et al., 2015; Tarantino et al., 2021). This is complicated by the fact that boundaries between the patches are often not discrete (Rocchini et al., 2013b). Thus, innovative approaches should be tested (Schuster et al., 2015).

The use of multi-temporal data has been proven to facilitate the differentiation of habitats in areas with a seasonal variability (Rapinel et al., 2019; Schuster et al., 2015). This approach, indeed, accounts for phenological differences among vegetation types, that can be the key to distinguish spectrally similar habitats, especially when the most appropriate dates are selected (Senf et al., 2015). However, there are many possible ways to include the multi-temporal information in the classification process. For example, Schuster et al. (2015) found that the accuracy of a classification in grassland habitats was increased by the number of used images, but with differences according to the type of data source. Tarantino et al. (2021) compared the effect of using a time series of a single vegetation index and a set of Sentinel-2 images and found that the first method outperformed the latter. In another study, multiple Sentinel-2 seasonal composites were compared, and the highest accuracy was achieved using the summer mean composite (Praticò et al., 2021).

Image classification outcomes can also be improved by the integration of ancillary data, that modern classification algorithms are able to handle (Wulder et al., 2018). Topographic features such as slope and aspect are often relevant, since they influence the distribution of natural communities on fine scales (Bhatt et al., 2022). Data on vegetation structure derived from active sensors like LiDAR (Light Detection And Ranging) can also facilitate habitat mapping, as demonstrated for example for a semi-arid region of Brazil by da Silveira et al. (2018) and for non-forest Natura 2000 habitats in Poland by Osińska-Skotak et al. (2021). However, some of the greatest improvements in image classifications are achieved when texture information is

103 included, as was highlighted in a recent meta-analysis (Khatami et al., 2016). Image texture metrics measure the spatial  
104 arrangement and variation of pixel values, and thus provide valuable information on the homogeneity of areas (Haralick et al.,  
105 1973). For this reason, they can facilitate the differentiation of spectrally similar habitats (e.g. Bhatt et al., 2022).

106 The spatial variability of the remotely sensed signal is also the basis for the assessment of plant biodiversity from remote  
107 sensing (Rocchini et al., 2010a). The so-called spectral diversity, or spectral heterogeneity, has been directly related to  
108 environmental heterogeneity by the Spectral Variation Hypothesis (Palmer et al., 2002; 2000). Moreover, spectral heterogeneity  
109 can be considered a proxy for species diversity (Rocchini et al., 2010a), because environmentally heterogeneous areas have a  
110 large number of niches available and are expected to host a high species diversity (Stein et al., 2014). The relationship between  
111 spectral heterogeneity and species diversity has proved to be sensitive to many factors (Wang and Gamon, 2019), like spatial  
112 scale (Oldeland et al., 2010; Wang et al., 2018), spectral resolution (Rossi et al., 2021) and temporal scale (Fauvel et al., 2020),  
113 thus it cannot be considered universally valid (Fassnacht et al., 2022). However, spectral heterogeneity can be useful regardless  
114 of its relation with taxonomic diversity, since it encompasses also a functional and a phylogenetic dimension of biodiversity  
115 (Wang and Gamon, 2019).

116 Many indices have been proposed as measures of spectral heterogeneity (Wang and Gamon, 2019). The most traditional ones  
117 include metrics of variability of single wavebands or vegetation indices such as NDVI (Gillespie, 2005; Levin et al., 2007),  
118 and metrics that condense full-spectrum variability, such as the distance from spectral centroid (e.g. Palmer et al., 2002;  
119 Rocchini, 2007). Recently, two novel approaches have emerged to estimate spectral heterogeneity. The first one relies on  
120 information theory: diversity indices based on information theory are computed from spectral data, generally by applying the  
121 moving window approach (Rocchini et al., 2013a). The most common of these indices is Shannon entropy (Shannon, 1948),  
122 computed by considering the relative abundance and richness of reflectance values. However, indices that consider also the  
123 spectral distance among pixel values have some advantages, as was recently highlighted by Thouverai et al. (2021). Rao's  
124 quadratic entropy has been proposed for this reason and proved to perform well in natural areas (Rocchini et al., 2021a).

125 The second novel and powerful approach to estimate spectral heterogeneity is based on "spectral species", i.e. spectral types  
126 considered as proxies for biological species (Féret and Asner, 2014). Following this approach, each pixel of the image is  
127 assigned to a spectral species, generally through unsupervised k-means clustering, thanks to the fact that pixels from the same  
128 species tend to converge to the same cluster (Féret and Asner, 2014). The spatial variation in spectral species is then used to  
129 infer metrics of  $\alpha$ - and  $\beta$ -diversity (Féret and Boissieu, 2020). So far, the spectral species method has been applied to tropical  
130 forests, based on very high-resolution airborne imaging spectroscopy (Féret and Asner, 2014), to low-resolution MODIS  
131 images of the entire Europe (Rocchini et al., 2021c), and recently also to Sentinel-2 data (Féret and Boissieu, 2020) to assess  
132 biodiversity changes in secondary forests (Chraïbi et al., 2021) and to estimate plant diversity in an ecological network (Liccari  
133 et al., 2022).

134 In this light, measures of spectral heterogeneity have the potential to improve habitat mapping frameworks. Indeed, when  
135 vegetation types share similar spectral reflectance characteristics, considering additional levels of information may facilitate  
136 their differentiation (e.g. Bhatt et al., 2022). The variability of taxonomic, functional and phylogenetic traits, as expressed by

137 spectral heterogeneity (Wang and Gamon, 2019), may be such a type of information. Thus, including spectral heterogeneity  
138 measures in image classification procedures could increase their robustness and accuracy, especially in complex landscape  
139 mosaics. However, very few studies have tried to incorporate these measures (e.g. Marzialetti et al., 2020).

140 Interestingly, both Rao's entropy and spectral species-based metrics can be assessed in the temporal dimension. If a multi-  
141 temporal stack is provided as input instead of a multi-spectral image, in fact, temporal diversity will be computed (Marzialetti  
142 et al., 2020). This spectral temporal diversity will likely be useful to assess biodiversity, since differences in phenology can be  
143 important to estimate plant diversity (Fauvel et al., 2020).

144 Moving forward from these premises, the aim of this study was to test and discuss an integrated approach to map a complex  
145 mosaic of natural and semi-natural habitats through remote sensing, using the Classical Karst as case study. Specifically, the  
146 main objectives were:

- 147 1) quantify the importance of measures of spectral heterogeneity for habitat classification;
- 148 2) provide a robust framework to include multi-temporal remotely sensed data for habitat classification.

149 To achieve these goals, multiple sets of remote sensing derived variables, namely vegetation indices and spectral heterogeneity  
150 indices, along with their variation over one year, were computed based on a series of Sentinel-2 images covering the period  
151 March 2021 - February 2022. These variables were aggregated in four temporal configurations, for which separate  
152 classifications were performed. Classification accuracies were compared to find the most reliable approach.

153

## 154 **2. Materials and methods**

### 155 **2.1. Study area**

156 The study was carried out in the Italian part of the Classical Karst, a limestone plateau, with altitudes ranging from 0 to 600 m,  
157 located in the provinces of Trieste and Gorizia within Friuli-Venezia Giulia region (NE Italy; Fig. 1). Seven different territorial  
158 disjunct patches were considered, that cover a total surface of 60 ha and are involved in a restoration project called "Ecomosaico  
159 del Carso" (Appendix). These areas are partially included in two Natura 2000 sites: the special area of conservation "Carso  
160 Triestino e Goriziano" (IT3340006) and the special protection area "Aree carsiche della Venezia Giulia" (IT3341002).

161 Land cover is characterized by a fine mosaic of natural and semi-natural habitats, created by the long-lasting human presence  
162 in the region. The main vegetation types are grassland, downy oak woodland and black pine plantation. Karst grassland is an  
163 extremely species-rich gramineous herbaceous formation that evolved with the millenary action of grazing and is now being  
164 replaced by shrublands and woodlands due to pasture abandonment. Downy oak woodlands are expanding in abandoned  
165 pastures and cover 70% of the Karst nowadays. Black pine has been planted since the mid 19<sup>th</sup> century for reforestation purposes  
166 and from then on has spontaneously expanded creating species-poor pine forests (Poldini, 2009, 1989). Many conservation  
167 projects are being developed in recent years to maintain and restore Karst grassland (Marin and Altobelli, 2021), that is also a  
168 habitat of community interest (code 62A0 "Eastern sub-Mediterranean dry grasslands (*Scorzoneratalia villosae*)" in Annex I  
169 of the Habitat Directive; European Commission, 1992).

170 The climate is transitional between Mediterranean and continental (Poldini, 1989). The average rainfall is 1200 mm/year, and  
171 the mean annual temperature is 12.5°C, although there are large differences due to elevation and slope exposure (OSMER,  
172 2015). The dry and cold Bora from NE contributes to desiccation and soil erosion (Poldini, 1989).

173

## 174 **2.2. Field data collection**

175 Field surveys were carried out between March and May 2022. Habitats present in the intervention areas of “Ecomosaico del  
176 Carso” project were identified on the field. In a first phase, habitats were identified as vegetation types with a high level of  
177 detail, in most cases as associations, according to the phytosociological types described for Classical Karst by Poldini (1989;  
178 2009). In a second phase, habitats were defined on the basis of their structural-physiognomic and ecological characteristics,  
179 and some of the previous classes were aggregated into coarser classes. The two classifications account respectively for 26 and  
180 11 habitat classes. Specifically, for the first classification process, different classes of grassland were distinguished according  
181 to the following criteria:

- 182 1) type of grassland: thermophilous, mesophilous, on flysch;
- 183 2) degree of felting (i.e., presence of *Sesleria autumnalis*): pure grassland (no *S. autumnalis*), first degradation stage  
184 (few patches of *S. autumnalis*), second degradation stage (mosaic with ca. 50% grassland elements and 50% *S.*  
185 *autumnalis*), third degradation stage (felted grassland, completely covered by *S. autumnalis*);
- 186 3) stage of vegetational succession: no bushes (zero encroachment level, I-0), few bushes with low height (ca < 1.5 m)  
187 and widely spaced (first encroachment level, I-1), medium height bushes (ca 3-4 m) relatively close to each other  
188 (second encroachment level, I-2).

189 In the second classification, only the last criterion was considered, while shrublands, initially differentiated according to the  
190 vegetation type, were aggregated into a single class. The two classes of downy oak woodland – namely, a young class with low  
191 height individuals, and a mature class with individuals higher than 6 m – were also merged. Groves with *Ailanthus altissima*  
192 and *Robinia pseudoacacia* were aggregated into an invasive alien species class, while sessile oak woodlands, black pine  
193 plantations, hay meadows and pasture-grasslands were kept as separate classes. Finally, a grassland-woodland mosaic was  
194 defined as a dynamic stage with patches of grassland and well-spaced patches of woodland. The complete list of habitat classes  
195 considered in this study is in Table A2. Two areas were excluded from the analysis since vegetation could not be classified  
196 according to the defined scheme: area n.17, where vegetation was cut before field surveys, and a portion of area n.6, where a  
197 fire occurred on 14/08/2021.

198 The habitats present in the study areas were manually mapped based on field collected data using QGIS 3.16.14 software (QGIS  
199 Development Team, 2022). Maps of vegetation height derived from LiDAR data were used as a base for polygon drawing.  
200 LiDAR RAFVG survey has been conducted in Friuli-Venezia Giulia region in the years 2017-2020 by aerial means. LiDAR  
201 point clouds, that have an average density of 16 points/m<sup>2</sup>, were downloaded from Eagle FVG portal  
202 (<https://eaglefvg.regione.fvg.it>). Each point includes a classification field (1 – unclassified, 2 – ground, 3 – low vegetation, 4 –  
203 medium vegetation, 5 – high vegetation): points belonging to “Ground” class were extracted and interpolated to create a plan,



204 then the distance of the vegetation points from the plan was computed and maps of vegetation height were produced.  
205 Elaboration was performed in CloudCompare 2.11.1 (Cloud Compare, 2021).

206

### 207 **2.3. Satellite data collection and processing**

208 The workflow applied to manage satellite data and to derive input variables for classification is represented in Fig. 2. First,  
209 Sentinel-2 images were retrieved using Google Earth Engine platform (Gorelick et al., 2017). The Sentinel-2 level-2A image  
210 collection (“COPERNICUS/S2\_SR\_HARMONIZED”) was filtered by date (from 2021-03-01 to 2022-02-28), by area (the  
211 Trieste and Gorizia Karst) and by cloud coverage (cloudy pixel percentage < 50%). The less cloudy image of each month was  
212 manually selected, to produce a collection of 12 monthly images covering a whole year.

213 Then, the 12 Sentinel-2 images were divided into four seasonal groups: spring (March 2021-May 2021), summer (June 2021-  
214 August 2021), autumn (September 2021-November 2021), and winter (December 2021-February 2022). Each group was  
215 reduced to a single image by computing the median of each spectral band, so that, at each location in the output image, the  
216 pixel value of a band is the median of all pixel values of that band in the input group.

217

### 218 **2.4. Vegetation indices**

219 Vegetation indices were preferred over the original Sentinel-2 spectral bands as inputs for the classification because they allow  
220 to reduce the dimensionality of the dataset while being more strongly related to the temporal variation of vegetation (Coppin  
221 et al., 2004). Four vegetation indices were derived from each Sentinel-2 image (Table 1): Normalized Difference Vegetation  
222 Index (NDVI), Green Normalized Difference Vegetation Index (GNDVI), Normalized Difference Water Index (NDWI) and  
223 Inverted Red Edge Chlorophyll Index (IRECI).

224 NDVI (Rouse et al., 1975) includes the red and NIR bands, respectively sensitive to chlorophyll content and leaf structure. It  
225 has been proven to be correlated to biomass, LAI and photosynthetic activity (Gamon et al., 1995). GNDVI is an alternative to  
226 NDVI, with the green band instead of the red band, that has been proposed to avoid saturation in case of high chlorophyll  
227 content (Gitelson et al., 1996). NDWI, including the NIR and SWIR bands, is sensitive to water content and can be useful in  
228 assessing vegetation water status (Chen et al., 2005). Finally, IRECI uses Sentinel-2 red and red-edge bands and is very sensitive  
229 to LAI parameter and canopy chlorophyll content (Frampton et al., 2013).

230 These indices were computed from each image in the monthly dataset and then aggregated into seasonal median composites,  
231 following the same procedure used for Sentinel-2 reflectance bands. These operations were performed in Google Earth Engine.

232

### 233 **2.5. Metrics of spectral heterogeneity and diversity**

234 Rao’s quadratic entropy (Rao’s Q; Rao, 1982) is a diversity index that considers both relative abundances of pixel values ( $p_i$ ,  
235  $p_j$ ) and spectral distances among them ( $d_{ij}$ ):

236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266

$$Q = \sum_{i=1}^N \sum_{j=1}^N d_{ij} \times p_i \times p_j$$

In this study, spectral Rao's Q layers were separately obtained from each vegetation index raster through package `rasterdiv` (Rocchini et al., 2021c; Thouverai et al., 2021) in R software (R Core Team, 2022). Following the function specifications, each input raster layer was first rescaled in 8-bit and the moving window was set to 3.

Then, Rao's Q was computed also in the multi-temporal dimension, by setting a "multidimension" method. In this case, the distances among pixel values are calculated considering more than one layer. For each vegetation index, two layers of multi-temporal Rao's Q were produced, one from the stack of 12 monthly images and the other from the stack of 4 seasonal composites.

Spectral  $\alpha$ - and  $\beta$ -diversity metrics were calculated following the spectral species concept proposed by Féret and Asner (2014), using the R package `biodivMapR` (Féret and Boissieu, 2020). The used algorithm includes several steps: first, the multi-spectral images are filtered to remove irrelevant pixels (non-vegetated, shady, or cloudy). Then, a principal component analysis (PCA) is performed, and the relevant principal components (PCs) are manually selected based on visual analysis. Spectral species mapping is based on k-means clustering. A subset of pixels is randomly extracted from the image and used to define k clusters (i.e., "spectral species"). The number of clusters was set to 20 in this study as it was suggested as the optimal for moderately diverse temperate sites (Féret and Boissieu, 2020). Then, clustering is applied to the whole image, so that each pixel is assigned to a cluster. Finally,  $\alpha$ - and  $\beta$ -diversity maps are produced, basing on the distribution of spectral species in the window size, that was set to 3x3 pixels in this case, since habitat patches were small. Shannon index was chosen as indicator of  $\alpha$ -diversity, while  $\beta$ -diversity was derived from a pairwise Bray-Curtis (BC; Bray and Curtis, 1957) dissimilarity matrix obtained from the spectral species map. The BC matrix was then subjected to an ordination (Principal Coordinate Analysis, PCoA) to project it into a 3-dimensional space and obtain a visual representation of the results (larger BC dissimilarity between pixels corresponds to larger color differences in the RGB space).

This algorithm was applied on each Sentinel-2 monthly and seasonal image, to obtain maps of  $\alpha$ - and  $\beta$ -diversity for each month and season. Then, this procedure was applied on a multi-temporal level, using stacks of vegetation indices covering a whole year as input. For each vegetation index, a stack of 12 layers (each corresponding to a month) and then a stack of 4 layers (each corresponding to a season) were used as input. In this way, two sets of multi-temporal  $\alpha$ - and  $\beta$ -diversity maps were obtained for each vegetation index, one based on monthly values (multi-temporal monthly) and the other based on seasonal values (multi-temporal seasonal).

## 2.6. Satellite image classification

The remote sensing variables produced through the processing steps outlined before were aggregated in four temporal configurations, as listed in Table 2, that were used as input for distinct classifications.

267 Reference data for image classification were derived from field surveys. Training points were randomly extracted from a set  
268 of training areas selected on the field, outside the polygons of “Ecomosaico del Carso” project when possible, and mapped by  
269 acquiring their GPS location. Validation points, on the other hand, were randomly launched in the whole set of polygons created  
270 in QGIS after excluding the training areas. Through this procedure, training and validation points can be considered as  
271 independent.

272 Image classification was performed using a Random Forest (RF) classification algorithm. Random Forest (Breiman, 2001) is  
273 an ensemble machine-learning classifier that builds a large number of decision trees (DTs), each based on a random subset of  
274 the training data and of the predictor variables. The training data not used to build the model (i.e., the out-of-bag data, OOB)  
275 are used to evaluate the model performance. The results of the different DT models are then averaged to assign each pixel to a  
276 class. In this way, the overall result is more reliable than the one obtained from an individual DT and is less affected by  
277 correlation among predictors (Maxwell et al., 2018). The relative importance of the predictor variables is computed using the  
278 OOB data, by systematically comparing the performance of the DTs that include specific variables and of those that do not:  
279 variables with high importance have a positive effect on the prediction accuracy (Breiman, 2001).

280 For each temporal configuration, two alternative pathways were followed. In one case, the whole set of variables was used as  
281 input for the classification. In the other case, a subset of variables was extracted through Recursive Feature Elimination (RFE;  
282 (Guyon et al., 2002). RFE is a common feature selection algorithm based on backward elimination, that uses a RF classifier to  
283 determine variables permutation importance and remove the less important variables. The importance measures are updated  
284 after each deletion, making the method suitable also to highly correlated variables (Gregorutti et al., 2017).

285 In this study, RF classification was performed using R `caret` package (Kuhn, 2021). Training data were randomly partitioned  
286 into a training and a testing set, with respectively 80% and 20% of the data. The `mtry` parameter (the number of randomly  
287 selected predictors used at each node) was optimized through a 5-fold cross-validation, while the relative importance of  
288 variables was calculated with `varImp` function.

289 The accuracy of each classification procedure was evaluated using a set of validation points independent from the training data.  
290 A confusion matrix was computed, and the proportion of correctly classified pixels (overall accuracy, OA) was derived. OA is  
291 preferable to other common metrics such as the Kappa coefficient because it is easier to understand and more suited for  
292 comparisons (Foody, 2004). Performances for individual classes were also assessed by considering User's accuracy (UA) and  
293 Producer's accuracy (PA). For a given class  $i$ , UA is the proportion of pixels classified as  $i$  that have reference class  $i$ , while  
294 PA is the proportion of pixels of reference class  $i$  that are classified as  $i$  (Borra et al., 2019). Both metrics vary between 0 and  
295 1. The significance of differences in classification accuracy among the different pathways was tested with McNemar's test, as  
296 suggested by Foody (2004).

297 After the classifications were performed as described above and the best input configuration was identified, another  
298 classification was carried out using only vegetation indices as input, to assess the effect of excluding spectral heterogeneity on  
299 the results.

300 All classifications and accuracy assessment analyses were performed in R software (R Core Team, 2022).

301

## 302 **3. Results**

### 303 **3.1. Accuracy of image classification**

304 The values of overall accuracy (OA) and Kappa obtained from the RF classifications are presented in Table 3. The OA was  
305 significantly higher when 11 habitat classes were considered of 26 (p-value < 0.05; Fig. 3a), while there was no significant  
306 difference when the number of input variables was reduced through RFE (Fig. 3b). The use of different temporal configurations  
307 only had a slight effect on accuracy (Fig. 3c): in particular, there were no significant differences between the monthly and the  
308 seasonal configurations, while there were significant differences between each configuration and its respective multi-temporal  
309 version (p-value < 0.05).

310 An OA higher than 70% was achieved only with the monthly and the seasonal configurations, considering 11 habitat classes.  
311 For each of these configurations, an additional classification was performed after removing spectral heterogeneity layers (i.e.,  
312 with only vegetation indices). In both cases, the resulting accuracy was significantly lower (0.65 vs. 0.72 for the seasonal  
313 configuration, p-value < 0.05; 0.69 vs. 0.73 for the monthly configuration, p-value < 0.05).

314 The classifications that achieved an OA > 70% did not differ significantly among them. Thus, the seasonal configuration was  
315 chosen as the best one based on a simplicity criterion, since it had a lower number of predictors, especially after RFE (34  
316 predictors). This classification will be referred to as the “best classification” and explored in the next paragraphs, while the  
317 results of the other classifications are reported in the Appendix.

318 The habitat map resulting from the best classification is represented in Fig. 4. The most common habitat inside the study areas,  
319 as resulting from field surveys, is downy oak woodland (27.94%), followed by grasslands at successional stages I-2 (20.95%)  
320 and I-1 (11.97%). In the best RF classified map, on the other hand, grassland I-2 (24.80%) is more common than downy oak  
321 woodland (21.70%), and is followed by shrubland (16.10%), grassland-woodland mosaic (13.38%) and grassland I-1 (12.89%).  
322 The confusion matrix for the best classification is reported in Table 4. The rows of the matrix represent the results obtained  
323 from the classification, while the columns represent the validation data used as reference; the diagonal contains the correctly  
324 classified pixels. Class-specific accuracy parameters derived from the confusion matrix are reported in Table 5.

325 Black pine plantation was the class for which the best results were achieved considering both Producer’s accuracy (PA = 0.88)  
326 and User’s accuracy (UA = 0.92), followed by downy oak woodland (PA = 0.74, UA = 0.86). Among the grassland classes,  
327 pure grassland and grassland I-1 achieved relatively high UA (respectively 0.80 and 0.78) and PA (respectively 0.71 and 0.68),  
328 and most of the errors occurred with grassland-I2 and downy oak woodland. For grassland I-2, PA was relatively low (0.64)  
329 because some pixels were misclassified as grassland-woodland mosaic, while the UA (0.69) was mainly affected by some  
330 pixels belonging to grassland-I and downy oak woodland. For shrublands, higher values were obtained for PA (0.57) than for  
331 UA (0.21), since many pixels classified as shrubland belonged to grassland I-2 and downy oak woodland. A similar result was  
332 found for hay meadow (PA = 0.71, UA = 0.59), that was mainly confused with grassland I-2. For pasture-grassland, UA (1.00)  
333 was higher than PA (0.13), and most pixels were misclassified as hay meadow. For grassland-woodland mosaic, both UA and

334 PA were quite low (0.28 and 0.72), and most of the errors occurred for pixels that either belonged to or were misclassified as  
335 grasslands and downy oak woodlands. The lowest values of accuracy were obtained for sessile oak woodlands (UA = 0.36, PA  
336 = 0.27), that were mainly confused with downy oak woodlands, and for invasive species groves, for which all validation pixels  
337 were misclassified as shrublands, downy oak woodlands or grassland-II (UA = 0.00, PA = 0.00).

338

### 339 **3.2. Variable importance**

340 The relative importance of the variables used as input for the best classification is shown in Fig. 5. The most important variable,  
341 present in every RF model, is the PCo2 of the  $\beta$ -diversity computed from the autumn composite (100.00%). Other important  
342 variables are, in order, PCo1 of the winter  $\beta$ -diversity (94.19%), PCo1 of the autumn  $\beta$ -diversity (88.93%), GNDVI and IRECI  
343 of the summer (respectively 79.39% and 72.29%).

344 In the other classifications, the most important variable is almost always a  $\beta$ -diversity, with the only exception of the monthly  
345 classification with 11 classes, in which vegetation indices are at the first places. The relative importance of  $\alpha$ -diversity and  
346 Rao's Q indices is low in all the classifications: the maximum values are respectively 42.56% for  $\alpha$ -diversity (in the monthly  
347 26-classes classification) and 64.91% for Rao's Q (in the monthly 11-classes classification). A detailed description of the input  
348 variables is presented in the Appendix.

349

## 350 **4. Discussion**

### 351 **4.1. Accuracy of image classification**

352 In this study, multiple RF classifications were performed to test different combinations of vegetation and spectral heterogeneity  
353 indices, using as study area a complex mosaic of habitats in Classical Karst. The small spatial extent of the habitat patches,  
354 combined with their spectral similarity and the high variability of vegetation, make this type of landscape particularly  
355 challenging to map from remote sensing (Corbane et al., 2015; Tarantino et al., 2021). The maximum overall accuracy achieved  
356 in this study was thus relatively low (0.73 for the 11-class classifications and 0.65 for the 26-class classifications). However,  
357 other studies using similar types of data did not achieve much higher levels of accuracy. For example, Rapinel et al. (2019)  
358 managed to map seven wet grassland plant communities with an accuracy of 0.78, by using Sentinel-2 time series and a SVM  
359 classifier. Tarantino et al. (2021) achieved an accuracy of 0.95, by using a SVM classifier and a set of input variables that  
360 included multi-seasonal Sentinel-2 images, a time series of MSAVI index and a DTM. However, they only mapped four  
361 Mediterranean grassland types. Bhatt et al. (2022), that used very high-resolution imagery (60 cm) to map nine heterogeneous  
362 habitats going from forests to open water, only reached a maximum accuracy of 0.79.

363 Moreover, some additional factors increased the complexity of the classification in the present work. Firstly, the analyzed areas  
364 are distributed over a relatively wide region (the Italian part of the Classical Karst), where the differences in altitude and  
365 substrate composition increase intra-habitat variability (Poldini, 1989). Villoslada et al. (2020) found that the spectral  
366 heterogeneity of the training samples can affect the accuracy of the classification, and that generally homogeneous classes are

367 more accurately mapped. The results observed for Classical Karst seem in line with this finding, since the best performances  
368 were observed for the most spectrally homogeneous habitats, namely black pine plantations and downy oak woodlands. Finally,  
369 most habitat patches in the study area had a small spatial scale, thus the proportion of mixed pixels was probably high, and this  
370 complicated habitat class separation (Rocchini et al., 2013b). The lowest class-specific accuracy, indeed, was found for invasive  
371 alien species groves, that were present in the smallest areas. Mapping invasive alien species from remote sensing is generally  
372 a challenging task, and the most promising results have been obtained using hyperspectral instead of multispectral imagery, to  
373 facilitate the differentiation of target species from others (He et al., 2011; Rocchini et al., 2015). For these reasons, the accuracy  
374 achieved in this study can be considered relatively high.

375 Spectral heterogeneity measures had an important role in improving the capacity of classifying habitats from satellite data. The  
376 classifications performed without spectral heterogeneity measures (maximum OA = 0.69) were significantly less accurate than  
377 the others (maximum OA = 0.73), and the resulting maps also had a noisier aspect. Spectral heterogeneity is mainly investigated  
378 nowadays for its relationship with species richness, that has been tested in many environments (Wang and Gamon, 2019), but  
379 it has rarely been used as an additional level of information for the classification of habitats. The results obtained in this study  
380 suggest that image classification frameworks could benefit from the inclusion of spectral heterogeneity measures, although  
381 with caution about which type of metric is being used, as will be discussed below.

382 The most important variables in almost all the classifications were metrics of spectral  $\beta$ -diversity. In this study, the spectral  $\beta$ -  
383 diversity was referred to the pairwise Bray-Curtis dissimilarity computed from the distribution of spectral species, as defined  
384 by Féret and Asner (2014). Spectral species are distinct spectral entities, that can be considered as proxies for individual plant  
385 species only with very high-resolution remote sensing data (Féret and Asner, 2014). In case of coarser spatial resolutions, such  
386 as the ones used in this study, spectral species can be related to higher levels of biological organization, such as plant  
387 communities or habitats (Rocchini et al., 2021c). Although both  $\alpha$ - and  $\beta$ -diversity in this study were based on the spectral  
388 species approach, the latter was far more important than the former for habitat classification. This can be explained considering  
389 what these metrics represent:  $\alpha$ -diversity measures the diversity within a single community, while  $\beta$ -diversity represents the  
390 degree of differentiation between communities, or their compositional dissimilarity (Whittaker, 1960). Different habitats can  
391 share a similar  $\alpha$ -diversity despite having different species; on the other hand,  $\beta$ -diversity allows more easily to differentiate  
392 habitats based on their dissimilarity. Here, the values of spectral  $\beta$ -diversity projected in the PCoA space clearly separated the  
393 three main groups of habitats present in Karst eco-mosaic: habitats dominated by woody deciduous plants (woodlands and  
394 shrublands), habitats dominated by herbaceous plants (grasslands and meadows) and habitats dominated by evergreens (pine  
395 forests).

396 The use of metrics based on the spectral species approach has some additional advantages. One of the first steps in the algorithm,  
397 indeed, is a PCA (Féret and Asner, 2014), that is one of the most widely used methods in image classifications to reduce feature  
398 dimensionality while maximizing spectral separability (Richards, 2013). Then, a k-means clustering is performed to distinguish  
399 and map the so-called spectral species. To some extent, this procedure is equivalent to a hybrid classification approach, in  
400 which an unsupervised classification is carried out before the application of supervised algorithms, to identify the main groups

401 of pixels basing only on their spectral similarity (Borra et al., 2019). Usually, this step is applied to choose appropriate classes  
402 and guide the collection of training samples (e.g. Lane et al., 2014). This procedure has some advantages: the full spectral  
403 information of satellite is exploited (Baldeck and Asner, 2013), and spectrally extreme pixels do not unproportionally affect  
404 the results, but are simply grouped into separate classes (Fassnacht et al., 2022). Moreover, the computation of spectral  $\beta$ -  
405 diversity requires another ordination (a PCoA), that further maximizes the separation of groups of similar pixels.

406 Moreover, the relationship between spectral and species diversity can be different for  $\alpha$ - and  $\beta$ - components. In many studies,  
407 the estimation of  $\alpha$ -diversity from remote required very high-resolution data (e.g. 1 m<sup>2</sup> in Wang et al. 2016a). In this study, for  
408 example, the spectral  $\alpha$ -diversity was similar in black pine plantations and pasture-grasslands, although they have very different  
409 species richnesses (Poldini, 1989). A study by Fassnacht et al. (2022) also pointed out that, at the spatial resolution of Sentinel-  
410 2 images, intensively used agricultural patches can show higher spectral diversity compared to species-rich grasslands. For  $\beta$ -  
411 diversity, on the other hand, a good agreement between spectral and field-based metrics was obtained also at relative coarse  
412 spatial resolution (e.g. Rocchini et al., 2010b), although few studies focused on this component of biodiversity (Wang & Gamon  
413 2019). For example, Rocchini et al. (2010) found that the relation between field and spectral  $\beta$ -diversity is even greater at larger  
414 grain sizes (20x50 m instead of 10x10 m). In another study (Hoffmann et al., 2019), most of the  $\beta$ -diversity of different plant  
415 communities distributed along an elevational gradient could be explained using Sentinel-2 data with 10 m spatial resolution.  
416 Thus, the link between species and spectral diversity seems to be generally stronger for  $\beta$ - than for  $\alpha$ -diversity.

417 The other spectral heterogeneity index considered in this study, spectral Rao's Q, had a low relative importance in all the  
418 classifications. This index is a measure of the heterogeneity of a pixel with respect to its surroundings (Thouverai et al., 2021),  
419 and has been proven to match species diversity in natural areas but not in agricultural lands with high heterogeneity (Rocchini  
420 et al., 2021b). In this study, the lowest Rao's Q values were found for black pine plantations and downy oak woodlands, that  
421 do host a low species diversity, while the highest values were found for pasture-grasslands and pure grasslands, that are species-  
422 rich habitats (Poldini, 2009). However, high values of Rao were found also for invasive species groves, thus the relation  
423 between spectral Rao's Q and species diversity was not clear. One possible reason is related to the spatial extent of the habitat  
424 patches: in this study, Rao's Q was computed with moving windows of 3x3 pixels (900 m<sup>2</sup>), thus, the habitats present in smaller  
425 patches were more likely to border with other habitats inside this window, resulting in higher spectral heterogeneity (i.e., high  
426 Rao's Q values). Using remote sensing data with higher spatial resolutions would probably improve this aspect.

427 However, the approach used to calculate Rao's Q may itself be a problem, since it highlights the differences among close  
428 pixels, and thus maximizes the noise, instead of minimizing it. Therefore, while the Rao's Q index can be used to estimate  
429 species diversity in some cases (Rocchini et al., 2021b), it might be less useful in the context of habitat mapping.

430 The results of this study show that some spectral heterogeneity metrics might be more useful than others in the context of  
431 habitat mapping. The relationship between spectral and species diversity is still not clear in many cases, but, as pointed out by  
432 Fassnacht et al. (2022), these measures can be useful regardless of their link with actual species diversity, since they allow to  
433 exploit the main strength of remote sensing: measures can be repeated over time, to capture habitat specific variations and  
434 monitor landscape evolution.

435

## 436 **4.2. Importance of vegetation indices**

437 Vegetation indices were the most important variables after  $\beta$ -diversity metrics in all the monthly and seasonal classifications.  
438 In particular, summer GNDVI, summer IRECI and autumn NDVI were the most important vegetation indices in the best  
439 classification. NDVI, with its variant GNDVI, is the most commonly used index and has been found useful in many studies  
440 (e.g. Schuster et al., 2015).

441 IRECI is the only index considered in this study that includes the Red Edge Sentinel-2 bands and has a strong linear relationship  
442 with canopy chlorophyll content and LAI (Frampton et al., 2013). The results presented here seem to confirm this relationship.  
443 Indeed, the temporal variation of IRECI follows the expected seasonal changes of canopy chlorophyll content, with an increase  
444 in spring, a maximum in summer and a decrease in autumn (Gara et al., 2019). In summer, the highest values were found for  
445 downy oak woodlands, in agreement with the fact that broadleaves species have a higher chlorophyll content compared to  
446 conifers (Li et al., 2018). Conversely, in winter IRECI was relatively high only for the evergreens black pine plantations.  
447 Moreover, the differences of IRECI among habitats might also reflect the variation of LAI across ecosystems: mean LAI  
448 generally increases from grasslands ( $1.7 \pm 1.2$ ) and shrublands ( $2.1 \pm 1.6$ ), to temperate deciduous broadleaves ( $5.1 \pm 1.6$ ) and  
449 evergreen needleleaves ( $5.5 \pm 3.4$ ) forests (Asner et al., 2003). Optical traits like chlorophyll content can improve the estimation  
450 and mapping of species composition over space, as demonstrated by Feilhauer et al. (2017) in semi-natural temperate  
451 grasslands. Although IRECI itself has not been tested much in the context of habitat mapping, other indices using the Red Edge  
452 spectrum have been shown to be useful for this purpose. For example, Schuster et al. (2012) found that the Red Edge channel  
453 of the RapidEye satellite had a positive influence on the overall accuracy of a land cover classification in a mosaic of natural  
454 and agricultural areas in Germany, especially for the bush vegetation and dry grassland classes. In another study, Alpine  
455 grasslands were distinguished from shrublands relying on the Sentinel-2 Red Edge bands, by detecting the seasonal anthocyanin  
456 accumulation in the shrub species (Bayle et al., 2019). A Red Edge-based index was also found to be more useful than NDVI  
457 to map plant communities in coastal meadows (Villoslada et al., 2020). These examples are in line with the results of this study,  
458 that confirm the role of the Red Edge spectrum for the distinction of habitats.

459

## 460 **4.3. Inclusion of multi-temporal data**

461 The aggregation of monthly data in seasonal composites using the median statistical operator allowed to reduce the number of  
462 input layers without losing information. The levels of accuracy achieved with the monthly and the seasonal temporal  
463 configurations, indeed, were not significantly different, while the number of input layers was reduced from 144 to 48. This  
464 method of reducing data dimensionality can be complemented with variable selection through RFE, that did not have a  
465 significant effect on accuracy. The use of seasonal composites for habitat mapping is known to be useful because it reduces the  
466 problem of cloudy images but maintains the advantage given by multi-temporal data (Kollert et al., 2021). In a recent work by  
467 (Praticò et al., 2021), the mean turned out to produce slightly better results than other statistical operators such as the median.



468 In this study the median was chosen because it is less sensitive to outliers and is generally the most common way to perform  
469 image reduction (Kollert et al., 2021), but other statistical operators could be investigated.

470 The multi-temporal configurations generally led to worse results than the other configurations, as was evident both from the  
471 accuracy values (mean OA = 0.59 for 26 classes and 0.65 for 11 classes) and from the visual assessment of the classified maps.

472 The temporal Rao's Q computed for different vegetation indices over a year was successfully used by Marzialetti et al. (2020)  
473 to map coastal dune habitats, but also the mean, the 10<sup>th</sup> and the 90<sup>th</sup> percentiles of vegetation indices were included in that  
474 case. In this study, only temporal heterogeneity layers were used as input in the multi-temporal configurations, and this may  
475 have reduced the capacity of distinguishing habitats. Including other measures that summarize the annual variation of  
476 vegetation indices could be a possible improvement.

477 The most relevant seasons for distinguishing vegetation types in Karst eco-mosaic were summer, autumn and winter, as was  
478 found by comparing the most important variables for the seasonal classifications. Spring, however, appeared among the most  
479 important variables in some classifications, and the month of May was important in multiple monthly classifications. This  
480 suggests that there is not one single period better than the others, and that multiple seasons should be considered. The advantage  
481 of using multi-temporal data for habitat classification has been proven in many cases, because data from different seasons  
482 provide different information (e.g. Feilhauer et al., 2013; Rapinel et al., 2019; Schuster et al., 2015). For example, Soubry and  
483 Guo (2021) found that the best season to distinguish shrubs in grasslands changed according to the spectral bands considered.  
484 In spring the most important bands were the red and blue bands, because the peak in growth was reached by shrubs but not by  
485 grasses; in summer a good separation was achieved only in the NIR region, due to the differences in leaf structures typical of  
486 woody and herbaceous plants; in autumn the most important bands were the SWIR and red, related to greenness and moisture.  
487 In the case of Karst eco-mosaic, autumn and winter generally allowed to distinguish evergreens from deciduous plants, while  
488 in summer there was the greatest separation among the different deciduous habitats especially with the NDVI and IRECI  
489 indices.

490

## 491 **Conclusion**

492 In this study, novel spectral heterogeneity indices were tested in a multi-temporal classification framework, and their potential  
493 to improve habitat mapping in complex landscapes was demonstrated, using the Classical Karst as testing area.

494 The aim of the study was generally achieved, but several improvements could be made. For example, different remote sensing  
495 data sources could be used, including hyperspectral sensors, active sensors such as LiDAR and radar, or sensors with very high  
496 spatial resolutions (Nagendra et al., 2013). In this way, the estimation of spectral heterogeneity could be improved. Moreover,  
497 input variables have been combined in a limited number of ways in this study and testing other configurations can possibly  
498 produce better results.

499 The framework presented here was applied to some areas of Classical Karst, but could be extended to test its validity on a  
500 larger scale. This approach based on remote sensing cannot replace field work and requires field data for training and validation,  
501 though it can be a valid tool to map habitats in a cost- and time-effective way that is very well suitable for monitoring purposes.  
502

503 **References**

- 504 Asner, G.P., Scurlock, J.M.O., A. Hicke, J., 2003. Global synthesis of leaf area index observations: implications for ecological  
505 and remote sensing studies: *Global leaf area index*. *Glob. Ecol. Biogeogr.* 12, 191–205.  
506 <https://doi.org/10.1046/j.1466-822X.2003.00026.x>
- 507 Baldeck, C., Asner, G., 2013. Estimating Vegetation Beta Diversity from Airborne Imaging Spectroscopy and Unsupervised  
508 Clustering. *Remote Sens.* 5, 2057–2071. <https://doi.org/10.3390/rs5052057>
- 509 Bayle, A., Carlson, B., Thierion, V., Isenmann, M., Choler, P., 2019. Improved Mapping of Mountain Shrublands Using the  
510 Sentinel-2 Red-Edge Band. *Remote Sens.* 11, 2807. <https://doi.org/10.3390/rs11232807>
- 511 Bhatt, P., Maclean, A., Dickinson, Y., Kumar, C., 2022. Fine-Scale Mapping of Natural Ecological Communities Using  
512 Machine Learning Approaches. *Remote Sens.* 14, 563. <https://doi.org/10.3390/rs14030563>
- 513 Borra, S., Thanki, R., Dey, N., 2019. *Satellite Image Analysis: Clustering and Classification*, SpringerBriefs in Applied  
514 Sciences and Technology. Springer Singapore, Singapore. <https://doi.org/10.1007/978-981-13-6424-2>
- 515 Bray, J.R., Curtis, J.T., 1957. An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecol. Monogr.* 27,  
516 325–349. <https://doi.org/10.2307/1942268>
- 517 Breiman, L., 2001. Random Forests. *Mach. Learn.* 45, 5–32. <https://doi.org/10.1023/A:1010933404324>
- 518 Chen, D., Huang, J., Jackson, T.J., 2005. Vegetation water content estimation for corn and soybeans using spectral indices  
519 derived from MODIS near- and short-wave infrared bands. *Remote Sens. Environ.* 98, 225–236.  
520 <https://doi.org/10.1016/j.rse.2005.07.008>
- 521 Chraïbi, E., Arnold, H., Luque, S., Deacon, A., Magurran, A., Féret, J.-B., 2021. A Remote Sensing Approach to Understanding  
522 Patterns of Secondary Succession in Tropical Forest. *Remote Sens.* 13, 2148. <https://doi.org/10.3390/rs13112148>
- 523 Corbane, C., Lang, S., Pipkins, K., Alleaume, S., Deshayes, M., García Millán, V.E., Strasser, T., Vanden Borre, J., Toon, S.,  
524 Michael, F., 2015. Remote sensing for mapping natural habitats and their conservation status – New opportunities  
525 and challenges. *Int. J. Appl. Earth Obs. Geoinformation*, Special Issue on Earth observation for habitat mapping and  
526 biodiversity monitoring 37, 7–16. <https://doi.org/10.1016/j.jag.2014.11.005>
- 527 da Silveira, H.L.F., Galvão, L.S., Sanches, I.D., de Sá, I.B., Taura, T.A., 2018. Use of MSI/Sentinel-2 and airborne LiDAR  
528 data for mapping vegetation and studying the relationships with soil attributes in the Brazilian semi-arid region. *Int.*  
529 *J. Appl. Earth Obs. Geoinformation* 73, 179–190. <https://doi.org/10.1016/j.jag.2018.06.016>
- 530 European Commission, 2005. Note to the Habitats Committee. Assessment, monitoring and reporting of conservation status –  
531 Preparing the 2001–2007 report under Article 17 of the Habitats Directive (No. DocHab-04-03/03 rev.3.). European  
532 Commission, Brussels.
- 533 European Commission, 1992. Council Directive 92/43/EEC of 21 May 1992 on the conservation of natural habitats and of wild  
534 fauna and flora (OJ L 206 22.07.1992 p. 7).

535 Fassnacht, F.E., Müllerová, J., Conti, L., Malavasi, M., Schmidlein, S., 2022. About the link between biodiversity and spectral  
536 variation. *Appl. Veg. Sci.* 25. <https://doi.org/10.1111/avsc.12643>

537 Fauvel, M., Lopes, M., Dubo, T., Rivers-Moore, J., Frison, P.-L., Gross, N., Ouin, A., 2020. Prediction of plant diversity in  
538 grasslands using Sentinel-1 and -2 satellite image time series. *Remote Sens. Environ.* 237, 111536.  
539 <https://doi.org/10.1016/j.rse.2019.111536>

540 Feilhauer, H., Somers, B., van der Linden, S., 2017. Optical trait indicators for remote sensing of plant species composition:  
541 Predictive power and seasonal variability. *Ecol. Indic.* 73, 825–833. <https://doi.org/10.1016/j.ecolind.2016.11.003>

542 Feilhauer, H., Thonfeld, F., Faude, U., He, K.S., Rocchini, D., Schmidlein, S., 2013. Assessing floristic composition with  
543 multispectral sensors—A comparison based on monotemporal and multiseasonal field spectra. *Int. J. Appl. Earth  
544 Obs. Geoinformation* 21, 218–229. <https://doi.org/10.1016/j.jag.2012.09.002>

545 Féret, J., Boissieu, F., 2020. *biodivMapR*: An R package for  $\alpha$ - and  $\beta$ -diversity mapping using remotely sensed images. *Methods  
546 Ecol. Evol.* 11, 64–70. <https://doi.org/10.1111/2041-210X.13310>

547 Féret, J.-B., Asner, G.P., 2014. Mapping tropical forest canopy diversity using high-fidelity imaging spectroscopy. *Ecol. Appl.*  
548 24, 1289–1296. <https://doi.org/10.1890/13-1824.1>

549 Foody, G.M., 2004. Thematic Map Comparison: Evaluating the statistical significance of differences in classification accuracy.  
550 *Photogramm. Eng. Remote Sens.* 70, 627–633. <https://doi.org/10.14358/PERS.70.5.627>

551 Frampton, W.J., Dash, J., Watmough, G., Milton, E.J., 2013. Evaluating the capabilities of Sentinel-2 for quantitative  
552 estimation of biophysical variables in vegetation. *ISPRS J. Photogramm. Remote Sens.* 82, 83–92.  
553 <https://doi.org/10.1016/j.isprsjprs.2013.04.007>

554 Gamon, J.A., Field, C.B., Goulden, M.L., Griffin, K.L., Hartley, A.E., Joel, G., Penuelas, J., Valentini, R., 1995. Relationships  
555 Between NDVI, Canopy Structure, and Photosynthesis in Three Californian Vegetation Types. *Ecol. Appl.* 5, 28–41.  
556 <https://doi.org/10.2307/1942049>

557 Gara, T.W., Darvishzadeh, R., Skidmore, A.K., Wang, T., Heurich, M., 2019. Accurate modelling of canopy traits from  
558 seasonal Sentinel-2 imagery based on the vertical distribution of leaf traits. *ISPRS J. Photogramm. Remote Sens.*  
559 157, 108–123. <https://doi.org/10.1016/j.isprsjprs.2019.09.005>

560 Gillespie, T.W., 2005. Predicting woody-plant species richness in tropical dry forests: a case study from South Florida, USA.  
561 *Ecol. Appl.* 15, 27–37. <https://doi.org/10.1890/03-5304>

562 Gitelson, A.A., Kaufman, Y.J., Merzlyak, M.N., 1996. Use of a green channel in remote sensing of global vegetation from  
563 EOS-MODIS. *Remote Sens. Environ.* 58, 289–298. [https://doi.org/10.1016/S0034-4257\(96\)00072-7](https://doi.org/10.1016/S0034-4257(96)00072-7)

564 Gregorutti, B., Michel, B., Saint-Pierre, P., 2017. Correlation and variable importance in random forests. *Stat. Comput.* 27,  
565 659–678. <https://doi.org/10.1007/s11222-016-9646-1>

566 Guyon, I., Weston, J., Barnhill, S., Vapnik, V., 2002. Gene selection for cancer classification using support vector machines.  
567 *Mach. Learn.* 46, 389–422. <https://doi.org/10.1023/A:1012487302797>

568 Haralick, R.M., Shanmugam, K., Dinstein, I., 1973. Textural Features for Image Classification. *IEEE Trans. Syst. Man Cybern.*  
569 *SMC-3*, 610–621. <https://doi.org/10.1109/TSMC.1973.4309314>

570 He, K.S., Rocchini, D., Neteler, M., Nagendra, H., 2011. Benefits of hyperspectral remote sensing for tracking plant invasions:  
571 Plant invasion and hyperspectral remote sensing. *Divers. Distrib.* 17, 381–392. [https://doi.org/10.1111/j.1472-](https://doi.org/10.1111/j.1472-4642.2011.00761.x)  
572 [4642.2011.00761.x](https://doi.org/10.1111/j.1472-4642.2011.00761.x)

573 Hoffmann, S., Schmitt, T.M., Chiarucci, A., Irl, S.D.H., Rocchini, D., Vetaas, O.R., Tanase, M.A., Mermoz, S., Bouvet, A.,  
574 Beierkuhnlein, C., 2019. Remote sensing of  $\beta$ -diversity: Evidence from plant communities in a semi-natural system.  
575 *Appl. Veg. Sci.* 22, 13–26. <https://doi.org/10.1111/avsc.12403>

576 Khatami, R., Mountrakis, G., Stehman, S.V., 2016. A meta-analysis of remote sensing research on supervised pixel-based land-  
577 cover image classification processes: General guidelines for practitioners and future research. *Remote Sens. Environ.*  
578 177, 89–100. <https://doi.org/10.1016/j.rse.2016.02.028>

579 Kollert, A., Bremer, M., Löw, M., Rutzinger, M., 2021. Exploring the potential of land surface phenology and seasonal cloud  
580 free composites of one year of Sentinel-2 imagery for tree species mapping in a mountainous region. *Int. J. Appl.*  
581 *Earth Obs. Geoinformation* 94, 102208. <https://doi.org/10.1016/j.jag.2020.102208>

582 Kuhn, M., 2021. *caret: Classification and Regression Training.*

583 Lane, C., Liu, H., Autrey, B., Anenkhonov, O., Chepinoga, V., Wu, Q., 2014. Improved Wetland Classification Using Eight-  
584 Band High Resolution Satellite Imagery and a Hybrid Approach. *Remote Sens.* 6, 12187–12216.  
585 <https://doi.org/10.3390/rs61212187>

586 Levin, N., Shmida, A., Levanoni, O., Tamari, H., Kark, S., 2007. Predicting mountain plant richness and rarity from space  
587 using satellite-derived vegetation indices: Predicting mountain biodiversity from space. *Divers. Distrib.* 13, 692–703.  
588 <https://doi.org/10.1111/j.1472-4642.2007.00372.x>

589 Li, Y., He, N., Hou, J., Xu, L., Liu, C., Zhang, J., Wang, Q., Zhang, X., Wu, X., 2018. Factors Influencing Leaf Chlorophyll  
590 Content in Natural Forests at the Biome Scale. *Front. Ecol. Evol.* 6, 64. <https://doi.org/10.3389/fevo.2018.00064>

591 Liccari, F., Sigura, M., Bacaro, G., 2022. Use of Remote Sensing Techniques to Estimate Plant Diversity within Ecological  
592 Networks: A Worked Example. *Remote Sens.* 14, 4933. <https://doi.org/10.3390/rs14194933>

593 Marin, A., Altobelli, A., 2021. Social ecology and traditional landscape enhancement. Some issues from a case study in the  
594 Gorizia Karst. *SMC Mag.* 55–60.

595 Marzialetti, F., Di Febbraro, M., Malavasi, M., Giulio, S., Acosta, A.T.R., Carranza, M.L., 2020. Mapping Coastal Dune  
596 Landscape through Spectral Rao's Q Temporal Diversity. *Remote Sens.* 12, 2315.  
597 <https://doi.org/10.3390/rs12142315>

598 Maxwell, A.E., Warner, T.A., Fang, F., 2018. Implementation of machine-learning classification in remote sensing: an applied  
599 review. *Int. J. Remote Sens.* 39, 2784–2817. <https://doi.org/10.1080/01431161.2018.1433343>

600 Oldeland, J., Wesuls, D., Rocchini, D., Schmidt, M., Jürgens, N., 2010. Does using species abundance data improve estimates  
601 of species diversity from remotely sensed spectral heterogeneity? *Ecol. Indic.* 10, 390–396.  
602 <https://doi.org/10.1016/j.ecolind.2009.07.012>

603 Osińska-Skotak, K., Radecka, A., Ostrowski, W., Michalska-Hejduk, D., Charyton, J., Bakula, K., Piórkowski, H., 2021. The  
604 Methodology for Identifying Secondary Succession in Non-Forest Natura 2000 Habitats Using Multi-Source  
605 Airborne Remote Sensing Data. *Remote Sens.* 13, 2803. <https://doi.org/10.3390/rs13142803>

606 OSMER, 2015. Schede climatiche territoriali del Friuli Venezia Giulia [WWW Document]. ARPA FVG. URL  
607 [https://www.clima.fvg.it/clima\\_schede.php?m=1](https://www.clima.fvg.it/clima_schede.php?m=1)

608 Palmer, M.W., Earls, P.G., Hoagland, B.W., White, P.S., Wohlgemuth, T., 2002. Quantitative tools for perfecting species lists.  
609 *Environmetrics* 13, 121–137. <https://doi.org/10.1002/env.516>

610 Palmer, M.W., Wohlgemuth, T., Earls, P., Arévalo, J.R., Thompson, S.D., 2000. Opportunities for long-term ecological  
611 research at the Tallgrass Prairie Preserve, Oklahoma, in: Lajtha, K., Vanderbilt, K. (Eds.), *Cooperation in Long Term*  
612 *Ecological Research in Central and Eastern Europe: Proceedings of ILTER Regional Workshop, Budapest, Hungary,*  
613 *22–25 June, 1999*, Pp. 123–128.

614 Poldini, L., 2009. *La diversità vegetale del Carso fra Trieste e Gorizia: lo stato dell’ambiente.*, Ed. Goliardiche. ed.

615 Poldini, L., 1989. *La vegetazione del Carso isontino e triestino: studio del paesaggio vegetale fra Trieste, Gorizia e i territori*  
616 *adiacenti.*, Lint. ed.

617 Praticò, S., Solano, F., Di Fazio, S., Modica, G., 2021. Machine Learning Classification of Mediterranean Forest Habitats in  
618 Google Earth Engine Based on Seasonal Sentinel-2 Time-Series and Input Image Composition Optimisation. *Remote*  
619 *Sens.* 13, 586. <https://doi.org/10.3390/rs13040586>

620 R Core Team, 2022. *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna,  
621 Austria. URL <https://www.R-project.org/>.

622 Rao, C.R., 1982. Diversity and dissimilarity coefficients: A unified approach. *Theor. Popul. Biol.* 21, 24–43.  
623 [https://doi.org/10.1016/0040-5809\(82\)90004-1](https://doi.org/10.1016/0040-5809(82)90004-1)

624 Rapinel, S., Mony, C., Lecoq, L., Clément, B., Thomas, A., Hubert-Moy, L., 2019. Evaluation of Sentinel-2 time-series for  
625 mapping floodplain grassland plant communities. *Remote Sens. Environ.* 223, 115–129.  
626 <https://doi.org/10.1016/j.rse.2019.01.018>

627 Richards, J.A., 2013. *Remote Sensing Digital Image Analysis.* Springer Berlin Heidelberg, Berlin, Heidelberg.  
628 <https://doi.org/10.1007/978-3-642-30062-2>

629 Rocchini, D., Andreo, V., Förster, M., Garzon-Lopez, C.X., Gutierrez, A.P., Gillespie, T.W., Hauffe, H.C., He, K.S.,  
630 Kleinschmit, B., Mairota, P., Marcantonio, M., Metz, M., Nagendra, H., Pareeth, S., Ponti, L., Ricotta, C., Rizzoli,  
631 A., Schaab, G., Zebisch, M., Zorer, R., Neteler, M., 2015. Potential of remote sensing to predict species invasions:  
632 A modelling perspective. *Prog. Phys. Geogr. Earth Environ.* 39, 283–309.  
633 <https://doi.org/10.1177/0309133315574659>

634 Rocchini, D., Balkenhol, N., Carter, G.A., Foody, G.M., Gillespie, T.W., He, K.S., Kark, S., Levin, N., Lucas, K., Luoto, M.,  
635 Nagendra, H., Oldeland, J., Ricotta, C., Southworth, J., Neteler, M., 2010a. Remotely sensed spectral heterogeneity  
636 as a proxy of species diversity: Recent advances and open challenges. *Ecol. Inform.* 5, 318–329.  
637 <https://doi.org/10.1016/j.ecoinf.2010.06.001>

638 Rocchini, D., Delucchi, L., Bacaro, G., Cavallini, P., Feilhauer, H., Foody, G.M., He, K.S., Nagendra, H., Porta, C., Ricotta,  
639 C., Schmidtlein, S., Spano, L.D., Wegmann, M., Neteler, M., 2013a. Calculating landscape diversity with  
640 information-theory based indices: A GRASS GIS solution. *Ecol. Inform.* 17, 82–93.  
641 <https://doi.org/10.1016/j.ecoinf.2012.04.002>

642 Rocchini, D., Foody, G.M., Nagendra, H., Ricotta, C., Anand, M., He, K.S., Amici, V., Kleinschmit, B., Förster, M.,  
643 Schmidtlein, S., Feilhauer, H., Ghisla, A., Metz, M., Neteler, M., 2013b. Uncertainty in ecosystem mapping by  
644 remote sensing. *Comput. Geosci.* 50, 128–135. <https://doi.org/10.1016/j.cageo.2012.05.022>

645 Rocchini, D., He, K.S., Oldeland, J., Wesuls, D., Neteler, M., 2010b. Spectral variation versus species  $\beta$ -diversity at different  
646 spatial scales: a test in African highland savannas. *J. Environ. Monit.* 12, 825. <https://doi.org/10.1039/b921835a>

647 Rocchini, D., Marcantonio, M., Da Re, D., Bacaro, G., Feoli, E., Foody, G.M., Furrer, R., Harrigan, R.J., Kleijn, D., Iannacito,  
648 M., Lenoir, J., Lin, M., Malavasi, M., Marchetto, E., Meyer, R.S., Moudry, V., Schneider, F.D., Šímová, P., Thornhill,  
649 A.H., Thouverai, E., Vicario, S., Wayne, R.K., Ricotta, C., 2021a. From zero to infinity: Minimum to maximum  
650 diversity of the planet by spatio-parametric Rao's quadratic entropy. *Glob. Ecol. Biogeogr.* 30, 1153–1162.  
651 <https://doi.org/10.1111/geb.13270>

652 Rocchini, D., Marcantonio, M., Da Re, D., Bacaro, G., Feoli, E., Foody, G.M., Furrer, R., Harrigan, R.J., Kleijn, D., Iannacito,  
653 M., Lenoir, J., Lin, M., Malavasi, M., Marchetto, E., Meyer, R.S., Moudry, V., Schneider, F.D., Šímová, P., Thornhill,  
654 A.H., Thouverai, E., Vicario, S., Wayne, R.K., Ricotta, C., 2021b. From zero to infinity: Minimum to maximum  
655 diversity of the planet by spatio-parametric Rao's quadratic entropy. *Glob. Ecol. Biogeogr.* 30, 1153–1162.  
656 <https://doi.org/10.1111/geb.13270>

657 Rocchini, D., Salvatori, N., Beierkuhnlein, C., Chiarucci, A., de Boissieu, F., Förster, M., Garzon-Lopez, C.X., Gillespie, T.W.,  
658 Hauffe, H.C., He, K.S., Kleinschmit, B., Lenoir, J., Malavasi, M., Moudry, V., Nagendra, H., Payne, D., Šímová, P.,  
659 Torresani, M., Wegmann, M., Féret, J.-B., 2021c. From local spectral species to global spectral communities: A  
660 benchmark for ecosystem diversity estimate by remote sensing. *Ecol. Inform.* 61, 101195.  
661 <https://doi.org/10.1016/j.ecoinf.2020.101195>

662 Rossi, C., Kneubühler, M., Schütz, M., Schaepman, M.E., Haller, R.M., Risch, A.C., 2021. Spatial resolution, spectral metrics  
663 and biomass are key aspects in estimating plant species richness from spectral diversity in species-rich grasslands.  
664 *Remote Sens. Ecol. Conserv.* rse2.244. <https://doi.org/10.1002/rse2.244>

665 Rouse, J.W., Haas, R.H., Schell, J.A.W., 1975. Monitoring vegetation systems in the great plains with ERTS. Presented at the  
666 Third ERTS Symposium.

667 Schuster, C., Förster, M., Kleinschmit, B., 2012. Testing the red edge channel for improving land-use classifications based on  
668 high-resolution multi-spectral satellite data. *Int. J. Remote Sens.* 33, 5583–5599.  
669 <https://doi.org/10.1080/01431161.2012.666812>

670 Schuster, C., Schmidt, T., Conrad, C., Kleinschmit, B., Förster, M., 2015. Grassland habitat mapping by intra-annual time  
671 series analysis – Comparison of RapidEye and TerraSAR-X satellite data. *Int. J. Appl. Earth Obs. Geoinformation*  
672 34, 25–34. <https://doi.org/10.1016/j.jag.2014.06.004>

673 Senf, C., Leitão, P.J., Pflugmacher, D., van der Linden, S., Hostert, P., 2015. Mapping land cover in complex Mediterranean  
674 landscapes using Landsat: Improved classification accuracies from integrating multi-seasonal and synthetic imagery.  
675 *Remote Sens. Environ.* 156, 527–536. <https://doi.org/10.1016/j.rse.2014.10.018>

676 Shannon, C.E., 1948. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* 27, 379–423.  
677 <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>

678 Soubry, I., Guo, X., 2021. Identification of the Optimal Season and Spectral Regions for Shrub Cover Estimation in Grasslands.  
679 *Sensors* 21, 3098. <https://doi.org/10.3390/s21093098>

680 Stein, A., Gerstner, K., Kreft, H., 2014. Environmental heterogeneity as a universal driver of species richness across taxa,  
681 biomes and spatial scales. *Ecol. Lett.* 17, 866–880. <https://doi.org/10.1111/ele.12277>

682 Tarantino, C., Forte, L., Blonda, P., Vicario, S., Tomaselli, V., Beierkuhnlein, C., Adamo, M., 2021. Intra-Annual Sentinel-2  
683 Time-Series Supporting Grassland Habitat Discrimination. *Remote Sens.* 13, 277.  
684 <https://doi.org/10.3390/rs13020277>

685 Thouverai, E., Marcantonio, M., Bacaro, G., Re, D.D., Iannacito, M., Marchetto, E., Ricotta, C., Tattoni, C., Vicario, S.,  
686 Rocchini, D., 2021. Measuring diversity from space: a global view of the free and open source rasterdiv R package  
687 under a coding perspective. *Community Ecol.* 22, 1–11. <https://doi.org/10.1007/s42974-021-00042-x>

688 Villoslada, M., Bergamo, T.F., Ward, R.D., Burnside, N.G., Joyce, C.B., Bunce, R.G.H., Sepp, K., 2020. Fine scale plant  
689 community assessment in coastal meadows using UAV based multispectral data. *Ecol. Indic.* 111, 105979.  
690 <https://doi.org/10.1016/j.ecolind.2019.105979>

691 Wang, R., Gamon, J., Emmerton, C., Li, H., Nestola, E., Pastorello, G., Menzer, O., 2016. Integrated Analysis of Productivity  
692 and Biodiversity in a Southern Alberta Prairie. *Remote Sens.* 8, 214. <https://doi.org/10.3390/rs8030214>

693 Wang, R., Gamon, J.A., 2019. Remote sensing of terrestrial plant biodiversity. *Remote Sens. Environ.* 231, 111218.  
694 <https://doi.org/10.1016/j.rse.2019.111218>

695 Wang, R., Gamon, J.A., Cavender-Bares, J., Townsend, P.A., Zyguelbaum, A.I., 2018. The spatial sensitivity of the spectral  
696 diversity–biodiversity relationship: an experimental test in a prairie grassland. *Ecol. Appl.* 28, 541–556.  
697 <https://doi.org/10.1002/eap.1669>

698 Whittaker, R.H., 1960. Vegetation of the Siskiyou Mountains, Oregon and California. *Ecol. Monogr.* 30, 279–338.  
699 <https://doi.org/10.2307/1943563>



700     Wulder, M.A., Coops, N.C., Roy, D.P., White, J.C., Hermosilla, T., 2018. Land cover 2.0. *Int. J. Remote Sens.* 39, 4254–4284.

701             <https://doi.org/10.1080/01431161.2018.1452075>

702

703 **Table Captions**

704 **Table 1:** List of vegetation indices used for the analysis.

705 **Table 2:** Input variables configurations used for image classification.

706 **Table 3:** Overall accuracy (OA) and Kappa values obtained from the different classification pathways.

707 **Table 4:** Confusion matrix for the best classification (seasonal classification performed with 11 classes). The rows represent  
708 the results obtained from the classification, while the columns represent the reference data. The values on the matrix diagonal  
709 are the correctly classified pixels. Habitats are abbreviated as follows: Shrubland (Shr), Downy oak woodland (DOW), Sessile  
710 oak woodland (SOW), Invasive alien species (IAS), Pure grassland (Gr\_I0), Grassland at successional stage 1 (Gr\_I1),  
711 Grassland at successional stage 2 (Gr\_I2), Grassland-woodland mosaic, Black pine plantation, Hay meadow, Pasture-grassland.

712 **Table 5:** Class-specific accuracy parameters (UA: user's accuracy, PA: producer's accuracy) obtained for the seasonal  
713 classification performed with 11 classes. Accuracy was assessed using independent validation data.

714

715 **Tables**

716 **Table 1**

<b>Index</b>	<b>Formula</b>	<b>Reference</b>
<b>NDVI</b>	$(\text{NIR}_{(B8)} - \text{Red}_{(B4)}) / (\text{NIR}_{(B8)} + \text{Red}_{(B4)})$	Rouse et al., 1975
<b>GNDVI</b>	$(\text{NIR}_{(B8)} - \text{Green}_{(B3)}) / (\text{NIR}_{(B8)} + \text{Green}_{(B3)})$	Gitelson et al., 1996
<b>NDWI</b>	$(\text{NIR}_{(B8)} - \text{SWIR}_{(B11)}) / (\text{NIR}_{(B8)} + \text{SWIR}_{(B11)})$	Chen et al., 2005
<b>IRECI</b>	$((\text{RedEdge}_{(B7)} - \text{Red}_{(B4)}) / ((\text{RedEdge}_{(B5)} / \text{RedEdge}_{(B6)})) \times 10000$	Frampton et al., 2013

717

718

719 **Table 1**

<b>Temporal configuration</b>	<b>Input variables</b>	<b>Number of input layers</b>
<b>Monthly</b>	Vegetation indices: 4 layers per month (NDVI, GNDVI, NDWI, IRECI) Rao's Q: 4 layers per month (NDVI, GNDVI, NDWI, IRECI) $\alpha$ -diversity: 1 layer per month $\beta$ -diversity (first 3 PCoA axes): 3 layers per month	144
<b>Seasonal</b>	Vegetation indices: 4 layers per season (NDVI, GNDVI, NDWI, IRECI) Rao's Q: 4 layers per season (NDVI, GNDVI, NDWI, IRECI) $\alpha$ -diversity: 1 layer per season $\beta$ -diversity (first 3 PCoA axes): 3 layers per season	48
<b>Multi-temporal monthly</b>	Multi-temporal Rao's Q: 4 layers per year (NDVI, GNDVI, NDWI, IRECI) Multi-temporal $\alpha$ -diversity: 4 layers per year (NDVI, GNDVI, NDWI, IRECI) Multi-temporal $\beta$ -diversity (first 3 PCoA axes): 3x4 layers per year	20
<b>Multi-temporal seasonal</b>	Multi-temporal Rao's Q: 4 layers per year (NDVI, GNDVI, NDWI, IRECI) Multi-temporal $\alpha$ -diversity: 4 layers per year (NDVI, GNDVI, NDWI, IRECI) Multi-temporal $\beta$ -diversity (first 3 PCoA axes): 3x4 layers per year	20

720

721

722 **Table 3**

<b>N° of classes</b>	<b>Input configuration</b>	<b>N° of predictors</b>	<b>OA</b>	<b>Kappa <sup>723</sup></b>
<b>26</b>	Monthly	144	0.65	0.58
	+ RFE	48	0.63	0.56
	Seasonal	48	0.63	0.56
	+ RFE	46	0.62	0.54
	Multi-temporal monthly	20	0.62	0.54
	+ RFE	20	0.61	0.53
	Multi-temporal seasonal	20	0.57	0.50
	+ RFE	20	0.58	0.51
<b>11</b>	Monthly	144	0.73	0.65
	+ RFE	100	0.73	0.65
	only vegetation indices	48	0.69	0.59
	Seasonal	48	0.72	0.64
	+ RFE	34	0.72	0.64
	only vegetation indices	16	0.65	0.56
	Multi-temporal monthly	20	0.66	0.57
	+ RFE	14	0.67	0.57
	Multi-temporal seasonal	20	0.64	0.55
	+ RFE	17	0.64	0.55

724

725

726 **Table 4**

	<b>Shr</b>	<b>DOW</b>	<b>SOW</b>	<b>IAS</b>	<b>Gr_I0</b>	<b>Gr_I1</b>	<b>Gr_I2</b>	<b>GWM</b>	<b>BPP</b>	<b>HM</b>	<b>PG</b>
<b>Shr</b>	8	6		2		2	20		1		
<b>DOW</b>	2	156	10	1			1	1	9	1	
<b>SOW</b>		7	4								
<b>IAS</b>		4		0			1		1		
<b>Gr_I0</b>		2		1	12						
<b>Gr_I1</b>	1					21	5				
<b>Gr_I2</b>	1	5	1		3	7	51	3		2	1
<b>GWM</b>	1	22			2	1	2	13	4	1	
<b>BPP</b>	1	7						1	107		
<b>HM</b>		1								10	6
<b>PG</b>											1

727

728

729 **Table 5**

<b>Class</b>	<b>UA</b>	<b>PA</b>
<b>Shrubland</b>	0.21	0.57
<b>Downy oak woodland</b>	0.86	0.74
<b>Sessile oak woodland</b>	0.36	0.27
<b>Invasive species</b>	0.00	0.00
<b>Grassland I0</b>	0.80	0.71
<b>Grassland I1</b>	0.78	0.68
<b>Grassland I2</b>	0.69	0.64
<b>Grassland-woodland mosaic</b>	0.28	0.72
<b>Black pine plantation</b>	0.92	0.88
<b>Hay meadow</b>	0.59	0.71
<b>Pasture-grassland</b>	1.00	0.13

730

731

732 **Figure Captions**

733 **Figure 1:** Location of the study area, represented on the Sentinel-2 median composite of summer 2021.

734 **Figure 2:** Workflow synthesizing the approach used to map natural habitats through a Random Forest classification and  
735 multiple combinations of input layers (vegetation and spectral heterogeneity indices).

736 **Figure 3:** Comparison of the overall accuracy achieved by considering different numbers of habitat classes (a), by performing  
737 or not a variable selection step through RFE (b), and by using different input variables configurations (c).

738 **Figure 4:** Habitat map resulting from the RF classification based on seasonal layers of vegetation and spectral heterogeneity  
739 indices. Among all the possible classifications, the selected is the one that resulted in the highest accuracy while minimising  
740 the amount of input layers. A total of 11 habitat classes was considered, based on structural-physiognomic and ecological  
741 characteristics. The areas are located in Monfalcone (a), Case Coisce (b), Opicina (c), Aurisina (d), San Lorenzo (e), San  
742 Giuseppe (f) and Bagnoli (g).

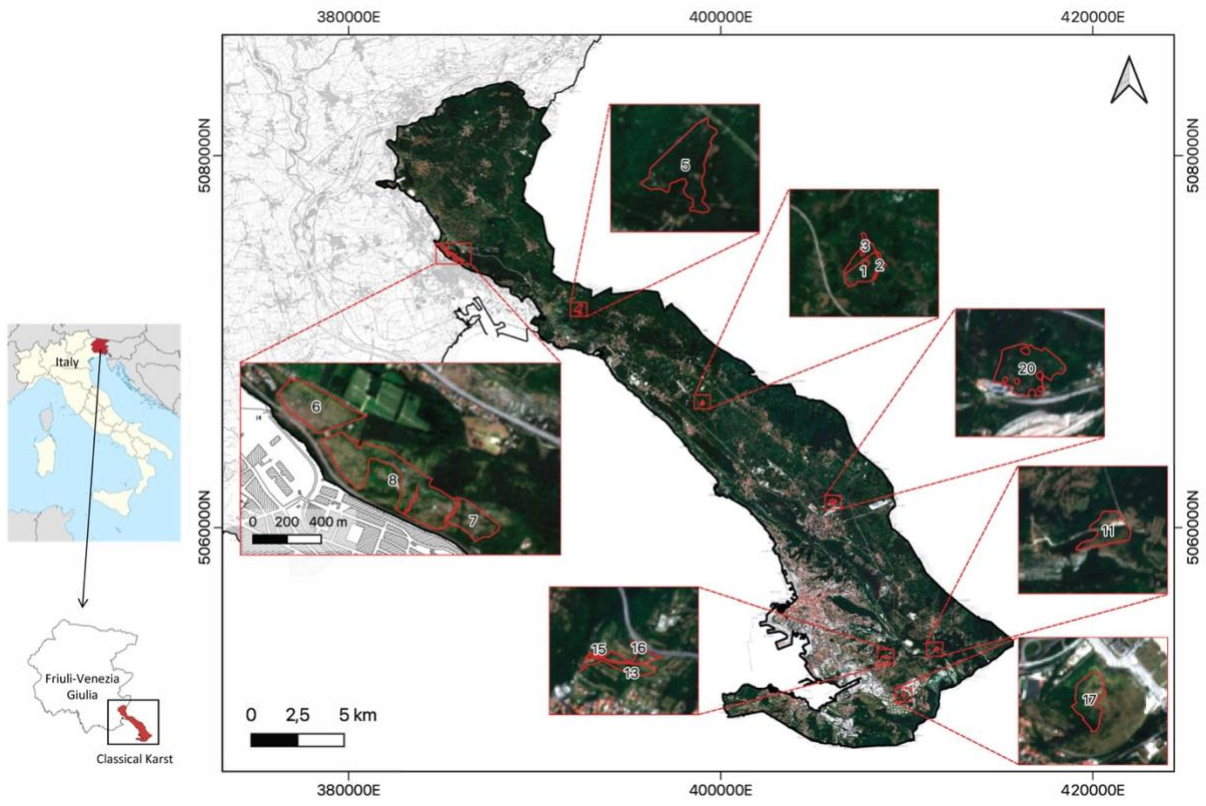
743 **Figure 5:** Relative importance of the variables used as input for the seasonal classifications with 11 classes. Classifications  
744 were performed with the whole set of input variables (a) and with a subset obtained by RFE (b). Only the first 20 variables are  
745 shown.

746

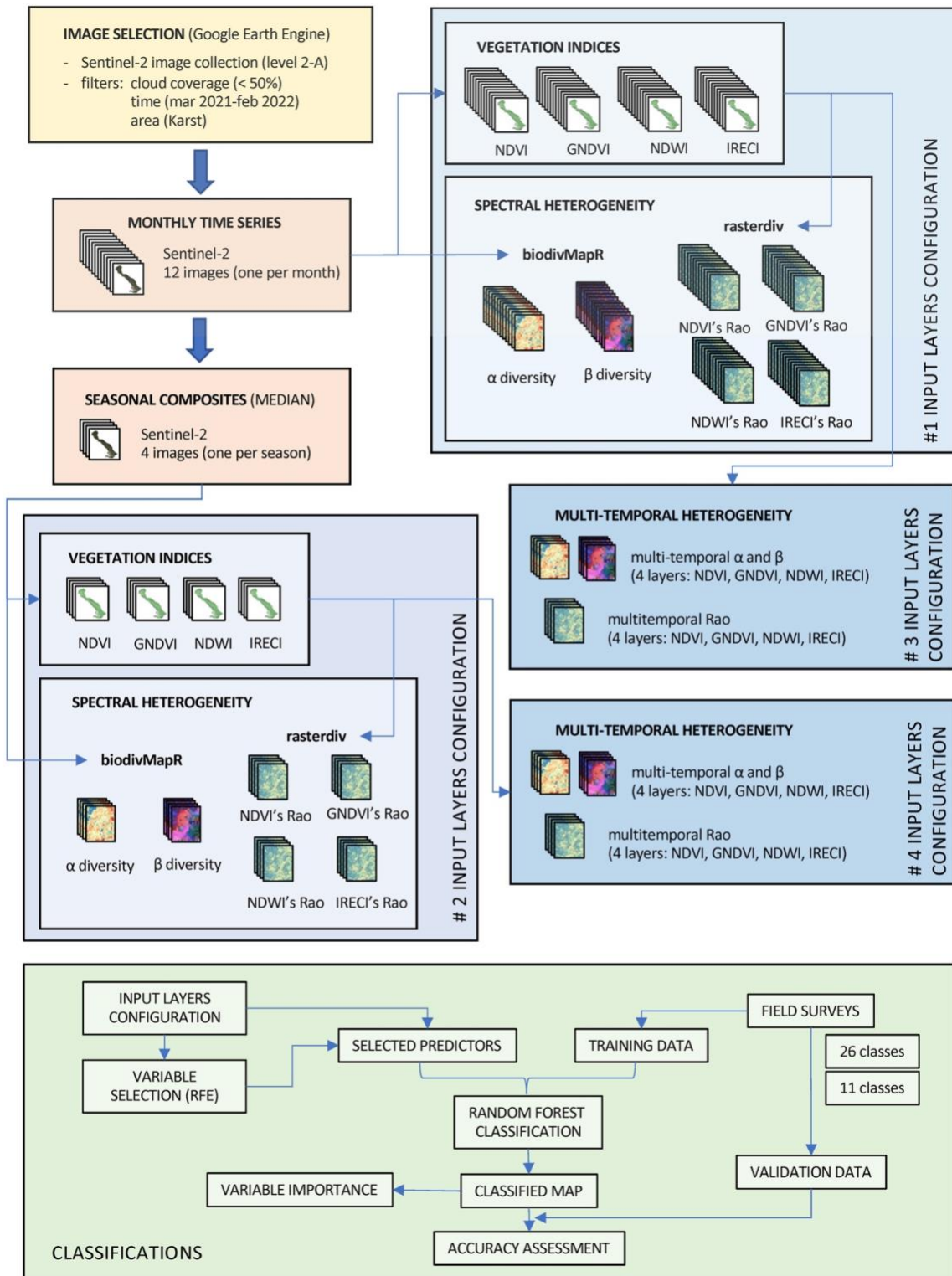


747 **Figures**

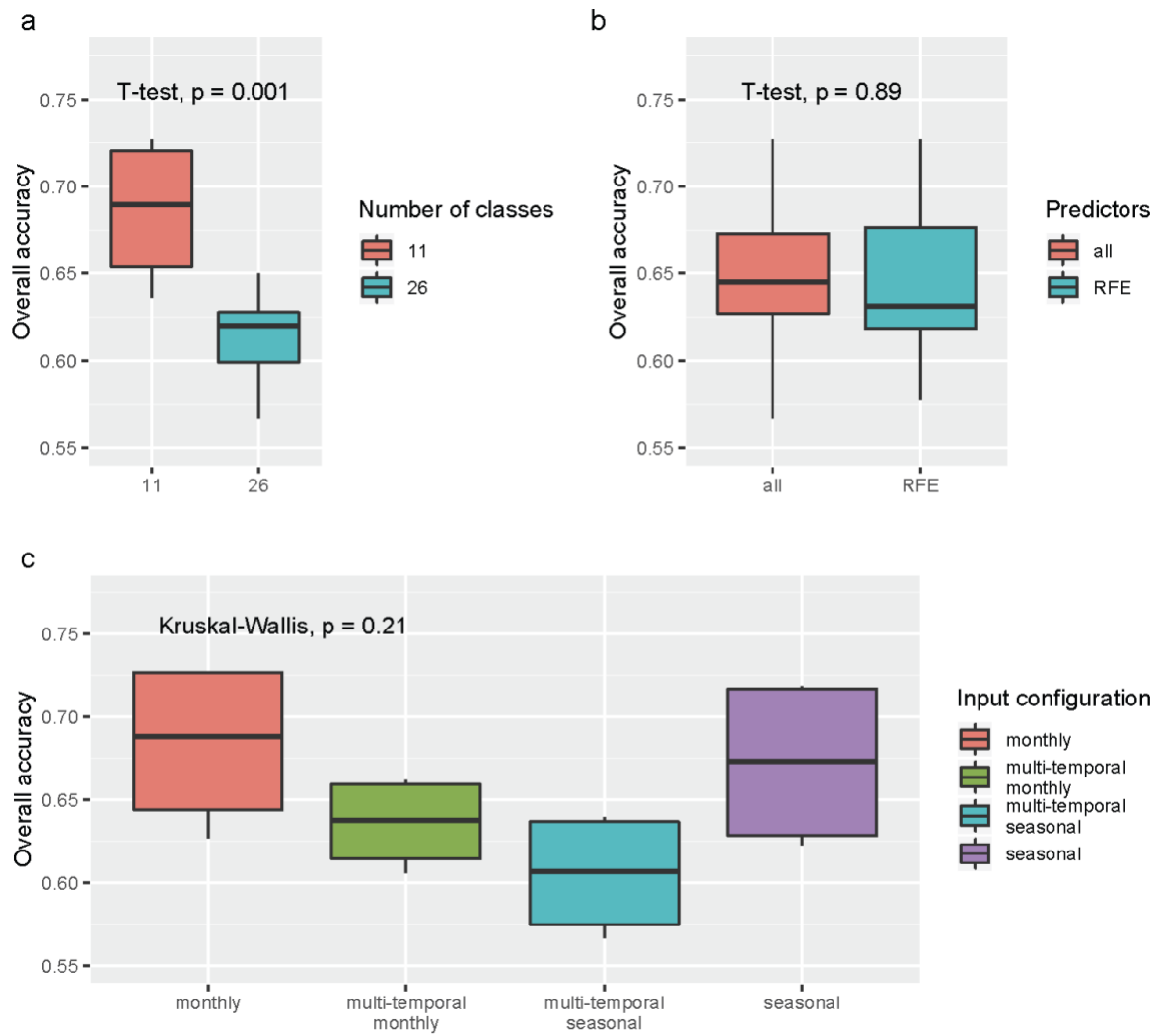
748 **Figure 1**



749

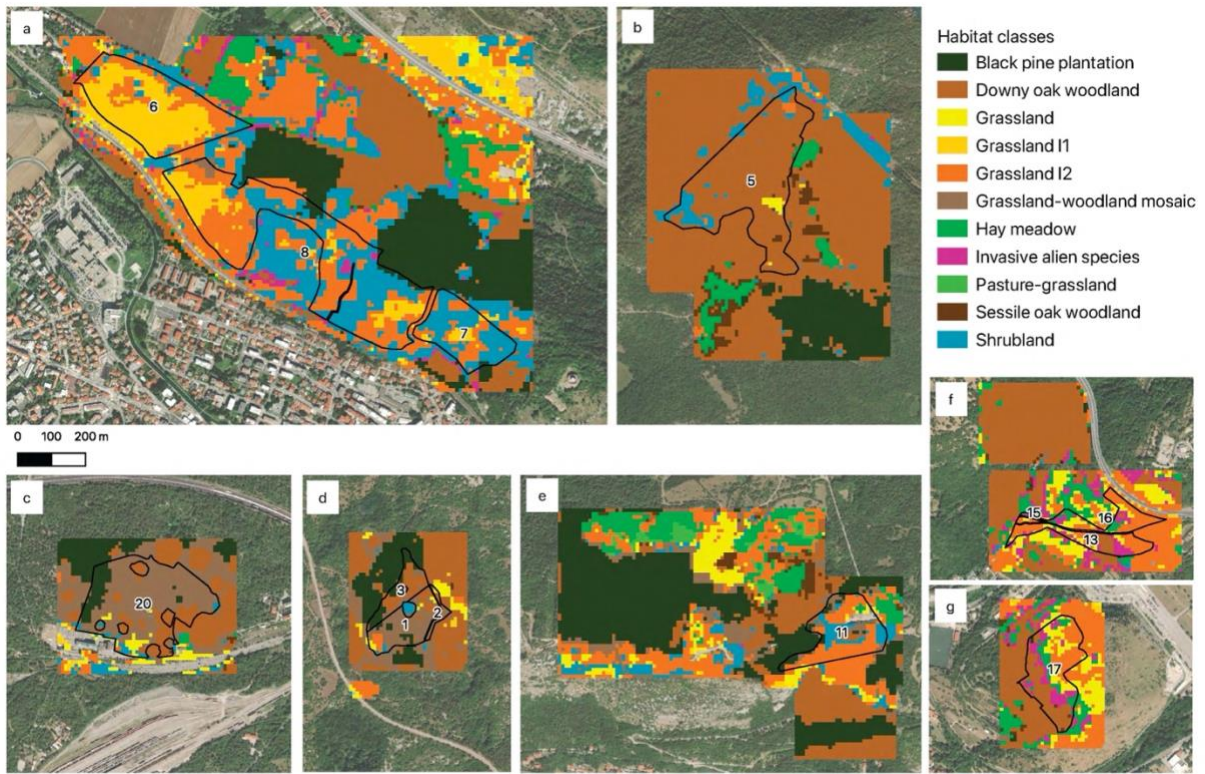


753 **Figure 3**

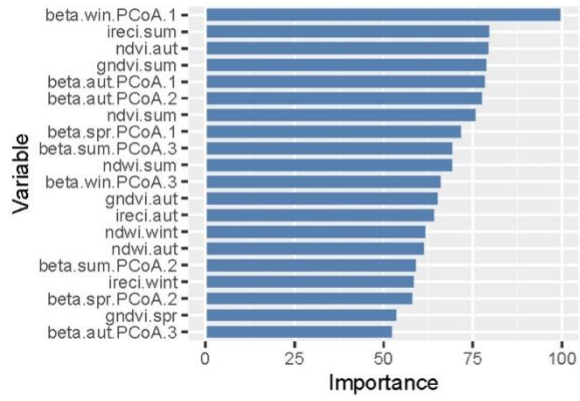


754

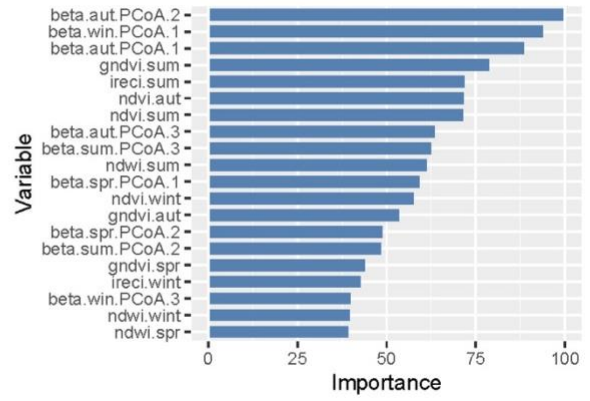
755



a



b



760

761