remote sensing

MDPI

Article

Evaluation and Selection of Multi-Spectral Indices to Classify Vegetation Using Multivariate Functional Principal Component Analysis

Simone Pesaresi [1,*], Adriano Mancini [2], Giacomo Quattrini [1] and Simona Casavecchia [1]

[1] Department of Agricultural, Food, and Environmental Sciences, D3A, Università Politecnica delle Marche, Via Brecce Bianche 12, 60131 Ancona, Italy; g.quattrini@pm.univpm.it (G.Q.); s.casavecchia@univpm.it (S.C.)
[2] Department of Information Engineering, DII, Università Politecnica delle Marche, Via Brecce Bianche 12, 60131 Ancona, Italy; a.mancini@univpm.it
* Correspondence: s.pesaresi@univpm.it

Abstract: The identification, classification and mapping of different plant communities and habitats is of fundamental importance for defining biodiversity monitoring and conservation strategies. Today, the availability of high temporal, spatial and spectral data from remote sensing platforms provides dense time series over different spectral bands. In the case of supervised mapping, time series based on classical vegetation indices (e.g., NDVI, GNDVI, ...) are usually input characteristics, but the selection of the best index or set of indices (which guarantees the best performance) is still based on human experience and is also influenced by the study area. In this work, several different time series, based on Sentinel-2 images, were created exploring new combinations of bands that extend the classic basic formulas as the normalized difference index. Multivariate Functional Principal Component Analysis (MFPCA) was used to contemporarily decompose the multiple time series. The principal multivariate seasonal spectral variations identified (MFPCA scores) were classified by using a Random Forest (RF) model. The MFPCA and RF classifications were nested into a forward selection strategy to identify the proper and minimum set of indices' (dense) time series that produced the most accurate supervised classification of plant communities and habitat. The results we obtained can be summarized as follows: (i) the selection of the best set of time series is specific to the study area and the habitats involved; (ii) well-known and widely used indices such as the NDVI are not selected as the indices with the best performance; instead, time series based on original indices (in terms of formula or combination of bands) or underused indices (such as those derivable with the visible bands) are selected; (iii) MFPCA efficiently reduces the dimensionality of the data (multiple dense time series) providing ecologically interpretable results representing an important tool for habitat modelling outperforming conventional approaches that consider only discrete time series.

Keywords: sentinel-2; time-series; functional data analysis; multivariate functional principal component analysis; habitat mapping; supervised classification; remote sensing; land surface phenology

Citation: Pesaresi, S.; Mancini, A.; Quattrini, G.; Casavecchia, S. Evaluation and Selection of Multi-Spectral Indices to Classify Vegetation Using Multivariate Functional Principal Component Analysis. *Remote Sens.* **2024**, *16*, 1224. https://doi.org/10.3390/rs16071224

Academic Editors: No-Wook Park, Kiwon Lee and Kwangseob Kim

Received: 20 February 2024
Revised: 27 March 2024
Accepted: 28 March 2024
Published: 30 March 2024

1. Introduction

Classifying and mapping plant communities and habitats are crucial for biodiversity monitoring and defining conservation strategies for Natura 2000 sites in Europe [1,2]. Currently, vegetation mapping benefits from the growing availability of high-quality data from remote sensing platforms such as Landsat, MODIS and Sentinel [3,4]. These platforms offer multi-temporal and multi-spectral time series data enabling the capture of seasonal variations in spectral reflectance related to the different phenological stages of vegetation (i.e., vegetation seasonality). These kinds of data are essential for an accurate supervised classification and mapping of plant communities and habitats [5–13]. Many studies have demonstrated the potential of direct machine learning applications for raw satellite multi-temporal data [14–18]. These models, which we can define as 'Pure Machine Learning'

according to Durell et al. [19], usually use time series of individual spectral bands or classic vegetation indices, such as the popular NDVI [20], consisting of a limited number of scenes within a single year. However, these models rely on human experience and prior knowledge of the best data acquisition time points and the most suitable set of indices to capture habitats during their optimal phenological stages. Therefore, these models face challenges in terms of transferability [21]. It is clear that recommending universal optimal time points and indices for all habitats across diverse study areas with varying vegetation and ecological characteristics is not feasible, despite the availability of indices tailored for specific applications [22,23]. In this context it is necessary to develop adaptable and transferable models that can autonomously select suitable indices and determine the ideal times for data acquisition based on the specific vegetation and ecological characteristics of a study area. A carefully selected set of area-specific indices offers significant advantages for land management organisations in compliance with national and international guidelines, such as the Habitats Directive [1,24,25]. These models should handle dense time series of remotely sensed data. Such data, which, in a specific time window, provide a richer wealth of information than multi-temporal data, are optimal for analysing seasonal changes in vegetation and improving classification accuracy [26,27].

Recently, promising methods known as 'Hybrid statistical-functional Machine Learning' [19], which combine machine learning with Functional Data Analysis (FDA) [28], have been employed to classify and map vegetation and habitats in two Natura 2000 sites [29,30]. Exploring such hybrid models is essential because they are capable of efficiently analysing dense time series of remote sensing data. The results are not only accurate but also facilitate interpretations and provide support to phytosociologists and ecologists in understanding the temporal spectral behaviour of plant associations (plant communities) [31–33]. The efficiency of analysing dense time series by FDA lies in its fundamental philosophy, which considers observed data functions as single entities, rather than merely as a sequence of individual observations [34]. In practice, if the entire time series of a pixel is expressed as a time function and considered as a single statistical unit, then a stack of remotely sensed images (a cube with x, y and t axes) is considered as a single temporal archive [35], essentially composed of as many functions as there are pixels in the area under test. The pixel-based functions (times series) of remotely sensed data can be thought of as points (or pixels) within a functional space [34]. The functional space can be univariate or multivariate, depending on the number of metrics (band or indices) used to describe and track the spectral variations within it (Figure 1).

Functional Principal Component Analysis (FPCA) is one of the most popular techniques in FDA for reducing the amount of functional data [36,37]. FPCA adapts traditional Principal Component Analysis (PCA) concepts to functions, allowing it to identify the main modes of variation among observations (functions) within a univariate functional space. It is evident that multivariate functional spaces are more natural and effective than univariate ones when describing spectral variations in vegetation (Figure 1). This is because seasonal patterns manifest differently across various spectral bands and vegetation indices, depending on the phenological stages of vegetation [26]. Multivariate Functional Principal Component Analysis (MFPCA) is well-suited for analysing multivariate functional spaces. MFPCA decomposes the multivariate functional space into a set of orthogonal multivariate functional principal components or modes of variation of functions (multivariate eigenfunctions), together with corresponding functional principal component scores (FPC scores). These FPC scores summarize the similarities between observations (functions), providing a compact representation of the data (one score value per multivariate principal component and per observation). In addition, these scores are uncorrelated by construction [38]. They can then serve as a building block for further statistical analyses such as unsupervised clustering, supervised classification methods or functional principal component regression with multiple covariates [39].

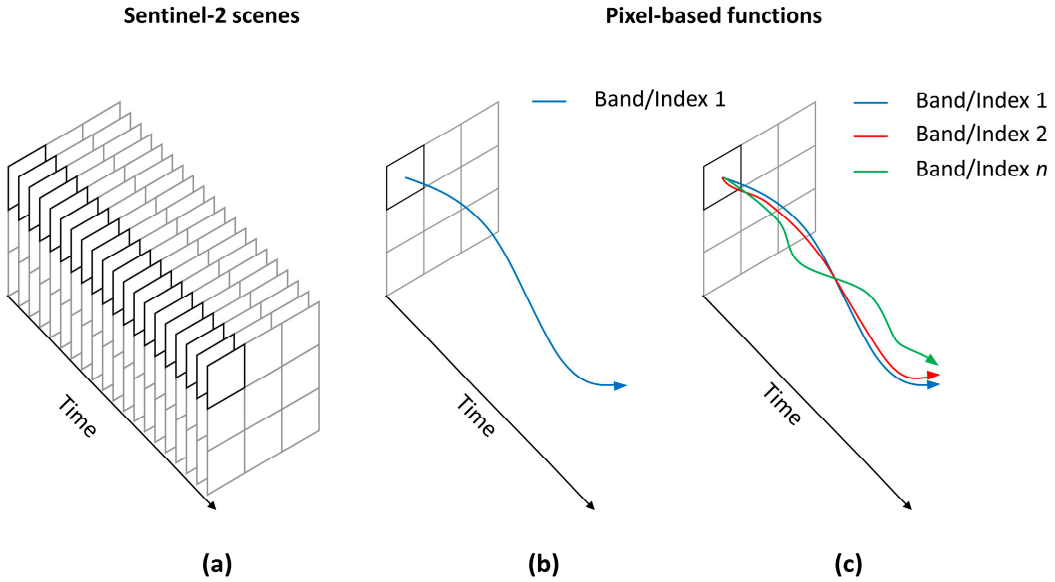

Figure 1. Spectral variations in remotely sensed images over time. (**a**) Finite discrete time series: this panel shows a typical representation of remotely sensed data captured at discrete points in time (raw data). Each point on the graph represents data from a specific moment. (**b**,**c**) Spectral variations in pixels as functions of time (smoothed representation of variations). These two panels show how individual pixel spectral characteristics evolve over time, simplifying trend observation. In detail (**b**) defines a univariate functional space that describe the spectral variations in pixels characterized by a single band or index, such as NDVI. This helps us to understand how one specific aspect of vegetation changes over time while (**c**) shows spectral variations in pixels characterized by multiple bands or indices, such as NDVI, GNDVI and NDWI, defining multivariate functional space (this allows us to study how different aspects of vegetation change together over time.

In this study, we develop new hybrid models that combine machine learning with MFPCA. MFPCA, the best of our knowledge, has not been previously used for supervised classification of habitats and vegetation. We believe that these models are valuable for analysing multivariate satellite dense time series, simultaneously considering seasonal spectral variations from different bands or vegetation indices, and for evaluating new vegetation indices through combinatorial calculations using different formulas to identify distinctive features for classification. To further improve classification performance and create interpretable models, we include a selection strategy to retain only relevant index time series and exclude unnecessary ones. Our study was conducted in two Natura 2000 sites in central Italy, characterized by different environmental conditions and vegetation types. We configured three distinct hybrid models by varying input data types and feature selection strategies and compared the results.

The objectives of this study aim to address the following questions:

1. Do supervised hybrid classification approaches based on FDA produce a higher accuracy compared to machine learning methods directly applied to raw multi-temporal data in both test sites?
2. Among the examined hybrid approaches, is there one that consistently achieves the highest accuracy in both test sites?
3. Among the explored formulas, is there one that consistently produces the highest accuracy in both test sites?
4. Can an appropriate set of indices be identified for each study site?

This work is structured as follows: in Section 2 we introduce the materials and methods, focusing on the study area and the 'hybrid statistical–functional–machine learning' models to analyse and classify dense remotely sensed time series. In Section 3 we present the results of our methodology applied to two different case studies. In Section 4 we discuss the results

and the impact of the developed approach, and in Section 5 we provide conclusions and outline future work.

2. Materials and Methods

In this section we present two distinct approaches for classifying remotely sensed data (see Figure 2). We begin by collecting Sentinel-2 satellite time series data, which can be directly classified using Random Forest (first approach: 'Pure Machine Learning'). Alternatively, spectral bands and indices created through combinatorial methods were transformed into continuous functions using Generalized Additive Models (GAM) and analysed with FDA (including FPCA and MFPCA). Random Forest can then be used to classify the FPCA-MFPCA scores (second approach: 'Hybrid statistical-functional-Machine Learning'). Further details are provided in the following sub-section. The developed R code is available in [40].

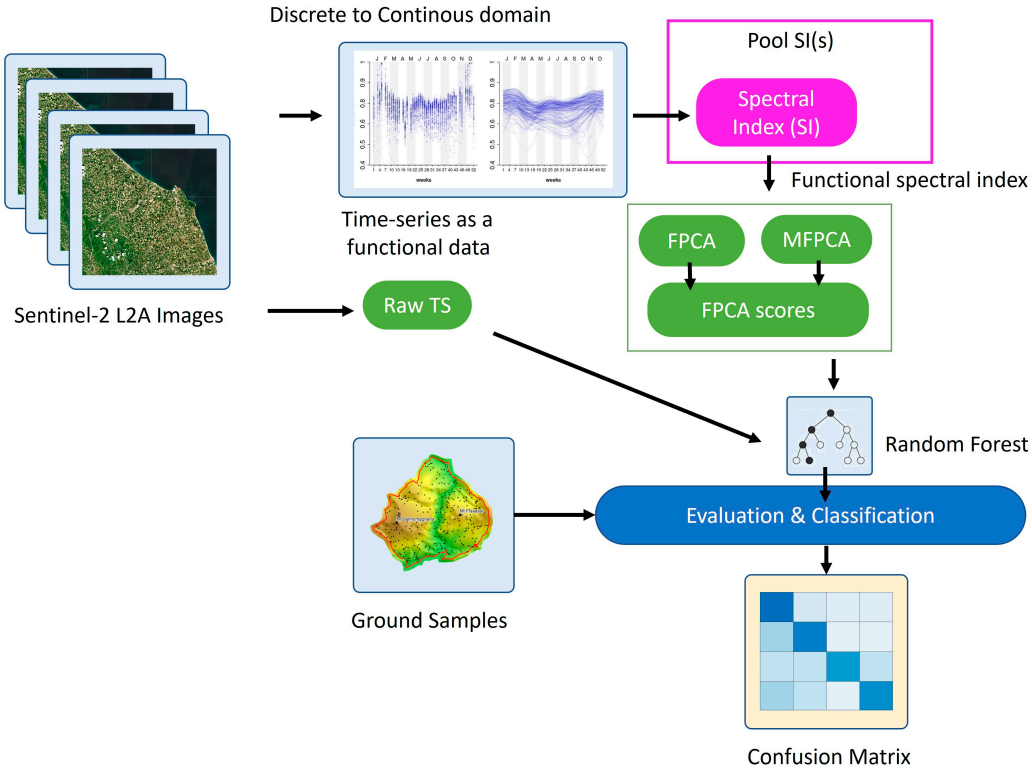

Figure 2. Starting from a set of Sentinel-2 images, we trigger a processing pipeline that extracts the most relevant vegetation indices that could be used to characterize the study area.

2.1. Study Area

This study focuses on two distinct areas of central Italy, specifically in the Marche region, which are part of the Natura 2000 network (Figure 3). The first area of interest is Mount Conero, situated in the coastal area of central Marche (43°33′00″N, 13°36′00″E). It is a Special Area of Conservation (SAC) known as 'Monte Conero' (code IT5320007) and covers an area of 650 hectares. Mount Conero has an elevation of 572 m above sea level, with an average annual precipitation of 710 mm and a mean annual temperature of 14.9 °C. The second study area is the 'Gola di Frasassi' (code IT5320003), also referred to as the Frasassi Gorge, located in the mountainous region of central Marche's Apennines (43°23′23″N, 12°57′36″E). This SAC spans an area of 728 hectares and reaches an altitude of 935 m above sea level. The average annual precipitation in this area is 1115 mm, while the mean annual temperature is 12.7 °C. According to the bioclimatic classification of Rivas-Martinez [41], both study areas belong to the temperate sub-Mediterranean macrobioclimate. The first area is characterised by a strong sub-Mediterranean level with pronounced summer aridity,

while the second area is characterised by a weak sub-Mediterranean level indicating lower summer aridity [42].

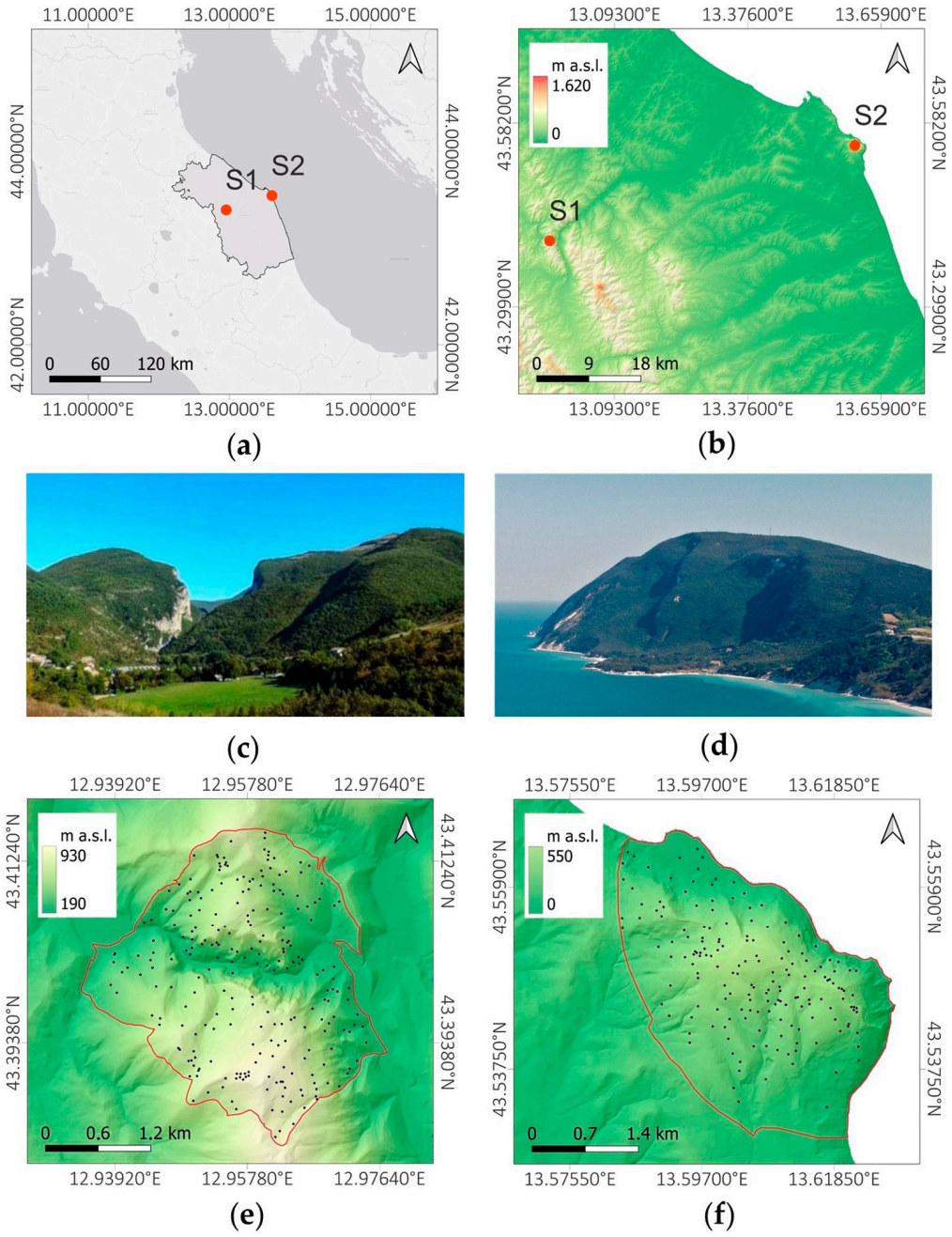

Figure 3. The two study areas: (**a**) national and (**b**) regional overview of the two study areas; S1 is the Frasassi Gorge, and S2 is Mount Conero. (**c**) Panoramic image of the Frasassi Gorge area. (**d**) Panoramic image of the Mount Conero area. (**e**) Reference data on the Digital Elevation Model with the boundary of the Frasassi Gorge Special Area of Conservation (SAC IT5320003). (**f**) Reference data on the Digital Elevation Model with the boundary of the Mount Conero area of interest.

2.2. Target Classes and Reference Data

Different vegetation types (recognised using the Braun-Blanquet approach) and the corresponding 92/43/EEC habitats are present in the two study areas. In the Mount Conero area, there are four different forest plant communities while the Frasassi Gorge area encompasses eight different vegetation typologies (four forests, two shrubs, one grassland and a mosaic of garrigue and chasmophitic vegetation). Detailed descriptions are provided in Table 1 and [29,30].

Table 1. Reference data for the study areas. Target classes for the supervised classification are listed. For plant associations, we report the syntaxa name and the corresponding habitat code (Annex 1 of the European Union Habitats Directive). The * denotes a priority habitat.

Class	Plant Association (Syntaxa)	Habitat Code	Plots
	Mount Conero area		172
	Woods		
c1	*Quercus ilex* evergreen forest with a high occurrence of Mediterranean species *Cyclamino hederifolii-Quercetum ilicis* [43].	9340	34
c2	*Quercus ilex* with deciduous trees mixed forest *Cephalanthero longifoliae-Quercetum ilicis* subass. *ruscetosum hypoglossy* [43].	9340	71
c3	*Ostrya carpinifolia* coastal deciduous forest *Asparago acutifolii–Ostryetum carpinifoliae* [44,45].	-	13
c4	Evergreen conifer forest plantations mostly dominated by *Pinus halepensis* and *P. pinea* [46].	-	54
	Frasassi Gorge area		241
	Woods		
v1	*Quercus ilex* (with deciduous trees) appenninic forest *Cephalanthero longifoliae-Quercetum ilicis* subass. *lathyretosum veneti* [43].	9340	34
v2	*Quercus pubescens* deciduous forest—*Cytiso sessilifolii-Quercetum pubescentis* [47,48].	91AA *	28
v3	*Ostrya carpinifolia* deciduous appenninic forest—*Scutellario columnae-Ostryetum carpinifoliae* [49].	-	56
v4	Evergreen conifer forest plantations mostly dominated *Pinus nigra* ssp. *nigra* and *P. halepensis* Mill. [50].	-	31
	Shrublands		
v5	*Spartium junceum* Shrub—*Spartio juncei-Cytisetum sessilifolii Spartium junceum* variant (Edoardo Biondi & Casavecchia, 2002).	-	16
v6	*Junyperus oxycedrus* shrub—*Spartio juncei-Cytisetum sessilifolii Juniperus oxycedrus* variant [51].	-	15
	Grasslands		
v7	*Bromus erectus* grassland—*Asperulo purpureae-Brometum erecti* [52].	6210 *	16
	Mosaic of garrigues and vegetation of rock and scree		
v8	*Satureja montana* Garrigues *Cephalario leucanthae-Saturejetum montanae* (could include 6110 and 6220 habitats); *Potentilla caulescens* and *Moehringia papulosa* chasmophytic vegetation of shady and wet rocky gorge's wall—*Moehringio papulosae-Potentilletum caulescentis* (habitat 8210 "Calcareous rocky slopes with chasmophytic vegetation") [52,53].	6110, 6220, 8210	46

The collected reference data, distributed over the two study areas are presented in Figure 3.

2.3. Remote Sensing Data Collection and Generation of Vegetation Indices

Sentinel-2 L2A images were acquired using the `Sen2r` package version 1.6.0 [54]. A total of 93 scenes (spanning from April 2017 to April 2020, as shown in Table A1) were collected for the two study areas, ensuring a cloud cover below 25% within the training plots. The images were pre-processed by masking the clouds and performing co-registration. A spatial resolution of 10 m was used, with the bands at 20 m being resampled using the nearest neighbours approach. Starting from the review of existing indices as in [55], we tried to summarize basic formulas, but we also considered other mapping functions. We considered up to 4 operands with basic rules to have a spectral order. The rules have been introduced to ensure a link with well-known indices such as the NDVI (type #3 in Table 2). The list of formulas is not related to a specific sensor/payload, and it could be applied to data acquired using aerial and satellite platforms. We considered Sentinel-2 bands, but the proposed approach can be applied to different types of platforms (e.g., Landsat-8).

Table 2. List of formulas for different types of indices. We analyse formulas with 2–4 operands and constraints on band order. We considered the following Sentinel-2 bands: B2, B3, B4, B5, B6, B7, B8*, B11, B12; * corresponds to B8–NIR (832.8 nm). More info of Sentinel-2 bands could be found here [56].

Formula #id	Formula	# of Operands	Constraint #1	Constraint #2	# of Combinations
0	A	1	-	-	9
1	$A-B$	2	$A>B$	-	36
2	A/B	2	$A>B$	-	36
3	$(A-B)/(A+B)$	2	$A>B$	-	36
4	$(A-B)/C$	3	$A>B$	$C>B$	84
5	$(A-B)/(C+B)$	3	$A>B$	$C>B$	84
6	$(A-B)/(C-B)$	3	$A>B$	$C>B$	84
7	$\frac{(A-B)/(A+B)}{(A+C)/(A-C)}$	3	$A>B$	$A>C$	84
8	$((A-B)/(A+B))((D-C)/(D+C))$	4	$A>B$	$D>C$	126
9	$A/B(C-D)/(C+D)$	4	$A>B$	$C>D$	126
10	$A/B(A-C)/(A+C)$	3	$A>C$	-	84
11	$A/B(B-C)/(B+C)$	3	$B>C$	-	84
12	$A/B{\cdot}C/D$	4	-	-	126
13	$(A-B)/(A+B+C+1e4)$	3	$A>B$	-	84
14	$((A-C)-(B-D))/((A-C)+(B-D))$	4	$A>C$	$B>D$	126
15	$(A-B)/(A+B+C)$	3	$A>B$	$B>C$	84
16	$(A-B)/((A+B-C)+1e4)$	3	$A>B$	$B>C$	84
17	$(2A-B-C)/(2A+B+C)$	3	$A>B$	$B>C$	84
18	$(A-(B+C))/(A+(B+C))$	3	$A>B$	$A>C$	84
19	$\log(A/B)$	2	-	-	36
20	$(A-B){\cdot}C$	3	$A>B$	-	84

2.4. Time Series as Functional Data

We arranged the 93 Sentinel-2 images chronologically by Day of the Year (DoY), Refs. [57–59] addressing outliers using the `clean.ts()` function from the R package `forecast` version 8.12 [60,61]. DoY values were aggregated into weekly averages (1–52 weeks) (e.g., Figure 4a). We interpolated and smoothed the weekly values using a GAM model with cyclic penalized cubic regression spline smooth (with default settings) [62]. GAMs have the advantage that they do not require measurements (like those of spectral bands) to be uniformly distributed, which is useful since clouds and other data issues cause random gaps in the data [63]. This process generated a weekly functional cubic cyclic spline representation of spectral variations in the plots (e.g., Figure 4b), and we applied it to all index formulas listed in Table 2. As mentioned in [36], the original discrete data were then set aside and the estimated curves (Figure 4b) were used for the rest of the analysis. The example R code for time series smoothing is available in [40] (repository 'habitatmapmfpca').

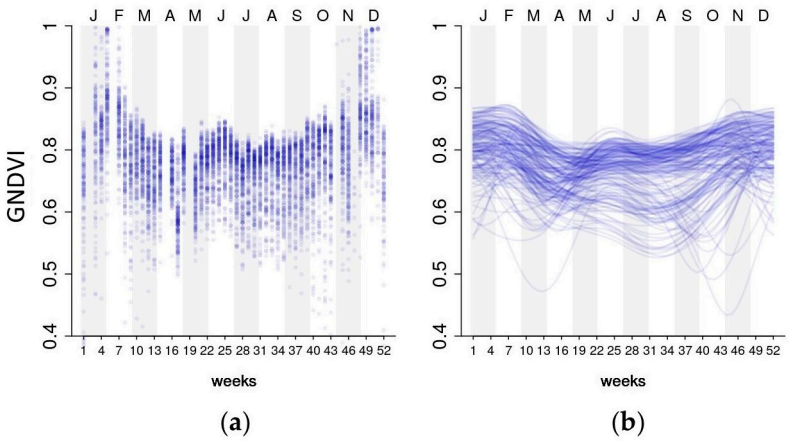

Figure 4. Example of derived time series considering mean weekly annual Sentinel-2 GNDVI variations (2017–2020) of the 172 plots of the Mount Conero study area. On the left (**a**) the discrete mean weekly time series, while on the right (**b**) the weekly functional cyclic cubic spline representation of the spectral plot variations. The letters at the top correspond to the initials of the months of the year.

2.5. Analysis of Functional Data Using FPCA and MFPCA

FPCA is a widely used FDA technique to reduce the amount of functional data [36,37]. It adapts traditional PCA concepts to functions, while preserving the functional structure (i.e., chronological order) of the observations (curves) [64]. FPCA extracts principal components (eigenfunctions representing the main modes of data variation) from the estimated curves, providing eigenvalues to quantify the captured variation and FPC scores to quantify curve similarities [32]. It is suitable for exploring and decomposing univariate functional spaces defined by a single variable. MFPCA extends FPCA to multivariate functional data, such as multiple bands or vegetation indices (Figure 1). It captures joint variations between functions, decomposing the data into orthogonal multivariate functional principal components (multivariate eigenfunctions) with eigenvalues and component scores. This provides a parsimonious data representation, with one score value per multivariate principal component per observation. The MFPCA scores, uncorrelated by construction, could be used for further statistical analyses (e.g., unsupervised functional clustering, supervised functional classification) [38] and graphical representation of the results for interpretation [32]. Univariate FPCA used the `fdaPace` R package version 0.5.5 [65] while MFPCA used the approach from [38] implemented in the associated R package version 1.3.6 [66].

2.6. Random Forest Classifier

Random Forest (RF) is a powerful ensemble learning classifier commonly used in habitat mapping studies based on remote sensing data [67]. We optimized RF performance by adjusting two key parameters: `ntree` (set to 1500) and `mtry` (evaluated from 1 to the square root of input variables) [68]. Imbalanced training and validation data can bias RF models in vegetation-related studies, over-predicting majority classes and under-predicting minority classes. To address this, we employed down-sampling in RF to balance class frequencies [29,69]. Additionally, we applied Recursive Feature Elimination to select important predictors and reduce input data dimensionality, enhancing model efficiency. These settings were maintained for all different supervised classification approaches (see following section).

2.7. Supervised Classification Approaches

We conducted supervised vegetation classification using Sentinel-2 temporal spectral variations through two approaches: 'Pure Machine Learning' and 'Hybrid Statistical-Functional Machine Learning' [19]. In the 'Pure Machine Learning' approach, we directly applied the RF classifier to raw Sentinel-2 multi-temporal imagery. The 'hybrid' approach integrated RF with FDA of dense time series, utilizing FPCA and MFPCA analyses for supervised classification. Specifically, we designed three hybrid models, each generating distinct input datasets for the classifier, consisting of separate FPCA and MFPCA scores. Details of these models are provided in the following subsections.

2.7.1. Pure Machine Learning Approach

Applying RF (or other machine learning methods) directly to raw satellite multi-temporal imagery data from discrete time series is a common method for vegetation and habitat mapping. These time series, typically based on a limited number of cloud-free scenes (e.g., <15%) selected within one year, can be constructed using individual spectral bands or predefined vegetation indices chosen by the authors [6,14,15,17,18,70–72]. In our study we used Sentinel-2 spectral bands discrete time series as input data for RF, avoiding an uncritical pre-selection among various available vegetation indices. We selected cloud-free images from 2019 according to the criteria discussed above, providing the broadest temporal coverage across different months for our study areas. For the Frasassi Gorge study area we selected 9 images (excluding January, May, November, and December due to cloud cover), and for the Mount Conero Area we selected 12 images (excluding January, September and December due to cloud cover) (see Table A1). This approach, considered as a baseline model, is referred to as B.

2.7.2. Hybrid Statistical–Functional–Machine Learning Approach

The first hybrid model used is the one proposed in [29], and referred to as mF. It involves analysing Multivariate Functional Spaces using multiple univariate FPCAs, one for each weekly vegetation index time series. The input data for RF consists of all univariate FPCA component scores. While mF models can be effective in terms of Overall Accuracy, it is important to note that the dimensionality of the input data can increase rapidly since univariate FPCA can extract about 6–7 components from each weekly vegetation index time series. The R code was developed in [29] and is available in [40] (repository 'habitatmapfrasassi').

For the second hybrid model, we applied MFPCA to simultaneously analyse and compress all weekly vegetation index time series generated by specific formulas (e.g., 36 indices for formula id #3—Table 2). We decided to extract a maximum of 36 multivariate functional principal components, balancing computational efficiency with effective vegetation characterization and classification. This decision was guided by the fact that, as previously mentioned, univariate FPCA typically only extracts about 6–7 components [29]. The resulting MFPCA components (multivariate eigenfunctions) and their scores offer a concise data representation [38]. The MFPCA scores for these 36 components served as input for the RF model, and this approach is denoted as M.

The third strategy aims to enhance vegetation classification accuracy by selecting a reduced set of time series indices specific to the study areas. This approach combines FPCA, MFPCA and RF through forward selection. For each iteration, an index time series was added and classified by RF (initially decomposed with univariate FPCA and subsequently with MFPCA). This process continued until no additional time series improved the model, with improvement assessed using the Overall Accuracy metric. As in the case of the M models, we limited MFPCA to extract a maximum of 36 components. The MFPCA scores from the selected index time series served as RF input data. This strategy is labelled Ms. The R code is available in [40] (repository 'habitatmapmfpca').

2.8. Accuracy Evaluation and Models Comparison

We assessed model accuracy using Overall Accuracy (OA), Producer Accuracy (PA), User Accuracy (UA) and the κ coefficient [73,74]. More details are reported in Table S1. To ensure robust estimates and minimize bias, we conducted 10-fold cross-validation five times, resulting in a cross-validated confusion matrix. RF models and accuracies were evaluated using the R `caret` package version 6.0.86 [75]. To compare all models simultaneously in terms of accuracy and complexity, we recorded OA, PA, the number of selected predictors (pr) and the final `mtry` of the RF model as columns in a data matrix. Each model (B, Ms, M, mF applied to each formula) was represented as a row in the matrix. Subsequently, we conducted a standardized Principal Component Analysis (PCA) on the data matrix.

3. Results

3.1. Models Performance and Comparison

The OA of the models is presented for both study areas, categorized into Pure Machine Learning and Hybrid Machine Learning approaches. Within the Hybrid Machine Learning category, the results are further detailed based on the different modelling strategies and indices formula ids. See Table 3 and Figure 5 for a summary of the results.

In the Mount Conero area, the baseline B model achieved an OA of 81.8%. Among the hybrid models, mF models exhibited an average OA of 84.3%, with the highest OA of 86% achieved using formula id #11 and the lowest at 81.6% with formula id #1. The M models had an average OA of 78.6%, with the highest OA of 85.6% obtained with formula id #18 and the lowest at 66.3% with formula id #8. The Ms models achieved an average OA of 84.4%, with the highest OA of 87.2% linked to formula id #15 and the lowest at 77.9% with formula id #4.

For the Frasassi Gorge area, the B model achieved an OA of 76.9%. Among the hybrid models, the mF models showed an average OA of 80.9%, with the highest OA of 82.9% achieved using formula id #3 and the lowest at 77.3% with formulas ids #0 and #1. The M models had an average OA of 74.2%, with the highest OA of 82.3% using formula id #7 and the lowest at 63.4% with formula id #17. Additionally, the Ms models obtained an average OA of 83.1%, with the highest OA of 86.5% linked to formula id #15 and the lowest at 81.1% with formula id #19.

Table 3. Comparison of model and formula performances in the two study areas based on Overall Accuracy. B—baseline model (Pure Machine Learning approach). mF, M, Ms—RF models based on Functional Data Analysis (Hybrid statistical—functional–Machine Learning approach). Formula id represents the different formulas used to generate indices detailed in Table 2. CO—Mount Conero area. VM—Frasassi Gorge area. In grey if the accuracy exceeds that of B. In bold, the best performance for each distinct hybrid approach.

	Mount Conero				Frasassi Gorge			
Formula #id	**B**	**mF**	**M**	**Ms**	**B**	**mF**	**M**	**Ms**
0	0.818	0.826	0.812	0.812	0.769	0.773	0.785	0.812
1		0.816	0.838	0.835		0.773	0.778	0.845
2		0.844	0.768	0.839		0.816	0.675	0.824
3		0.849	0.825	0.849		**0.829**	0.817	0.829
4		0.857	0.790	0.779		0.811	0.733	0.832
5		0.857	0.793	0.859		0.819	0.731	0.842
6		0.841	0.675	0.842		0.808	0.646	0.815
7		0.854	0.802	0.860		0.818	**0.823**	0.836
8		0.831	0.663	0.838		0.816	0.644	0.840
9		0.856	0.797	0.848		0.816	0.754	0.840
10		0.835	0.778	0.840		0.792	0.667	0.811
11		**0.860**	0.790	0.860		0.825	0.708	0.828
12		0.842	0.732	0.851		0.825	0.668	0.814
13		0.828	0.826	0.844		0.784	0.764	0.832
14		0.844	0.819	0.838		0.802	0.778	0.840
15		0.847	0.838	**0.872**		0.813	0.810	**0.865**
16		0.832	0.814	0.843		0.783	0.787	0.828
17		0.845	0.671	0.847		0.798	0.634	0.856
18		0.845	**0.856**	0.857		0.806	0.798	0.835
19		0.850	0.829	0.850		0.805	0.813	0.811
20		0.852	0.794	0.851		0.820	0.764	0.822
mean	0.818	0.843	0.786	0.844	0.8	0.806	0.742	0.831

In both study areas, the Ms and mF models consistently outperformed the M and B models, achieving a higher Overall Accuracy of 9.6 percentage points in the Frasassi Gorge area, and 5.4 percentage points in the Mount Conero area (see Figure 5 and Table 3). Furthermore, using indices (formula ids #1–#20 in Table 3) in the Ms and mF models demonstrated superior performance compared to using individual bands (formula id #0 in Table 3). In both study areas, the highest OA was achieved by the Ms models applied to vegetation indices with formula id #15 (see Tables 2 and 3 for its definition).

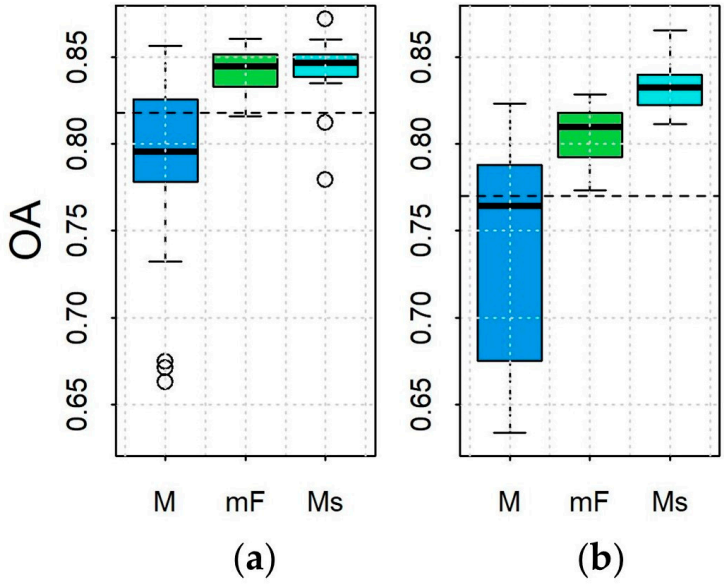

Figure 5. Comparison of Overall Accuracy (OA) among different model strategies for the two study areas. The dashed line represents the OA achieved by the baseline B model using a Pure Machine Learning approach. M, mF and Ms are three hybrid model strategies combining Random Forest with Functional Data Analysis (Hybrid statistical–functional–Machine Learning approach). (**a**) Mount Conero area. (**b**) Frasassi Gorge area.

Tables A2 and A3 offer a comprehensive overview of all models for both the Mount Conero and Frasassi Gorge areas providing accuracy (OA and PA), and complexity metrics (number of selected predictors, pr, and the final mtry of the RF). PCA of these tables (Figure 6) allows for a visual representation that facilitates model comparison based on their multivariate (inter- and intra-group) variability. Similar models are close together, and dissimilar models are further apart. The properties of the models are indicated by black arrows. The B model is represented by a red triangle, while the mF, M and Ms models applied to different formulas are represented in spider plots with distinct colours. The first principal component (PC1) axis, accounting for 49.5% and 43.8% of the total variation in the Mount Conero and Frasassi Gorge areas, respectively, indicates an increasing gradient of accuracy among the models. It clearly shows that the Ms and mF models outperform the B and M models in both OA (as shown in Table 3 and Figure 5) and PA. The second principal component (PC2) axis, which accounts for 22.5% and 17.0% of the total variation in the Mount Conero and Frasassi Gorge areas, respectively, is directly related to the increasing number of predictors used as input data (pr) and the `mtry` value.

PCA analysis reveals that the Ms models are the most parsimonious, achieving the highest OA and PA accuracy while using the fewest predictors and `mtry` (Figure 6).

Tables S2 and S3 provide details from the forward selection procedure used by Ms models. These tables outline the selected bands and indices that constitute the minimal set needed to optimize model performance in each formula and study area. The number of time series (bands or indices) selected ranged from 1 to 9 (1 to 7 for the Mount Conero area and 2 to 9 for the Frasassi Gorge area). The most frequently involved bands in the selected indices (in descending order) for the Frasassi Gorge area were B7, B5, B11, B4, B3, B12, while band B8 was the least utilized. For the Mount Conero area, the most utilized bands were B7, B6, B11, while bands B8 and B5 were less utilized.

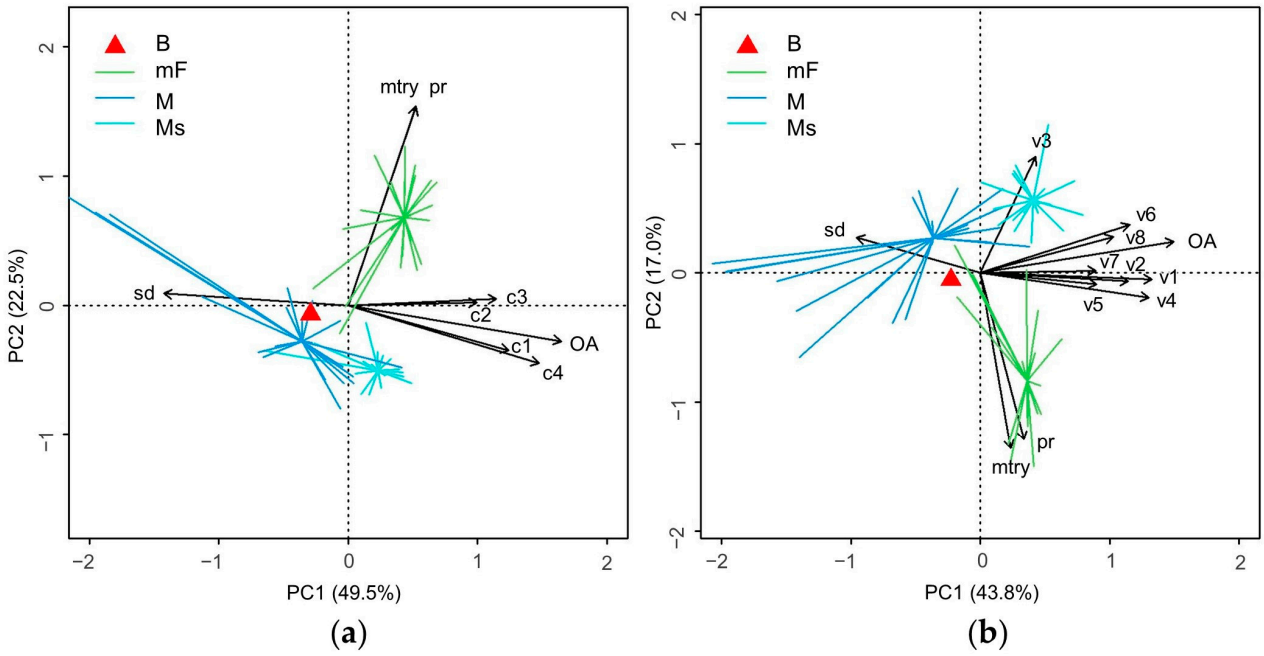

Figure 6. Principal Component biplot relating properties of accuracy and model complexity (black arrows) to the different supervised classification models (B, mF, M, Ms) applied to all distinct formulas. (**a**) Mount Conero Area. PCA axis 1 accounts for 49.5% of the multivariate variation and axis 2 for 22.5%. (**b**) Frasassi Gorge Area. PCA axis 1 accounts for 43.8% of the multivariate variation and axis 2 for 17.0%. Labels: OA–Overall Accuracy; sd–standard deviation; pr–number of input variables selected; mtry–final Random Forest mtry parameter; v1–v8 and c1–c4 are Producer Accuracy of vegetation types (listed in Table 1) for Frasassi Gorge and Mount Conero areas, respectively.

3.2. Best Models

The Ms models applied to formula id #15 (see Tables 2 and 3) achieved the highest OA in both study areas. Below, we summarise the accuracy results of these models and compare them to the B models by showing the error matrices (Tables 4 and 5). In the Supplementary Materials, detailed graphical representations of the two Ms models are provided (Figures S1 and S2), illustrating the selected time series and functional decomposition via MFPCA with the most discriminating components (seasonal variation) for the different vegetation types.

Table 4. Cross-validated confusion matrix (10-fold, repeated five times) for predicted target classes in the Mount Conero area. The table includes Overall Accuracy, Producer Accuracy, User Accuracy (expressed in percentage) and the κ statistic. The rows and columns (c1–c4) represent the plant associations and habitats listed in Table 1. B—baseline model (Pure Machine Learning approach). Ms-F15 (Ms model with the Formula id #15) is the top-performing model in terms of Overall Accuracy among the RF models based on Functional Data Analysis (Hybrid statistical–functional–Machine Learning approach). Pred stands for prediction.

		B						**Ms-Formula id #15**					
		Reference						**Reference**					
		c1	**c2**	**c3**	**c4**	**UA**			**c1**	**c2**	**c3**	**c4**	**UA**
Pred	**c1**	16.2	3.2	0.0	2.1	75.5	**Pred**	**c1**	39.2	3.7	3.1	3.4	79.4
	c2	4.0	36.2	3.9	3.0	76.9		**c2**	1.3	16.9	0.0	0.7	89.7
	c3	0.0	0.3	3.5	0.0	91.2		**c3**	0.0	0.0	4.3	0.0	100.0
	c4	0.9	0.8	0.0	25.8	93.8		**c4**	0.1	0.6	0.0	26.7	97.5
	PA	76.8	89.3	47.7	83.7			**PA**	96.6	80.0	58.5	86.7	
	OA	81.79 (±9.50)						**OA**	87.18 (±7.82)				
	K	0.72 (±0.14)						**K**	0.80 (±0.11)				

Table 5. Cross-validated confusion matrix (10-fold, repeated five times) for predicted target classes in the Frasassi Gorge area. The table includes Overall Accuracy, Producer Accuracy, User Accuracy (expressed in percentage) and the κ statistic. The rows and columns (v1–v8) represent the plant associations and habitats listed in Table 1. B—baseline model (Pure Machine Learning approach). Ms-F15 (Ms model with the Formula id #15) is the top-performing model in terms of Overall Accuracy among the RF models based on Functional Data Analysis (Hybrid statistical–functional–Machine Learning approach). Pred stands for prediction.

B										
		reference								
		v1	**v2**	**v3**	**v4**	**v5**	**v6**	**v7**	**v8**	
pred	**v1**	11.7	0	1.32	0.74	0	0	0	0	84.9
	v2	0	5.87	1.49	0	1.07	0	0.17	0	68.3
	v3	0.58	4.96	18.6	0.41	1.16	0	0	0	72.3
	v4	1.4	0	0.33	11.7	0	0.83	0	0	82.0
	v5	0.17	0.74	0.17	0	2.07	0	0.25	0.41	54.3
	v6	0	0	0	0	0.66	4.38	0	0.83	74.6
	v7	0	0	0.41	0	0.33	0	5.37	0.25	84.4
	v8	0.25	0	0.83	0	1.32	0.99	0.83	17.5	80.6
	PA	82.9	50.7	80.4	91.0	31.3	70.7	81.3	92.2	
	OA	76.99 (±7.07)								
	K	0.72 (±0.08)								
	Ms-Formula id #15									
		reference								
		v1	**v2**	**v3**	**v4**	**v5**	**v6**	**v7**	**v8**	**UA**
pred	**v1**	13.4	0.0	0.6	0.1	0.0	0.0	0.0	0.0	95.3
	v2	0.0	6.8	0.9	0.3	0.6	0.0	0.0	0.0	78.8
	v3	0.4	4.3	21.3	0.4	0.2	0.4	0.0	0.0	78.9
	v4	0.2	0.0	0.3	12.0	0.0	0.0	0.0	0.0	95.4
	v5	0.0	0.2	0.0	0.0	3.8	0.0	0.4	0.0	85.2
	v6	0.0	0.0	0.0	0.0	0.2	4.8	0.0	0.0	95.1
	v7	0.0	0.0	0.0	0.0	0.1	0.4	5.9	0.4	86.6
	v8	0.0	0.2	0.0	0.0	1.7	0.6	0.3	18.6	86.5
	PA	95.3	58.6	92.1	93.5	57.5	77.3	88.8	97.8	
	OA	86.51 (±6.99)								
	K	0.83 (±0.08)								

3.2.1. Mount Conero Area

The Ms model (applied to time series indices obtained with formula id #15) selected six time series for the Mount Conero area (A, B, C operators of the formula id #15 index), which were: (B12, B11, B03); (B07, B06, B04); (B11, B08, B07); (B08, B05, B04); (B07, B06, B03); (B12, B08, B06) (Table S2). Their seasonal variations and functional decomposition are depicted in Figure S1. With an OA of 87.18%, this model outperformed model B, which achieved 81.7%, and demonstrated a higher PA for the target classes c1, c3 and c4, as well as better UAs in all classes (Table 4).

3.2.2. Frasassi Gorge Area

The Ms model (applied to time series indices obtained with Formula id #15) selected nine time series for the Frasassi Gorge area (A, B, C operators of the Formula id #15 index), which were: (B10, B07, B04); (B08, B03, B02); (B07, B04, B02); (B07, B03, B02); (B10, B05, B04); (B11, B10, B04); (B06, B05, B04); (B08, B07, B04); (B07, B04, B03) (Table S3). Their seasonal variations and functional decomposition are depicted in Figure S2. With an OA of 86.5%, this model outperformed the 76.9% achieved by the B model. Furthermore, all PAs and UAs were higher for the Ms model compared to the B model (Table 5).

4. Discussion

4.1. Main Results

This study highlights the effectiveness of the 'Hybrid statistical–functional–Machine Learning' approach, which combines RF with an FDA of dense multispectral time series. The approach outperforms conventional methods that directly use RF on raw satellite multi-temporal images. Dense time series, when properly analysed and compressed, offer crucial information for characterizing seasonal spectral changes in vegetation, improving classification accuracy [26,27]. Ms models, which were the most accurate in both study areas, could be suitable tools with important practical implications for accurate classification, mapping and monitoring of vegetation and habitats included in Annex I of the 92/43/EEC Directive. Indeed, these models not only effectively process dense time series (increasingly accessible through web platforms like Google Earth Engine [76,77]) with FDA, but also independently identify sets of indices specific to the study area (through the forward selection strategy). The selection of location-specific indices plays a key role in optimizing the land management [24,25]. Thus, these models are adept at capturing vegetation and habitats during their optimal phenological stages without requiring prior knowledge of the best times for data acquisition or the most appropriate index sets, thus making them more transferable than conventional models [21]. In addition, the results of these models are graphically interpretable, contributing to a better understanding of critical seasonal multispectral variations among different plant communities and habitats (Figures S1 and S2).

Furthermore, the Ms models allowed us to employ new vegetation indices derived from a combinatorial approach and evaluate their effect on classification accuracy. The results revealed two aspects of particular interest. In both study areas, the most accurate models were the Ms models based on the formula id #15, an original index. In addition, rarely used indices based only on visible spectral bands played a significant role, confirming that classifications based only on known indices such as NDVI may not always be the most effective choice for classification purposes [20,78] or for characterizing plant communities. These results agree that specific plant communities and vegetation types have their own, specific multispectral profiles [24,26,79].

4.2. Models Comparison

4.2.1. Pure Machine Learning Approach: B Models

The B models demonstrated a lower accuracy, with a difference of up to 9.6% compared to the Hybrid statistical–functional–Machine Learning approach (see Table 3 and Figures 5 and 6). This lower performance was expected for several reasons. Model B typically employs input data based on time series of images selected for their cloud-free and low-cloud-cover conditions in a single reference year, reducing the data processing complexity, e.g., [14]. However, this approach often results in a limited number of images being available, with missing data for specific months. In our case, nine images were available for the Frasassi Gorge area and twelve for the Mount Conero area, covering different months depending on local weather conditions (e.g., excluding January, May, November and December for the Frasassi Gorge area and January, September and December for Mount Conero area due to cloud cover). This data gap may negatively impact the description of plant phenology and thus the accuracy of vegetation classification [80]. These models can be defined as "image-dependent" [81] since the timing and quality of image acquisition significantly impact classification accuracy [24]. Another crucial aspect to consider is that B models often skip important pre-processing steps aimed at noise detection, removal and reduction in time series, despite recommendations from [78,82], with a negative impact on accuracy.

4.2.2. Hybrid Statistical–Functional–Machine Learning Approach

Hybrid models that combine RF with FDA, overcoming the limitations of the B model, demonstrate a higher accuracy. The FDA approach treats temporal spectral variations as curves (smoothed functions) (e.g., Figures 1 and 4), allowing dense time series to be

analysed and offering richer information within a specific time window [32] than the B models for the classification stage. Unlike B models, hybrid models can be called 'image-independent' [81]. In these models, it is the quality of the functional data, which must adequately represent seasonal spectral variations in vegetation (e.g., Figure 4), that significantly influences the accuracy of the classification, rather than the timing and quality of the individual images used to create it. During the transformation of the raw data into a functional data using the GAM approach, it is essential to perform pre-processing steps to identify and remove outliers and reduce noise [83]. Another advantage over B models is that, to create pixel-based functions, it is better to exploit as much information as possible for each pixel. Thus, even images with only small areas without clouds or even one pixel without clouds can be used. In other words, if a part of an image is covered by clouds, this does not prevent the use of the part without clouds, whereas this is usually not the case for B models. We can assert that, if using dense time series data is an ideal choice for analysing seasonal variations in vegetation and achieving more accurate classifications [26,27,58], then FDA serves as an ideal tool for compressing and analysing dense time series data.

Ms, mF and M models have different characteristics and levels of accuracy. The Ms models are consistently better than the others in terms of Overall Accuracy for both study areas (Figure 5, Table 3). The superior performance is particularly evident, especially when applied to indices generated with formula id #15, in a more complex study area, such as the Frasassi Gorge, which has a higher number of target classes (Table 3). These models also performed better compared to previous studies. In the Mount Conero area, they achieved an 87.2% accuracy, exceeding the 83.2% accuracy in [30], which used only NDVI seasonal variation data. In the Frasassi Gorge area, these models achieved an 86.5% accuracy, exceeding the 82.1% accuracy in [29], obtained with mF models based on six time series of preselected indices (see Table 3). It is important to note that the Ms models are parsimonious. They achieved such a high accuracy with the smallest number of predictors and `mtry` (Figure 6, Tables A2 and A3), and this means that they can select a tailored and mutually complementary set of indices that best align with area-specific characteristics by capturing crucial seasonal multispectral variations. The key to this capability lies in the incorporation of two wrapper methods within Ms models, operating at distinct levels. Forward selection works on the entire index time series, while Recursive Feature Elimination focuses on individual MFPCA components extracted from the progressively selected time series. In summary, Ms models improve the characterization and distinction of various plant communities and habitats, enabling more accurate and detailed classifications. Their parsimonious nature makes them interpretable, contributing to a better understanding of critical seasonal multispectral variation among different plant communities and habitats (Figures S1 and S2). These hybrid models can complement species-based approaches in plant community ecology [30,32,33,38,84]. Besides their strengths, Ms models have some limitations. Indeed, forward selection does not guarantee the identification of the best model since the final set of selected indices is highly dependent on the first index chosen [85]. Moreover, they may require long computation times for evaluation, especially when dealing with many time series, such as those generated by formula id #15 (126 time series of indices). However, to improve the efficiency of these models and reduce the number of models to be evaluated, a preliminary filtering method could be implemented in future analyses. This method aims to identify and remove strongly correlated time series, allowing Ms models to process a smaller and more focused set of candidate time series.

The mF models, in line with prior research [29], demonstrated their effectiveness by achieving high accuracies. However, they also exhibited complexity and a lack of parsimony due to the utilization of many predictors (see Figure 6, Tables A2 and A3). This complexity arises from the limitation of multiple separate FPCAs in adequately addressing joint variations among different time series, resulting in the extraction of numerous correlated and redundant components. This redundancy makes the interpretation of results complicated [38]. Each vegetation plot has multiple scores associated with different univariate FPCA analyses which cannot be synthesized into a single functional reduced-ordination

space [29]. Consequently, while effective, these models are not very efficient and do not facilitate the understanding of crucial seasonal multispectral variation among different plant communities and habitats.

Finally, among the hybrid models, the M models proved to be less accurate. Their accuracies were modest and highly variable, consistently lower than the mF and Ms models, and often inferior to the B models as well (Table 3, Figures 5 and 6). The M models compress all the time series of vegetation indices associated with a specific formula using a single MFPCA, and the corresponding scores serve as input data for RF. It is likely that the established number of components extracted (k = 36) proved inadequate and too low, probably discarding useful seasonal variations for RF. To increase the accuracy of the model, one solution would be to increase the number of MFPCA components. However, this approach, as in mF models, hinders the identification of the minimum set of time series and indices specific to the vegetation of the study area. This limitation prevents us from fully capturing the crucial seasonal multispectral variations among different plant communities and habitats. In contrast, this method is suitable when the time series and indices specific to the study area are few and known.

4.3. Formula Comparison

Ms models performed best in both study areas using formula id #15. Surprisingly, this formula performed better than the well-known and widely used normalized difference (NDVI, Formula id #3) and simple difference (DVI, Formula id #1) formulas (Table 3). To our knowledge, formula id #15 is an original index that has not been found in the literature or common databases. It can be considered an extension of the normalized difference index, as it uses the difference between two bands in the numerator and the sum of the same two bands plus a third one in the denominator.

The results presented in Tables S2 and S3 show the final indices selected from the Ms models in the two study areas. In particular, the frequent use of Red Edge spectral bands (B5 and B7), SWIR (B11, B12) and, especially in the Frasassi Gorge area, visible bands (green and red, B3 and B4) is evident. These results are in line with previous studies [70,79,85–88], which emphasized the importance of these bands for distinguishing and mapping tree species, vegetation and habitats. The importance of visible bands is evident in the Frasassi Gorge area, where, out of five indices selected through formula id #1, which achieved a satisfactory Overall Accuracy, three are based exclusively on visible bands. This result is significant for habitat mapping (Directive 92/43/EEC) because these indices, which are often overlooked, can improve the accuracy of classification and offer the advantage of an intuitive understanding of their variations [89].

In this study, the NIR had a lower contribution to classification accuracy than the other bands mentioned above, despite the fact that its important role in vegetation mapping is well known and proven [7,90]. NIR plays a key role in satellites with a higher spatial but lower spectral resolution than Sentinel-2, such as IKONOS-2 and WorldView-2 [91,92].

4.4. Limits and Future Works

The first step in FDA is to transform raw data into functional objects by fitting discrete observations with curves that approximate the underlying continuous process. Achieving a balance between data fit and avoiding overfitting or neglecting essential aspects of the estimated smooth function is a common goal in the smoothing process [36]. Developing appropriate curves to describe the seasonal dynamics of vegetation across spectral bands or indices is crucial for accurate supervised vegetation classifications. Although promising results have been obtained in this and previous studies [29,30] using pixel-based functions interpolated with GAM (with default parameters: Knots = 10 and cross-validation for penalty value selection), future research could investigate how parameter variations and alternative smoothing methods [93,94] can improve classification accuracy. However, understanding the data-generating process and experimentation are fundamental tools in spline smoothing [28,36].

The Ms models demonstrated a superior performance in both study areas. However, the error matrices (Tables 3 and 4) revealed challenges in discriminating between some categories such as hornbeam and oak forests (e.g., 91AA* habitat). Incorporating topographical variables [29,95] and more extensive reference data could enhance model performance. The amount of reference data in our study, although well-distributed (see Figure 3), is relatively small and this may negatively affect the performance of classification [96] and the selection of time series. The main challenge, in fact, in mapping plant communities and habitats lies in the time required for field data collection [6]. The activities of "drone truthing", obtaining reference data through drones [97–99], offers a cost-effective way for biologists to verify satellite-derived maps, overcoming the limitations associated with ground-truthing for habitat mapping [100]. The acquired RGB images allow for the recognition of plant species [101], improving the efficiency of vegetation and habitat identification, even in complex environments, by recognizing indicator species of plant communities [102]. We are currently extending our analysis to other areas in the Central Apennines of Italy, where we have obtained extensive reference data through both 'ground-truthing' and 'drone-truthing'. Preliminary results confirm the effectiveness of the Ms models in selecting a minimal number of appropriate indices for the accurate classification of 16 different vegetation categories, demonstrating a significant level of discrimination for oak and hornbeam forests in this context.

In addition to statistical validation, the robustness of the model can be qualitatively assessed through the map generated by applying the model to all pixels [18,103]. In this study, we chose not to perform mapping. This is because, even if feasible, it would have been laborious given the numerous models developed. Our intention was to create a standardized and easily adaptable methodology that could select the most suitable indices for the study area.

Future developments will focus on evaluating other machine learning algorithms besides RF. One intriguing option could be the use of Linear Discriminant Analysis (LDA), which is also applicable in the functional context [104,105]. In the context of habitat and vegetation mapping, the adoption of a Hybrid statistical–functional model with LDA should ensure good classification results and at the same time identify the seasonal discriminant function that indicates the times when maximum differences between vegetation types emerge. This approach would improve the interpretability of the results from an ecological point of view, a crucial aspect for territorial entities engaged in habitat management and conservation, as required by the Habitats Directive.

5. Conclusions

In this paper we studied different approaches to supporting the classification of vegetation. These models combine machine learning, using RF, with the application of FDA to dense satellite time series. Our main goal was to improve the accuracy of vegetation and habitat classification in two different study areas. We achieved this by comparing the performance of these models to that of the most common classification methods, which apply machine learning directly to raw multi-temporal satellite data. Furthermore, we analysed the effect of different formulas for calculating vegetation indices, using a combinatorial approach. The goal was to identify the best approach and formula that consistently generated the best classification accuracies in both study areas. Now, analysing the results based on the research questions formulated at the beginning of this work, we derive the following conclusions:

1. The Hybrid supervised classification approaches based on FDA produce higher accuracy than common machine learning methods applied directly to raw multi-temporal data in both test areas.
2. Among the hybrid approaches examined, the Ms models achieve the highest accuracy in both test sites. These models effectively combine FDA, by exploiting MFPCA that compresses multiple time series based on different vegetation indices, with the use of RF. Using a forward selection strategy, we identified a limited set of indices that

meaningfully represent crucial multispectral seasonal variations obtaining really good results. Ms models are remarkably efficient, producing high accuracies with a low number of input data.

3. Among the formulas explored for calculating vegetation indices, the formula id #15 proved to be the best performing one in both study areas. However, other formulas have achieved good results (e.g., formula ids #17, #1), suggesting that further studies could be conducted in different study areas and with more reference data. In general, the use of indices rather than individual bands achieves better results.
4. This study demonstrated that Ms models can effectively identify a specific set of indices for each study area, adapting to the ecological characteristics and vegetation of the respective areas.

In conclusion, in scenarios characterized by an increasing availability of satellite data (and then dense time-series), we believe that Ms models could play a role of significant practical relevance in habitat monitoring and mapping. These models can identify the most suitable indices, based on the specific characteristics of the study site and the ecological and vegetation peculiarities of the analysed area, with the aim of maximizing the accuracy of the classifications. Furthermore, the results obtained can be integrated with the field data based on species recognition (for example, the Braun-Blanquet method), thus contributing to the understanding and conservation of biodiversity in the study areas. These models represent a promising contribution to overcoming the obstacle of transferability in remote sensing for the conservation of Natura 2000 habitats [21]. The R code for these models is available in [40] (repository habitatmapmfpca).

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/rs16071224/s1.

Author Contributions: Conceptualization S.P., A.M., G.Q. and S.C.; Data curation S.P. and A.M.; Formal analysis, S.P. and A.M.; Investigation, S.P., G.Q. and S.C.; Methodology, S.P., A.M. and G.Q.; Software, S.P. and A.M.; Supervision, S.P. and S.C.; Writing—original draft, S.P., A.M., G.Q. and S.C.; Writing—review and editing, S.P., A.M., G.Q. and S.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The R code for these models is available at https://github.com/geobotany/habitatmapmfpca (accessed on 27 March 2024).

Acknowledgments: The authors want to thank the Lorenzo Deplano, Riccardo Forconi and Cristian Colavito at the Department of Information Engineering (DII) of Università Politecnica delle Marche for their support to optimize the R code.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Table A1. Selection of Sentinel-2 Images: All images were employed to represent spectral seasonal variations as pixel-based functions, which were then used for Hybrid Statistical-Functional-Machine Learning models with RF Models based on Functional Data Analysis. The * and ** scenes from 2019 were used for the baseline model (Pure Machine Learning Approach) with Random Forest directly applied to raw time series for the Mount Conero and Frasassi Gorge areas, respectively.

Num	Date	Doy	Week	Month	Num	Date	Doy	Week	Month
1	21 April 2017	111	16	4	48	13 October 2018	286	41	10
2	1 May 2017	121	18	5	49	12 November 2018	316	46	11
3	31 May 2017	151	22	5	50	7 December 2018	341	49	12
4	20 June 2017	171	25	6	51	12 December 2018	346	50	12
5	10 July 2017	191	28	7	52	27 December 2018	361	52	12

Table A1. *Cont.*

Num	Date	Doy	Week	Month	Num	Date	Doy	Week	Month
6	20 July 2017	201	29	7	53	31 January 2019	31	5	1
7	30 July 2017	211	31	7	54	26 January 2019	26	4	1
8	9 August 2017	221	32	8	55	5 February 2019	36	6	2
9	19 August 2017	231	33	8	56	15 February 2019 **	46	7	2
10	29 August 2017	241	35	8	57	20 February 2019 *	51	8	2
11	18 September 2017	261	38	9	58	25 February 2019	56	8	2
12	8 October 2017	281	41	10	59	2 March 2019 **	61	9	3
13	18 October 2017	291	42	10	60	12 March 2019	71	11	3
14	28 October 2017	301	43	10	61	17 March 2019	76	11	3
15	27 November 2017	331	48	11	62	22 March 2019 *,**	81	12	3
16	7 December 2017	341	49	12	63	1 April 2019 **	91	13	4
17	22 December 2017	356	51	12	64	16 April 2019 *	106	16	4
18	6 January 2018	6	1	1	65	31 May 2019	151	22	5
19	15 February 2018	46	7	2	66	5 June 2019 *,**	156	23	6
20	6 April 2018	96	14	4	67	15 June 2019	166	24	6
21	16 April 2018	106	16	4	68	25 June 2019	176	26	6
22	21 April 2018	111	16	4	69	30 June 2019 *	181	26	6
23	26 April 2018	116	17	4	70	5 July 2019	186	27	7
24	11 May 2018	131	19	5	71	20 July 2019 *	201	29	7
25	16 May 2018	136	20	5	72	25 July 2019 **	206	30	7
26	21 May 2018	141	21	5	73	30 July 2019	211	31	7
27	31 May 2018	151	22	5	74	4 August 2019 *	216	31	8
28	10 June 2018	161	23	6	75	9 August 2019	221	32	8
29	15 June 2018	166	24	6	76	14 August 2019	226	33	8
30	20 June 2018	171	25	6	77	19 August 2019 **	231	33	8
31	30 June 2018	181	26	6	78	24 August 2019	236	34	8
32	10 July 2018	191	28	7	79	29 August 2019 *	241	35	8
33	15 July 2018	196	28	7	80	8 September 2019	251	36	9
34	20 July 2018	201	29	7	81	13 September 2019	256	37	9
35	25 July 2018	206	30	7	82	18 September 2019 **	261	38	9
36	30 July 2018	211	31	7	83	8 October 2019 *	281	41	10
37	4 August 2018	216	31	8	84	23 October 2019 **	296	43	10
38	9 August 2018	221	32	8	85	7 November 2019	311	45	11
39	19 August 2018	231	33	8	86	1 January 2020	1	1	1
40	24 August 2018	236	34	8	87	6 January 2020	6	1	1
41	29 August 2018	241	35	8	88	5 February 2020	36	6	2
42	3 September 2018	246	36	9	89	15 February 2020	46	7	2
43	8 September 2018	251	36	9	90	20 February 2020	51	8	2
44	18 September 2018	261	38	9	91	11 March 2020	71	11	3
45	23 September 2018	266	38	9	92	16 March 2020	76	11	3
46	28 September 2018	271	39	9	93	21 March 2020	81	12	3
47	3 October 2018	276	40	10					

Table A2. List of models for the Mount Conero area, displaying their accuracy (OA—Overall Accuracy and sd—standard deviation; for c1–c4 vegetation types Producer's Accuracy was reported) and model complexity (pr—number of input predictors and Random Forest's mtry value for tree splits).

Model	Formula	pr	mtry	OA	sd	c1	c2	c3	c4
B	0	38	4	0.818	0.095	0.768	0.893	0.477	0.837
M	0	6	1	0.812	0.076	0.692	0.901	0.538	0.844
M	1	2	1	0.838	0.085	0.730	0.887	0.754	0.867
M	2	18	1	0.768	0.082	0.757	0.887	0.015	0.800
M	3	6	1	0.825	0.076	0.654	0.941	0.462	0.878
M	4	34	1	0.790	0.081	0.714	0.930	0.031	0.841
M	5	36	2	0.793	0.075	0.768	0.899	0.015	0.859
M	6	30	5	0.675	0.092	0.400	0.893	0.138	0.704

Table A2. *Cont.*

Model	Formula	pr	mtry	OA	sd	c1	c2	c3	c4
M	7	26	5	0.802	0.082	0.703	0.927	0.385	0.807
M	8	34	5	0.663	0.120	0.454	0.930	0.185	0.567
M	9	36	3	0.797	0.091	0.703	0.935	0.262	0.807
M	10	14	1	0.778	0.087	0.768	0.918	0.000	0.789
M	11	10	2	0.790	0.088	0.686	0.868	0.508	0.830
M	12	22	3	0.732	0.088	0.708	0.882	0.000	0.730
M	13	10	1	0.827	0.074	0.719	0.955	0.354	0.844
M	14	34	4	0.671	0.098	0.562	0.859	0.385	0.570
M	15	10	3	0.856	0.072	0.762	0.938	0.615	0.870
M	16	6	2	0.814	0.076	0.659	0.904	0.615	0.848
M	17	6	2	0.838	0.081	0.730	0.899	0.585	0.893
M	18	10	2	0.829	0.081	0.714	0.913	0.492	0.878
M	19	18	4	0.794	0.081	0.735	0.904	0.354	0.796
M	20	36	6	0.793	0.090	0.703	0.921	0.215	0.826
mF	0	46	2	0.826	0.084	0.751	0.910	0.477	0.852
mF	1	258	11	0.816	0.082	0.714	0.893	0.492	0.863
mF	2	274	13	0.844	0.079	0.773	0.921	0.554	0.859
mF	3	290	7	0.849	0.074	0.719	0.944	0.615	0.870
mF	4	290	7	0.857	0.070	0.746	0.938	0.631	0.881
mF	5	630	15	0.857	0.070	0.751	0.955	0.585	0.867
mF	6	674	21	0.841	0.074	0.741	0.930	0.585	0.856
mF	7	294	15	0.854	0.066	0.724	0.924	0.738	0.878
mF	8	954	22	0.831	0.084	0.757	0.930	0.508	0.830
mF	9	818	19	0.856	0.071	0.730	0.972	0.523	0.870
mF	10	518	21	0.835	0.077	0.757	0.907	0.554	0.863
mF	11	658	20	0.860	0.072	0.751	0.963	0.631	0.856
mF	12	910	28	0.842	0.078	0.746	0.941	0.477	0.867
mF	13	118	2	0.828	0.084	0.703	0.907	0.631	0.859
mF	14	710	20	0.845	0.071	0.762	0.938	0.615	0.833
mF	15	634	24	0.845	0.072	0.730	0.927	0.662	0.863
mF	16	674	11	0.833	0.088	0.724	0.907	0.600	0.867
mF	17	610	7	0.847	0.070	0.730	0.932	0.646	0.863
mF	18	610	3	0.850	0.066	0.730	0.949	0.631	0.856
mF	19	250	6	0.852	0.077	0.735	0.944	0.600	0.870
mF	20	122	3	0.818	0.094	0.708	0.882	0.692	0.841
Ms	0	14	3	0.812	0.086	0.730	0.913	0.354	0.848
Ms	1	6	1	0.835	0.075	0.719	0.938	0.431	0.878
Ms	2	19	3	0.839	0.073	0.762	0.944	0.369	0.867
Ms	3	10	2	0.849	0.082	0.746	0.932	0.615	0.867
Ms	4	8	3	0.779	0.086	0.719	0.834	0.338	0.856
Ms	5	10	2	0.859	0.072	0.751	0.941	0.677	0.870
Ms	6	14	2	0.842	0.071	0.697	0.961	0.585	0.848
Ms	7	6	1	0.860	0.082	0.697	0.958	0.769	0.863
Ms	8	10	2	0.838	0.082	0.778	0.927	0.431	0.859
Ms	9	10	3	0.848	0.072	0.708	0.966	0.523	0.870
Ms	10	24	4	0.840	0.081	0.751	0.944	0.631	0.815
Ms	11	14	2	0.860	0.074	0.757	0.938	0.646	0.881
Ms	12	10	1	0.851	0.074	0.686	0.972	0.662	0.852
Ms	13	10	3	0.844	0.084	0.751	0.932	0.523	0.870
Ms	14	10	2	0.838	0.077	0.762	0.913	0.554	0.859
Ms	15	10	2	0.872	0.078	0.800	0.966	0.585	0.867
Ms	16	10	3	0.844	0.079	0.795	0.938	0.446	0.848
Ms	17	6	1	0.847	0.079	0.757	0.938	0.554	0.856
Ms	18	10	3	0.857	0.073	0.773	0.921	0.677	0.874
Ms	19	10	2	0.850	0.075	0.741	0.930	0.646	0.867
Ms	20	6	2	0.851	0.075	0.697	0.941	0.754	0.863

Table A3. List of models for the Frasassi Gorge area, displaying their accuracy (OA—Overall Accuracy and sd—standard deviation; for v1–v8 vegetation types Producer's Accuracy was reported) and model complexity (pr—number of input predictors and Random Forest's mtry value for tree splits).

Model	Formula	pr	mtry	OA	sd	v1	v2	v3	v4	v5	v6	v7	v8
B	0	62	3	0.770	0.071	0.829	0.507	0.804	0.910	0.313	0.707	0.813	0.922
M	0	26	4	0.785	0.070	0.771	0.571	0.857	0.813	0.350	0.707	0.850	0.974
M	1	26	5	0.778	0.076	0.882	0.436	0.864	0.871	0.287	0.653	0.825	0.935
M	2	30	5	0.675	0.083	0.676	0.371	0.696	0.794	0.550	0.600	0.775	0.791
M	3	30	5	0.817	0.063	0.853	0.600	0.839	0.877	0.400	0.827	0.875	0.974
M	4	34	5	0.733	0.080	0.682	0.421	0.718	0.903	0.550	0.653	0.813	0.930
M	5	34	5	0.731	0.079	0.700	0.529	0.721	0.826	0.525	0.707	0.800	0.883
M	6	34	5	0.646	0.078	0.682	0.314	0.857	0.587	0.137	0.293	0.675	0.887
M	7	34	5	0.823	0.080	0.924	0.500	0.879	0.865	0.413	0.893	0.800	0.978
M	8	36	6	0.644	0.081	0.618	0.300	0.893	0.613	0.187	0.093	0.588	0.948
M	9	22	2	0.754	0.077	0.747	0.386	0.843	0.761	0.425	0.747	0.850	0.957
M	10	36	4	0.667	0.081	0.506	0.393	0.718	0.710	0.512	0.640	0.750	0.900
M	11	36	6	0.708	0.075	0.712	0.164	0.768	0.839	0.463	0.680	0.775	0.952
M	12	36	4	0.668	0.090	0.588	0.407	0.743	0.626	0.375	0.613	0.763	0.913
M	13	30	3	0.764	0.074	0.835	0.307	0.896	0.787	0.300	0.747	0.775	0.974
M	14	34	3	0.634	0.072	0.659	0.179	0.882	0.690	0.050	0.053	0.838	0.870
M	15	30	4	0.798	0.069	0.841	0.543	0.829	0.897	0.375	0.773	0.788	0.978
M	16	18	3	0.788	0.071	0.771	0.536	0.904	0.787	0.400	0.693	0.875	0.948
M	17	18	2	0.810	0.071	0.812	0.521	0.843	0.839	0.562	0.893	0.825	0.978
M	18	18	4	0.813	0.078	0.924	0.586	0.807	0.852	0.463	0.787	0.813	0.978
M	19	30	4	0.764	0.071	0.841	0.486	0.825	0.761	0.338	0.773	0.813	0.935
M	20	34	4	0.786	0.070	0.771	0.493	0.861	0.890	0.325	0.760	0.800	0.978
mF	0	58	2	0.773	0.073	0.735	0.614	0.839	0.839	0.312	0.493	0.938	0.970
mF	1	250	3	0.773	0.066	0.806	0.550	0.850	0.819	0.350	0.640	0.863	0.922
mF	2	275	16	0.816	0.067	0.924	0.571	0.814	0.903	0.613	0.680	0.813	0.948
mF	3	202	7	0.829	0.062	0.912	0.579	0.807	0.942	0.588	0.707	0.875	0.978
mF	4	550	22	0.811	0.065	0.924	0.550	0.821	0.890	0.488	0.733	0.813	0.961
mF	5	550	9	0.819	0.073	0.935	0.600	0.821	0.890	0.525	0.720	0.813	0.952
mF	6	202	12	0.808	0.072	0.947	0.564	0.768	0.903	0.450	0.760	0.913	0.943
mF	7	606	15	0.818	0.066	0.953	0.579	0.789	0.897	0.563	0.693	0.813	0.978
mF	8	530	2	0.816	0.065	0.894	0.500	0.893	0.884	0.350	0.707	0.938	0.970
mF	9	998	17	0.816	0.062	0.853	0.529	0.900	0.871	0.475	0.720	0.813	0.978
mF	10	470	21	0.792	0.065	0.894	0.536	0.782	0.865	0.550	0.707	0.850	0.935
mF	11	606	12	0.825	0.065	0.912	0.536	0.882	0.890	0.500	0.707	0.813	0.978
mF	12	886	20	0.825	0.065	0.935	0.529	0.879	0.923	0.550	0.707	0.813	0.935
mF	13	498	1	0.785	0.064	0.853	0.493	0.879	0.832	0.375	0.627	0.863	0.935
mF	14	782	26	0.798	0.071	0.947	0.571	0.786	0.871	0.350	0.720	0.875	0.948
mF	15	646	25	0.806	0.067	0.935	0.557	0.761	0.910	0.475	0.707	0.850	0.978
mF	16	470	1	0.784	0.068	0.835	0.464	0.868	0.839	0.413	0.653	0.863	0.943
mF	17	510	10	0.813	0.066	0.906	0.600	0.786	0.903	0.500	0.693	0.875	0.978
mF	18	438	12	0.805	0.066	0.912	0.571	0.779	0.871	0.488	0.707	0.875	0.978
mF	19	202	4	0.820	0.063	0.906	0.607	0.814	0.890	0.575	0.693	0.813	0.978
mF	20	474	6	0.789	0.067	0.788	0.557	0.846	0.858	0.338	0.707	0.925	0.952
Ms	0	22	3	0.812	0.076	0.865	0.586	0.821	0.839	0.625	0.733	0.813	0.970
Ms	1	22	4	0.845	0.065	0.929	0.550	0.839	0.923	0.663	0.827	0.825	0.987
Ms	2	26	1	0.824	0.073	0.924	0.471	0.893	0.884	0.475	0.840	0.925	0.922
Ms	3	18	2	0.829	0.075	0.953	0.543	0.839	0.942	0.713	0.693	0.762	0.930
Ms	4	22	1	0.832	0.081	0.912	0.579	0.871	0.910	0.425	0.800	0.863	0.970
Ms	5	14	2	0.842	0.070	0.924	0.714	0.846	0.813	0.587	0.773	0.863	0.978
Ms	6	22	4	0.815	0.065	0.941	0.464	0.814	0.903	0.437	0.813	0.925	0.970
Ms	7	18	3	0.836	0.065	0.971	0.514	0.868	0.890	0.525	0.867	0.775	0.974
Ms	8	34	3	0.840	0.075	0.900	0.679	0.879	0.897	0.437	0.733	0.925	0.952
Ms	9	18	1	0.840	0.060	0.906	0.486	0.936	0.806	0.437	0.973	0.875	1.000
Ms	10	18	2	0.811	0.071	0.959	0.579	0.846	0.852	0.400	0.747	0.825	0.930

Table A3. *Cont.*

Model	Formula	pr	mtry	OA	sd	v1	v2	v3	v4	v5	v6	v7	v8
Ms	11	22	2	0.828	0.075	0.853	0.450	0.893	0.935	0.625	0.787	0.813	0.978
Ms	12	18	2	0.814	0.087	0.900	0.529	0.843	0.890	0.437	0.907	0.813	0.939
Ms	13	22	4	0.833	0.074	0.924	0.600	0.864	0.865	0.538	0.787	0.813	0.970
Ms	14	22	3	0.840	0.074	0.865	0.664	0.868	0.897	0.525	0.840	0.888	0.948
Ms	15	22	3	0.865	0.070	0.953	0.586	0.921	0.935	0.575	0.773	0.888	0.978
Ms	16	22	4	0.828	0.064	0.906	0.493	0.896	0.845	0.425	0.880	0.900	0.974
Ms	17	22	4	0.856	0.055	1.000	0.550	0.893	0.910	0.550	0.827	0.838	0.978
Ms	18	22	2	0.835	0.055	0.953	0.457	0.889	0.910	0.613	0.827	0.838	0.943
Ms	19	26	4	0.811	0.063	0.894	0.493	0.864	0.890	0.425	0.773	0.863	0.957
Ms	20	22	2	0.822	0.064	0.747	0.621	0.929	0.865	0.550	0.707	0.763	1.000

References

1. The Habitats Directive. Council Directive 92/43/EEC of 21 May 1992 on the Conservation of Natural Habitats and of Wild Fauna and Flora. *Off. J. L* **1992**, *206*, 7–50.
2. Evans, D. The Habitats of the European Union Habitats Directive. *Biol. Environ. Proc. R. Irish Acad.* **2006**, *106B*, 167–173. [CrossRef]
3. Corbane, C.; Lang, S.; Pipkins, K.; Alleaume, S.; Deshayes, M.; García Millán, V.E.; Strasser, T.; Vanden Borre, J.; Toon, S.; Michael, F. Remote Sensing for Mapping Natural Habitats and Their Conservation Status—New Opportunities and Challenges. *Int. J. Appl. Earth Obs. Geoinf.* **2015**, *37*, 7–16. [CrossRef]
4. Vanden Borre, J.; Paelinckx, D.; Mücher, C.A.; Kooistra, L.; Haest, B.; De Blust, G.; Schmidt, A.M. Integrating Remote Sensing in Natura 2000 Habitat Monitoring: Prospects on the Way Forward. *J. Nat. Conserv.* **2011**, *19*, 116–125. [CrossRef]
5. Schmidt, T.; Schuster, C.; Kleinschmit, B.; Förster, M. Evaluating an Intra-Annual Time Series for Grassland Classification—How Many Acquisitions and What Seasonal Origin Are Optimal? *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 3428–3439. [CrossRef]
6. Rapinel, S.; Rozo, C.; Delbosc, P.; Bioret, F.; Bouzillé, J.B.; Hubert-Moy, L. Contribution of Free Satellite Time-Series Images to Mapping Plant Communities in the Mediterranean Natura 2000 Site: The Example of Biguglia Pond in Corse (France). *Mediterr. Bot.* **2020**, *41*, 181–191. [CrossRef]
7. Marzialetti, F.; Giulio, S.; Malavasi, M.; Sperandii, M.G.; Acosta, A.T.R.; Carranza, M.L. Capturing Coastal Dune Natural Vegetation Types Using a Phenology-Based Mapping Approach: The Potential of Sentinel-2. *Remote Sens.* **2019**, *11*, 1506. [CrossRef]
8. Bajocco, S.; Ferrara, C.; Alivernini, A.; Bascietto, M.; Ricotta, C. Remotely-Sensed Phenology of Italian Forests: Going beyond the Species. *Int. J. Appl. Earth Obs. Geoinf.* **2019**, *74*, 314–321. [CrossRef]
9. Grignetti, A.; Salvatori, R.; Casacchia, R.; Manes, F. Mediterranean Vegetation Analysis by Multi-Temporal Satellite Sensor Data. *Int. J. Remote Sens.* **1997**, *18*, 1307–1318. [CrossRef]
10. Marzialetti, F.; Di Febbraro, M.; Malavasi, M.; Giulio, S.; Acosta, A.T.R.; Carranza, M.L. Mapping Coastal Dune Landscape through Spectral Rao's Q Temporal Diversity. *Remote Sens.* **2020**, *12*, 2315. [CrossRef]
11. Sittaro, F.; Hutengs, C.; Semella, S.; Vohland, M. A Machine Learning Framework for the Classification of Natura 2000 Habitat Types at Large Spatial Scales Using MODIS Surface Reflectance Data. *Remote Sens.* **2022**, *14*, 823. [CrossRef]
12. Mahmud, S.; Redowan, M.; Ahmed, R.; Khan, A.A.; Rahman, M.M. Phenology-Based Classification of Sentinel-2 Data to Detect Coastal Mangroves. *Geocarto Int.* **2022**, *37*, 14335–14354. [CrossRef]
13. Raab, C.; Stroh, H.G.; Tonn, B.; Meißner, M.; Rohwer, N.; Balkenhol, N.; Isselstein, J. Mapping Semi-Natural Grassland Communities Using Multi-Temporal RapidEye Remote Sensing Data. *Int. J. Remote Sens.* **2018**, *39*, 5638–5659. [CrossRef]
14. Hubert-Moy, L.; Fabre, E.; Rapinel, S. Contribution of SPOT-7 Multi-Temporal Imagery for Mapping Wetland Vegetation. *Eur. J. Remote Sens.* **2020**, *53*, 201–210. [CrossRef]
15. Jarocińska, A.; Kopeć, D.; Niedzielko, J.; Wylazłowska, J.; Halladin-Dąbrowska, A.; Charyton, J.; Piernik, A.; Kamiński, D. The Utility of Airborne Hyperspectral and Satellite Multispectral Images in Identifying Natura 2000 Non-Forest Habitats for Conservation Purposes. *Sci. Rep.* **2023**, *13*, 4549. [CrossRef] [PubMed]
16. Tarantino, C.; Forte, L.; Blonda, P.; Vicario, S.; Tomaselli, V.; Beierkuhnlein, C.; Adamo, M. Intra-Annual Sentinel-2 Time-Series Supporting Grassland Habitat Discrimination. *Remote Sens.* **2021**, *13*, 277. [CrossRef]
17. Buck, O.; Millán, V.E.G.; Klink, A.; Pakzad, K. Using Information Layers for Mapping Grassland Habitat Distribution at Local to Regional Scales. *Int. J. Appl. Earth Obs. Geoinf.* **2015**, *37*, 83–89. [CrossRef]
18. Rapinel, S.; Mony, C.; Lecoq, L.; Clément, B.; Thomas, A.; Hubert-Moy, L. Evaluation of Sentinel-2 Time-Series for Mapping Floodplain Grassland Plant Communities. *Remote Sens. Environ.* **2019**, *223*, 115–129. [CrossRef]
19. Durell, L.; Scott, J.T.; Hering, A.S. Hybrid Forecasting for Functional Time Series of Dissolved Oxygen Profiles. *Data Sci. Sci.* **2023**, *2*, 2152401. [CrossRef]

20. Huang, S.; Tang, L.; Hupy, J.P.; Wang, Y.; Shao, G. A Commentary Review on the Use of Normalized Difference Vegetation Index (NDVI) in the Era of Popular Remote Sensing. *J. For. Res.* **2021**, *32*, 1–6. [CrossRef]
21. Vanden Borre, J.; Spanhove, T.; Haest, B. Towards a Mature Age of Remote Sensing for Natura 2000 Habitat Conservation: Poor Method Transferability as a Prime Obstacle. In *The Roles of Remote Sensing in Nature Conservation*; Springer International Publishing: Cham, Switzerland, 2017; pp. 11–37.
22. Xue, J.; Su, B. Significant Remote Sensing Vegetation Indices: A Review of Developments and Applications. *J. Sens.* **2017**, *2017*, 1353691. [CrossRef]
23. Fatima, N.; Javed, A. Assessment of Land Use Land Cover Change Detection Using Geospatial Techniques in Southeast Rajasthan. *J. Geosci. Environ. Prot.* **2021**, *9*, 299–319. [CrossRef]
24. Barrett, B.; Raab, C.; Cawkwell, F.; Green, S. Upland Vegetation Mapping Using Random Forests with Optical and Radar Satellite Data. *Remote Sens. Ecol. Conserv.* **2016**, *2*, 212–231. [CrossRef] [PubMed]
25. Nagendra, H.; Lucas, R.; Honrado, J.P.; Jongman, R.H.G.; Tarantino, C.; Adamo, M.; Mairota, P. Remote Sensing for Conservation Monitoring: Assessing Protected Areas, Habitat Extent, Habitat Condition, Species Diversity, and Threats. *Ecol. Indic.* **2013**, *33*, 45–59. [CrossRef]
26. Pasquarella, V.J.; Holden, C.E.; Kaufman, L.; Woodcock, C.E. From Imagery to Ecology: Leveraging Time Series of All Available Landsat Observations to Map and Monitor Ecosystem State and Dynamics. *Remote Sens. Ecol. Conserv.* **2016**, *2*, 152–170. [CrossRef]
27. Gillanders, S.N.; Coops, N.C.; Wulder, M.A.; Gergel, S.E.; Nelson, T. Multitemporal Remote Sensing of Landscape Dynamics and Pattern Change: Describing Natural and Anthropogenic Trends. *Prog. Phys. Geogr. Earth Environ.* **2008**, *32*, 503–528. [CrossRef]
28. Ramsay, J.O.; Silverman, B.W. *Functional Data Analysis*; Ramsay, R., Silverman, B., Eds.; Springer Series in Statistics; Springer: New York, NY, USA, 2005; ISBN 978-0-387-40080-8.
29. Pesaresi, S.; Mancini, A.; Quattrini, G.; Casavecchia, S. Functional Analysis for Habitat Mapping in a Special Area of Conservation Using Sentinel-2 Time-Series Data. *Remote Sens.* **2022**, *14*, 1179. [CrossRef]
30. Pesaresi, S.; Mancini, A.; Quattrini, G.; Casavecchia, S. Mapping Mediterranean Forest Plant Associations and Habitats with Functional Principal Component Analysis Using Landsat 8 NDVI Time Series. *Remote Sens.* **2020**, *12*, 1132. [CrossRef]
31. Coviello, L.; Martini, F.M.; Cesaretti, L.; Pesaresi, S.; Solfanelli, F.; Mancini, A. Clustering of Remotely Sensed Time Series Using Functional Principal Component Analysis to Monitor Crops. In Proceedings of the 2022 IEEE Workshop on Metrology for Agriculture and Forestry (MetroAgriFor), Perugia, Italy, 3–5 November 2022; pp. 141–145.
32. Hurley, M.A.; Hebblewhite, M.; Gaillard, J.; Dray, S.; Taylor, K.A.; Smith, W.K.; Zager, P.; Bonenfant, C. Functional Analysis of Normalized Difference Vegetation Index Curves Reveals Overwinter Mule Deer Survival Is Driven by Both Spring and Autumn Phenology. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2014**, *369*, 20130196. [CrossRef]
33. Pesaresi, S.; Mancini, A.; Casavecchia, S. Recognition and Characterization of Forest Plant Communities through Remote-Sensing NDVI Time Series. *Diversity* **2020**, *12*, 313. [CrossRef]
34. Ramsay, J.O. When the Data Are Functions. *Psychometrika* **1982**, *47*, 379–396. [CrossRef]
35. Kennedy, R.E.; Andréfouët, S.; Cohen, W.B.; Gómez, C.; Griffiths, P.; Hais, M.; Healey, S.P.; Helmer, E.H.; Hostert, P.; Lyons, M.B.; et al. Bringing an Ecological View of Change to Landsat-Based Remote Sensing. *Front. Ecol. Environ.* **2014**, *12*, 339–346. [CrossRef] [PubMed]
36. Levitin, D.J.; Nuzzo, R.L.; Vines, B.; Ramsay, J.O. Introduction to Functional Data Analysis. *Can. Psychol.* **2007**, *48*, 135–155. [CrossRef]
37. Ramsay, J.O.; Dalzell, C.J. Some Tools for Functional Data Analysis. *J. R. Stat. Soc. Ser. B* **1991**, *53*, 539–572. [CrossRef]
38. Happ, C.; Greven, S. Multivariate Functional Principal Component Analysis for Data Observed on Different (Dimensional) Domains. *J. Am. Stat. Assoc.* **2018**, *113*, 649–659. [CrossRef]
39. Wang, J.-L.; Chiou, J.-M.; Müller, H.-G. Functional Data Analysis. *Annu. Rev. Stat. Its Appl.* **2016**, *3*, 257–295. [CrossRef]
40. Geobotanic Group at Università Politecnica delle Marche. Dataset and R Code Related to the Habitat Mapping with Functional Hybrid Machine Learning. Available online: https://github.com/geobotany (accessed on 15 January 2024).
41. Rivas-Martínez, S.; Sáenz, S.R.; Penas, A. Worldwide Bioclimatic Classification System. *Glob. Geobot.* **2011**, *1*, 1–634.
42. Pesaresi, S.; Biondi, E.; Casavecchia, S. Bioclimates of Italy. *J. Maps* **2017**, *13*, 955–960. [CrossRef]
43. Biondi, E.; Casavecchia, S.; Gigante, D. Contribution to the Syntaxonomic Knowledge of the Quercus Ilex L. Woods of the Central European Mediterranean Basin. *Fitosociologia* **2003**, *40*, 129–156.
44. Biondi, E.; Gubellini, L.; Pinzi, M.; Casavecchia, S. The Vascular Flora of Conero Regional Nature Park (Marche, Central Italy). *Flora Mediterr.* **2012**, *22*, 67–167. [CrossRef]
45. Biondi, E. L'ostrya Carpinifolia Scop. Sul Litorale Delle Marche (Italia Centrale). *Stud. Geobot.* **1982**, *2*, 141–147.
46. Baiocco, M.; Casavecchia, S.; Biondi, E.; Pietracapina, A. Indagini Geobotaniche per Il Recupero Del Rimboschimento Del Monte Conero (Italia Centrale). *Doc. Phytosociol.* **1996**, *16*, 387–425.
47. Blasi, C.; Di Pietro, R.; Filesi, L. Syntaxonomical Revision of Quercetalia Pubescenti-Petraeae in the Italian Peninsula. *Fitosociologia* **2004**, *41*, 87–164.
48. Blasi, C.; Feoli, E.; Avena, G.C. Due Nuove Associazioni Dei Quercetalia Pubescentis Dell'Appennino Centrale. *Stud. Geobot.* **1982**, *2*, 155–167.

49. Pedrotti, F.; Ballelli, S.; Biondi, E.; Cortini Pedrotti, C.; Orsomando, E. Resoconto Dell'escursione Della Società Italiana Di Fitosociologia Nelle Marche Ed in Umbria (11–14 Giugno 1979). *Not. Fitosociologico* **1980**, *16*, 73–75.
50. Allegrezza, M.; Pesaresi, S.; Ballelli, S.; Tesei, G.; Ottaviani, C. Influences of Mature Pinus Nigra Plantations on the Floristic-Vegetational Composition along an Altitudinal Gradient in the Central Apennines, Italy. *iForest* **2020**, *13*, 279–285. [CrossRef]
51. Biondi, E.; Casavecchia, S. Inquadramento Fitosociologico Della Vegetazione Arbustiva Di Un Settore Dell'Appennino Settentrionale. *Fitosociologia* **2002**, *39*, 65–73.
52. Biondi, E.; Allegrezza, M.; Zuccarello, V. Syntaxonomic Revision of the Apennine Grasslands Belonging to Brometalia Erecti, and an Analysis of Their Relationships with the Xerophilous Vegetation of Rosmarinetea Officinalis (Italy). *Phytocoenologia* **2005**, *35*, 129–164. [CrossRef]
53. Allegrezza, M.; Biondi, E.; Ballelli, S.; Formica, E. La Vegetazione Dei Settori Rupestri Calcarei Dell'Italia Centrale. *Fitosociologia* **1997**, *32*, 91–120.
54. Ranghetti, L.; Boschetti, M.; Nutini, F.; Busetto, L. "Sen2r": An R Toolbox for Automatically Downloading and Preprocessing Sentinel-2 Satellite Data. *Comput. Geosci.* **2020**, *139*, 104473. [CrossRef]
55. Zeng, Y.; Hao, D.; Huete, A.; Dechant, B.; Berry, J.; Chen, J.M.; Joiner, J.; Frankenberg, C.; Bond-Lamberty, B.; Ryu, Y.; et al. Optical Vegetation Indices for Monitoring Terrestrial Ecosystems Globally. *Nat. Rev. Earth Environ.* **2022**, *3*, 477–493. [CrossRef]
56. ESA. Sentinel-2 User Handbook. Available online: https://sentinel.esa.int/documents/247904/685211/sentinel-2_user_handbook (accessed on 15 January 2024).
57. Fisher, J.I.; Mustard, J.F.; Vadeboncoeur, M.A. Green Leaf Phenology at Landsat Resolution: Scaling from the Field to the Satellite. *Remote Sens. Environ.* **2006**, *100*, 265–279. [CrossRef]
58. Schuster, C.; Schmidt, T.; Conrad, C.; Kleinschmit, B.; Förster, M. Grassland habitat mapping by intra-annual time series analysis—Comparison of RapidEye and TerraSAR-X satellite data. *Int. J. Appl. Earth Obs. Geoinf.* **2015**, *34*, 25–34. [CrossRef]
59. Lambert, J.; Drenou, C.; Denux, J.; Balent, G.; Cheret, V. Monitoring Forest Decline through Remote Sensing Time Series Analysis. *GISci. Remote Sens.* **2013**, *50*, 437–457. [CrossRef]
60. Hyndman, R.; Athanasopoulos, G.; Bergmeir, C.; Caceres, G.; Chhay, L.; O'Hara-Wild, M.; Petropoulos, F.; Razbash, S.; Wang, E.; Yasmeen, F. Forecast: Forecasting Functions for Time Series and Linear Models. R Package Version 8.6. Available online: https://cran.r-project.org/package=forecast (accessed on 3 August 2020).
61. Hyndman, R.J.; Khandakar, Y. Automatic Time Series Forecasting: The Forecast Package for R. *J. Stat. Softw.* **2008**, *27*, 1–22. [CrossRef]
62. Wood, S.N. *Generalized Additive Models: An Introduction with R*; Chapman and Hall/CRC: New York, NY, USA, 2017; ISBN 9781315370279.
63. Younes, N.; Joyce, K.E.; Maier, S.W. All Models of Satellite-Derived Phenology Are Wrong, but Some Are Useful: A Case Study from Northern Australia. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *97*, 102285. [CrossRef]
64. Di Salvo, F.; Ruggieri, M.; Plaia, A. Functional Principal Component Analysis for Multivariate Multidimensional Environmental Data. *Environ. Ecol. Stat.* **2015**, *22*, 739–757. [CrossRef]
65. Dai, X.; Hadjipantelis, P.Z.; Han, K.; Ji, H. Fdapace: Functional Data Analysis and Empirical Dynamics. R Package Version 0.5.5. Available online: https://cran.r-project.org/package=fdapace (accessed on 3 August 2020).
66. Happ-Kurz, C. MFPCA: Multivariate Functional Principal Component Analysis for Data Observed on Different Dimensional Domains. R Package Version 1.3-6. Available online: https://cran.r-project.org/web/packages/MFPCA/index.html (accessed on 22 March 2022).
67. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
68. Belgiu, M.; Drăguţ, L. Random Forest in Remote Sensing: A Review of Applications and Future Directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31. [CrossRef]
69. Evans, J.S.; Cushman, S.A. Gradient Modeling of Conifer Species Using Random Forests. *Landsc. Ecol.* **2009**, *24*, 673–683. [CrossRef]
70. Le Dez, M.; Robin, M.; Launeau, P. Contribution of Sentinel-2 Satellite Images for Habitat Mapping of the Natura 2000 Site 'Estuaire de La Loire' (France). *Remote Sens. Appl. Soc. Environ.* **2021**, *24*, 100637. [CrossRef]
71. Marcinkowska-Ochtyra, A.; Ochtyra, A.; Raczko, E.; Kopeć, D. Natura 2000 Grassland Habitats Mapping Based on Spectro-Temporal Dimension of Sentinel-2 Images with Machine Learning. *Remote Sens.* **2023**, *15*, 1388. [CrossRef]
72. Wakulińska, M.; Marcinkowska-Ochtyra, A. Multi-Temporal Sentinel-2 Data in Classification of Mountain Vegetation. *Remote Sens.* **2020**, *12*, 2696. [CrossRef]
73. Congalton, R.G. A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data. *Remote Sens. Environ.* **1991**, *37*, 35–46. [CrossRef]
74. Cohen, J. A Coefficient of Agreement for Nominal Scales. *Educ. Psychol. Meas.* **1960**, *20*, 37–46. [CrossRef]
75. Kuhn, M. Building Predictive Models in R Using the Caret Package. *J. Stat. Softw.* **2008**, *28*, 1–26. [CrossRef]
76. Pham-Duc, B.; Nguyen, H.; Phan, H.; Tran-Anh, Q. Trends and Applications of Google Earth Engine in Remote Sensing and Earth Science Research: A Bibliometric Analysis Using Scopus Database. *Earth Sci. Inform.* **2023**, *16*, 2355–2371. [CrossRef]
77. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-Scale Geospatial Analysis for Everyone. *Remote Sens. Environ.* **2017**, *202*, 18–27. [CrossRef]

78. Pettorelli, N.; Vik, J.O.; Mysterud, A.; Gaillard, J.-M.; Tucker, C.J.; Stenseth, N.C. Using the Satellite-Derived NDVI to Assess Ecological Responses to Environmental Change. *Trends Ecol. Evol.* **2005**, *20*, 503–510. [CrossRef]
79. Grabska, E.; Hostert, P.; Pflugmacher, D.; Ostapowicz, K. Forest Stand Species Mapping Using the Sentinel-2 Time Series. *Remote Sens.* **2019**, *11*, 1197. [CrossRef]
80. Vrieling, A.; Meroni, M.; Darvishzadeh, R.; Skidmore, A.K.; Wang, T.; Zurita-Milla, R.; Oosterbeek, K.; O'Connor, B.; Paganini, M. Vegetation Phenology from Sentinel-2 and Field Cameras for a Dutch Barrier Island. *Remote Sens. Environ.* **2018**, *215*, 517–529. [CrossRef]
81. Pasquarella, V.J.; Holden, C.E.; Woodcock, C.E. Improved Mapping of Forest Type Using Spectral-Temporal Landsat Features. *Remote Sens. Environ.* **2018**, *210*, 193–207. [CrossRef]
82. Alvera-Azcárate, A.; Sirjacobs, D.; Barth, A.; Beckers, J.-M. Outlier Detection in Satellite Data Using Spatial Coherence. *Remote Sens. Environ.* **2012**, *119*, 84–91. [CrossRef]
83. Balestra, M.; Pierdicca, R.; Cesaretti, L.; Quattrini, G.; Mancini, A.; Galli, A.; Malinverni, E.S.; Casavecchia, S.; Pesaresi, S. A comparison of pre-processing approaches for remotely sensed time series classification based on functional analysis. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2023**. [CrossRef]
84. Liu, C.; Ray, S.; Hooker, G.; Friedl, M. Functional Factor Analysis for Periodic Remote Sensing Data. *Ann. Appl. Stat.* **2012**, *6*, 601–624. [CrossRef]
85. Fassnacht, F.E.; Neumann, C.; Forster, M.; Buddenbaum, H.; Ghosh, A.; Clasen, A.; Joshi, P.K.; Koch, B. Comparison of Feature Reduction Algorithms for Classifying Tree Species with Hyperspectral Data on Three Central European Test Sites. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 2547–2561. [CrossRef]
86. Saini, R.; Ghosh, S.K. Analyzing the Impact of Red-Edge Band on Land Use Land Cover Classification Using Multispectral RapidEye Imagery and Machine Learning Techniques. *J. Appl. Remote Sens.* **2019**, *13*, 044511. [CrossRef]
87. Schuster, C.; Förster, M.; Kleinschmit, B. Testing the Red Edge Channel for Improving Land-Use Classifications Based on High-Resolution Multi-Spectral Satellite Data. *Int. J. Remote Sens.* **2012**, *33*, 5583–5599. [CrossRef]
88. Immitzer, M.; Vuolo, F.; Atzberger, C. First Experience with Sentinel-2 Data for Crop and Tree Species Classifications in Central Europe. *Remote Sens.* **2016**, *8*, 166. [CrossRef]
89. Meyer, G.E.; Neto, J.C. Verification of Color Vegetation Indices for Automated Crop Imaging Applications. *Comput. Electron. Agric.* **2008**, *63*, 282–293. [CrossRef]
90. Alcaraz-Segura, D.; Cabello, J.; Paruelo, J. Baseline Characterization of Major Iberian Vegetation Types Based on the NDVI Dynamics. *Plant Ecol.* **2009**, *202*, 13–29. [CrossRef]
91. Saini, R. Integrating Vegetation Indices and Spectral Features for Vegetation Mapping from Multispectral Satellite Imagery Using AdaBoost and Random Forest Machine Learning Classifiers. *Geomat. Environ. Eng.* **2022**, *17*, 57–74. [CrossRef]
92. Illarionova, S.; Shadrin, D.; Trekin, A.; Ignatiev, V.; Oseledets, I. Generation of the NIR Spectral Band for Satellite Images with Convolutional Neural Networks. *Sensors* **2021**, *21*, 5646. [CrossRef] [PubMed]
93. Chen, J.; Jo, P. A Simple Method for Reconstructing a High-Quality NDVI Time-Series Data Set Based on the Savitzky–Golay Filter. *Remote Sens. Environ.* **2004**, *91*, 332–344. [CrossRef]
94. Li, S.; Xu, L.; Jing, Y.; Yin, H.; Li, X.; Guan, X. High-Quality Vegetation Index Product Generation: A Review of NDVI Time Series Reconstruction Techniques. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *105*, 102640. [CrossRef]
95. Marcinkowska-Ochtyra, A.; Gryguc, K.; Ochtyra, A.; Kopeć, D.; Jarocińska, A.; Sławik, Ł. Multitemporal Hyperspectral Data Fusion with Topographic Indices—Improving Classification of Natura 2000 Grassland Habitats. *Remote Sens.* **2019**, *11*, 2264. [CrossRef]
96. Tuia, D.; Persello, C.; Bruzzone, L. Domain Adaptation for the Classification of Remote Sensing Data: An Overview of Recent Advances. *IEEE Geosci. Remote Sens. Mag.* **2016**, *4*, 41–57. [CrossRef]
97. Piel, A.K.; Crunchant, A.; Knot, I.E.; Chalmers, C.; Fergus, P.; Mulero-Pázmány, M.; Wich, S.A. Noninvasive Technologies for Primate Conservation in the 21st Century. *Int. J. Primatol.* **2022**, *43*, 133–167. [CrossRef]
98. Suir, G.; Saltus, C.; Sasser, C.; Harris, J.; Reif, M.; Diaz, R.; Giffin, G. *Evaluating Drone Truthing as an Alternative to Ground Truthing: An Example with Wetland Plant Identification*; Engineer Research and Development Center (U.S.): Vicksburg, MS, USA, 2021.
99. Szantoi, Z.; Smith, S.E.; Strona, G.; Koh, L.P.; Wich, S.A.; Szantoi, Z.; Smith, S.E.; Strona, G.; Koh, L.P.; Serge, A. Mapping Orangutan Habitat and Agricultural Areas Using Landsat OLI Imagery Augmented with Unmanned Aircraft System Aerial Photography. *Int. J. Remote Sens.* **2017**, *38*, 2231–2245. [CrossRef]
100. Wich, S.A.; Koh, L.P. *Conservation Drones: Mapping and Monitoring Biodiversity*; Oxford University Press: Oxford, UK, 2018; pp. 51–54.
101. Onishi, M.; Ise, T. Explainable Identification and Mapping of Trees Using UAV RGB Image and Deep Learning. *Sci. Rep.* **2021**, *11*, 903. [CrossRef]
102. Gigante, D.; Attorre, F.; Venanzoni, R.; Acosta, A.T.R.; Agrillo, E.; Aleffi, M.; Alessi, N.; Allegrezza, M.; Angelini, P.; Angiolini, C.; et al. A Methodological Protocol for Annex I Habitats Monitoring: The Contribution of Vegetation Science. *Plant Sociol.* **2016**, *53*, 77–87. [CrossRef]

103. Correll, M.D.; Hantson, W.; Hodgman, T.P.; Cline, B.B.; Elphick, C.S.; Gregory Shriver, W.; Tymkiw, E.L.; Olsen, B.J. Fine-Scale Mapping of Coastal Plant Communities in the Northeastern USA. *Wetlands* **2019**, *39*, 17–28. [CrossRef]
104. Epifanio, I.; Ventura-Campos, N. Hippocampal Shape Analysis in Alzheimer's Disease Using Functional Data Analysis. *Stat. Med.* **2014**, *33*, 867–880. [CrossRef] [PubMed]
105. Ramsay, J.O.; Silverman, B.W. *Applied Functional Data Analysis: Methods and Case Studies*; Ramsay, J.O., Silverman, B.W., Eds.; Springer Series in Statistics; Springer: New York, NY, USA, 2002; Volume 45, ISBN 978-0-387-95414-1.