Criteria Comparison for Classifying Peatland Vegetation Types Using In Situ Hyperspectral Measurements

Thierry Erudel, Sophie Fabre, Thomas Houet, Florence Mazier, Xavier Briottet

MDPI

Article

Criteria Comparison for Classifying Peatland Vegetation Types Using *in situ* Hyperspectral Measurements

Thierry Erudel [1,2,3,4]*, Sophie Fabre [4], Thomas Houet [5], Florence Mazier [3] and Xavier Briottet [4]

[1] LabEx DRIIHM (Programme "investissements d'avenir" : ANR-11-LABX-0010), INEE-CNRS 3 rue Michel-Ange, 75016 Paris, France
[2] Université de Toulouse, Institut Supérieur de l'Aéronautique et de l'Espace (ISAE), Toulouse 31055, France
[3] GEODE UMR 5602 CNRS, Université Toulouse Jean Jaurès, 5 allées Antonio Machado, 31058 Toulouse Cedex 1, France; florence.mazier@univ-tlse2.fr
[4] ONERA, Optics and Associated Techniques Department, 2 avenue Edouard Belin, 31005 Toulouse Cedex, France; Sophie.Fabre@onera.fr, Xavier.Briottet@onera.fr
[5] LETG-Rennes UMR 6554 CNRS, Université Rennes 2, Place du recteur Henri le Moal, 35043 Rennes Cedex, France; thomas.houet@univ-rennes2.fr
* Correspondence: Thierry.Erudel@onera.fr

Academic Editor: name
Version July 13, 2017 submitted to Remote Sens.

Abstract: This study aims at evaluating three classes of methods to discriminate 13 peatland vegetation types using reflectance data. These vegetation types were empirically defined according to their composition, strata and biodiversity richness. On one hand, it is assumed that same vegetation type spectral signatures have similarities. Consequently they can be compared to a reference spectral database. To catch those similarities, several similarities criteria (related to distances (Euclidean distance, Manhattan distance, Canberra distance) or spectral shapes (Spectral Angle Mapper) or probabilistic behaviour (Spectral Information Divergence) and several mathematical transformations of spectral signatures enhancing absorption features (such as the first derivative or the second derivative, the normalized spectral signature, the continuum removal, the continuum removal derivative reflectance, the log transformation) were investigated. Furthermore those similarity measures were applied on spectral ranges which characterize specific biophysical properties. On the other hand, we suppose that specific biophysical properties/components may help to discriminate vegetation types applying supervised classification such as Random Forest (RF), Support Vector Machines (SVM), Regularized Logistic Regression (RLR), Partial Least Squares-Discriminant Analysis (PLS-DA). Biophysical components can be used in a local way considering vegetation spectral indices or in a global way considering spectral ranges and transformed spectral signatures as explained above. RLR classifier applied on spectral vegetation indices (training size = 25 %) was able to achieve 77.21 % overall accuracy in discriminating peatland vegetation types. It was also able to discriminate 83.95 % vegetation types considering specific spectral range [350–1350 nm], first derivative of spectral signatures and training size = 25 %. Conversely, similarity criterion was able to achieve 81.70 % overall accuracy using the Canberra distance computed on the full spectral range [350–2500 nm]. The results of this study suggest that RLR classifier and similarity criteria are promising to map the different vegetation types with high ecological values despite vegetation heterogeneity and mixture.

Keywords: Biodiversity; peatland; vegetation type; classification; hyperspectral; *in situ* measurements

1. Introduction

Peatlands represent a diverse array of wetlands that accumulate partially decomposed organic material. Whilst they may only cover a small proportion ($\sim$ 3 %) of the Earth's land surface, these ecosystems are highly

important in terms of functional and ecological values. Indeed, undisturbed, global peatland systems act as net atmospheric carbon sinks, storing approximately a third of the world's soil organic carbon [1], the vast majority of which (450–547 GtC (Gigatons of Carbon)) is held in northern peatlands (those above 45°N [2]). From an ecological perspective, these environments also provide important habitats for a number of rare plant and animal species [3].

Traditionally, species discrimination for floristic mapping needs intensive field work, including taxonomical information and the visual estimation of percentage cover for each species which are costly and time-consuming and sometimes inapplicable due to their poor accessibility [4]. Remote sensing is a technique that gathers data regularly about the earth's features. The main advantages that make remote sensing preferable to field-based methods in land cover classification, are that it has repeat coverage potential, allowing continuous monitoring, and its digital data can be easily integrated into a geographic information system (GIS) for more analysis which is less costly and less time-consuming [5,6].

Historically, aerial photography was the first remote sensing method to be employed for mapping wetland vegetation [7]. Currently, a variety of remotely sensed images are available for mapping wetland vegetation thanks to of airborne and space-borne vectors with multi-spectral sensors or hyperspectral sensors which operate within the different optical spectrum [8].

Mapping and monitoring wetlands (and even though peatland) floristic diversity is really challenging. Indeed, both temporal and spatial resolutions of remotely sensed imageries and *in situ* plant diversity and mixing contribute to the limitation of such techniques. Wetland plants are not as easily detectable as terrestrial plants since herbaceous wetland vegetations exhibits high spectral and spatial variabilities because of its steep environmental gradients [5,8]. Besides, the reflectance spectra of wetland vegetation canopies are often very similar and can be combined with reflectance spectra of the underlying soil, hydrologic regime and atmospheric vapour [9,10].

However, plant species have been successfully classified in estuarine [11], palustrine [12] and riparian habitats [13], as well in saltmarsh [5], in mangrove [14,15], in swamp [16] but not in peatlands, to our knowledge. Peatland mapping faces two great challenges at local and global scales due to their high environmental function (biodiversity hotspot, greenhouse gas fluxes, etc.): characterizing their internal diversity [8] and delineating their extent [17]. This study focuses on the first challenge for which only high - spectral or spatial - resolution imageries appear appropriate (see for instance [18–20]).

Plant species classification can benefit from several existing and recent techniques commonly used in remote sensing. Two main methods are applied for vegetation discrimination: the similarity measurement techniques and the supervised classification methods with sometimes application of a preliminary spectral band reduction technique. On one hand, similarity measures enable to discriminate similar classes from a set of spectra, extracted from images or acquired on the field. Some spectral measures, such as the Spectral Angle Mapper (SAM) are related to the difference of the spectral shape (*e.g.* [21] identified forests of the Liege oaks from other forests, grain crops and steppes using the multispectral Advanced Very High Resolution Radiometer (AVHRR) with five bands from 580 nm to 1250 nm, 1 km spatial resolution (Overall Accuracy (OA) = 94.10 %, $\kappa = 0.93$); [22] discriminated tree species using the multispectral Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) sensor with 9 spectral bands from 520 nm to 2430 nm and a spatial resolution of 15 m or 30 m ($\kappa = 0.66$)). Other spectral measures, such as the Spectral Information Divergence (SID) are related to probabilistic behaviour (*e.g.* [23] classified different tree species at leaf and vegetation cover scales using the hyperspectral HyMap sensor: 126 spectral bands from 436 nm to 2485 nm and a spatial resolution of 4 m (OA = 91.10 %, $\kappa = 0.87$)). On the other hand, the supervised classification methods may contribute as well to discriminate (group of) spectral signatures for plant species discrimination. The Linear Discriminant Analysis (LDA) is a method assuming that independent variables are normally distributed and which attempts to look for linear combination of variables to model the difference between the classes of the data (*e.g.* [24] succeeded in classifying different tree species at leaf and vegetation cover scales using the HYperspectral Digital Imagery Collection Experiment (HYDICE) sensor with 210 spectral bands from 400 nm to 2500 nm, 1.6 m spatial resolution (OA = 86 % using an object-based approach)). The Random Forest is an ensemble learning method

based on the construction of multiple decision trees (*e.g.* [25] succeeded in mapping invasive plants using the hyperspectral Probe-1 sensor: 128 bands from 450 nm to 2507 nm, 5 m spatial resolution (OA = 86 % for the leafy spurge classification)). The Support Vector Machines (SVM) is a classifier that looks for the best separating hyperplane (*e.g.* [26] succeeded in classifying different tree species in boreal forest using HySpex VNIR-1600-instrument: 160 spectral bands ranging from 410 nm to 990 nm , with a spatial resolution of 0.4 m (OA = 79.2 %); [27] classified successfully tropical vegetation using the Hyperion (EO-1) sensor (OA = 80 %)). The Regularized Logistic Regression (RLR) is the combination of a linear model (logistic regression) and a regularization term. It is usually used for feature selection (*e.g.* [28] applied it to reduce the 64 spectral bands from the hyperspectral AisaEAGLE II sensor to classify tree species in boreal forest using SVM; [29] applied it for reducing the 79 bands from the hyperspectral Digital Airborne Imaging Spectrometer (DAIS) sensor and the 220 bands from the hyperspectral Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor to classify different land covers using SVM) is investigated in this paper as a classifier.

Discriminating and classifying plant species can be done. Firstly, using different techniques hyperspectral measurements thanks to a portable spectroradiometer (FieldSpec Pro FR, Analytical Spectral Devices – ASD) which ranges on the reflective domain ([350–2500 nm] with a spectral resolution of 3 nm in Visible and Near InfraRed (VNIR) and approximatively 10 nm in the ShortWave InfraRed (SWIR)) either on laboratory [14] or immediately after the leaf was cut using the leaf clip accessory [16]. This can be an indicator of the ability of discriminating plant species using specific wavelengths or evaluating the performance of a classifier. Then, the wetlands heterogeneity mixing vegetation types can be catched still using a portable spectroradiometer: [12] used the ASD spectroradiometer, Ground Field of View (GFOV) = 0.43 m; [5] used the GER 3700 (Geophysical and Environmental Research Corporation) which ranges from 350 nm to 2509 nm) with a spectral resolution of 2 nm below 1000 nm and from 6 to 10 nm beyond 1000 nm, GFOV = 0.13 m. Secondly, with airborne imageries, hyperspectral sensors (SOC-700: 120 spectral bands between 394 and 890 nm with a 4 nm bandwidth and a spatial resolution of 0.5 m and a spatial resolution of 3 m [13]; HyMap: 128 bands in the visible and near infrared (VNIR: 0.45–1.50 µm with a 10 nm bandwidth) through the shortwave infrared (SWIR: 1.50–2.50 µm with a 15–20 nm bandwidth [11]). Thirdly, with spaceborne imageries using hyperspectral sensors (Hyperion: 242 spectral bands from 357 to 2756 nm with a spectral interval of 10 nm and a spatial resolution of 30 m [15]) or multispectral sensors (SPOT-5: 4 bands with 10 m resolution [15]) can be used to map wetlands.

This study aims at inventorying and evaluating the performance of discrimination techniques for peatland habitats based on *in situ* spectra. These habitats are characterized by more or less homogeneous vegetation mixing and have been chosen because of their ecological values (*i.e.* biodiversity). As defined by [30], mapping these habitats is therefore important to identify potential and/or effective areas with (at least) a floristic biodiversity function. For instance, we do not aim at detecting *Drosera rotundifolia* but at mapping the habitat favorable to this species (*Sphagnum* ...). Similarity measures and classifiers were applied on spectral signatures and some of their transformations (first and second derivatives, continuum removal, first derivative of continuum removal, normalized spectral signatures, log transformation). These transformations have been chosen because they enhance biophysical components which may help to distinguish plant species. These techniques were applied on different spectral ranges that either characterize specific biophysical components [31]. Classifiers were applied on spectral vegetation indices, characterizing specific biophysical components such as chlorophyll, pigments, nitrogen, cellulose, water.

This paper is organized as follows. After presenting the study site located in the Pyrenees (France) and associated data collection in Section 2, the methodology is detailed in Section 3. Then Section 4 presents and discussed the results of the different classifications that are suitable for distinguishing vegetation types. Finally, in Section 5, the conclusion summarizes main results and some perspectives are arisen for applying these techniques to hyperspectral imageries.

2. Material

2.1. Study site

The study site is the Bernadouze peatbog (Latitude: 42°47′ N , Longitude: 1°24′ E ; approximatively 2 ha), which is part of Human-Nature Observatory "Haut-Vicdessos" located in Ariège (Pyrénées, France) (Figure 1) and supported by the French CNRS and the LabEx DRIIHM. It is a long term monitored study site where hydrological, climatological, botanical, archeological, remotely sensed surveys are regularly conducted.

2.2. Field data collection

In this study, thirteen vegetation units with ecological values and potentials (*i.e.* biodiversity) have been identified in the Bernadouze peatbog. These units are named hereafter "vegetation types" according to the dominant land cover type or to the potential development of interesting plant species which may have ecological values (Table 1). For each type, several locations have been surveyed to characterize their plant species composition (Table A.1).

For all these 32 sample locations (Figure 1), radiances are measured at three different dates over 9 days in September 2014 (09/04/2014, 09/05/2014, 09/12/2014) under sunny and cloudless conditions between 10:00 a.m. and 1:00 p.m. and Sun's azimuth angle ranging from 106° and 160°. Data have been collected using an Analytical Spectral Device (ASD) spectroradiometer which ranges on the reflective domain (350–2500 nm) with a 3–12 nm spectral resolution depending on the spectral domain. Its spectral specifications are summarized in Table 2.

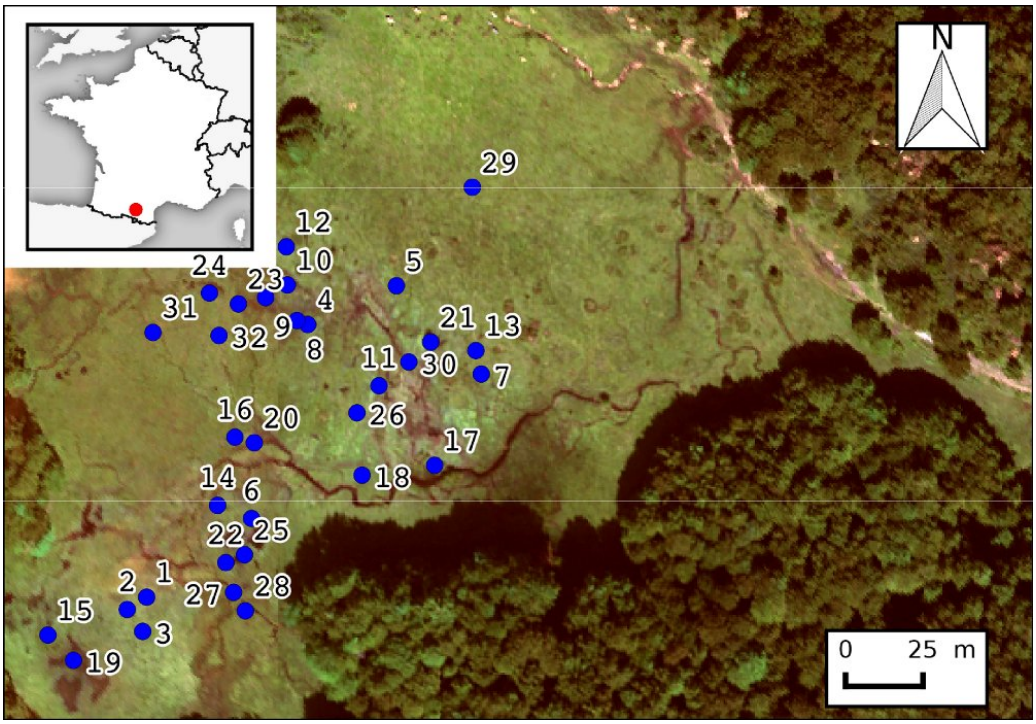


Figure 1. Location of the *in situ* spectroradiometer measurements – True color composite made from hyperspectral (HySpex) aerial imageries acquired on the 09/12/2014 (R = 639.98 nm, G = 549.06 nm, B = 461.79 nm).

To measure the reflectance of a sample plot (ρ) the reflectance of a white reference (ρ_{ref}) is required. This latter was obtained with a Spectralon (Labsphere, North Sutton, NH, USA) panel. Finally, after dark current correction, ρ is given by:

Table 1. Species names, number of measurements, number of locations and total number of spectra collected.

Vegetation types	Code	Measurements			No. of locations	No. of spectra
		09/04/2014	09/05/2014	09/12/2014		
Calluna vulgaris	CAVU		2	2	2	14
Sphagnum sp.	SPHA		2	4	5	22
Eleocharis quinqueflora	ELQU	1	2	1	2	15
Pinguicula sp.	PING	1	1		1	8
Menyanthes trifoliata	METR	1	1	1	1	12
Juniperus communis	JUCO	1	2	2	2	19
Rhododendron ferrugineum	RHFR		2	2	2	14
Salix sp.	SALI	1		3	3	17
Aquatic environment a	AQ_A	3	6	7	6	53
Aquatic environment b	AQ_B		1	1	1	7
Aquatic environment c	AQ_C	1	1	1	1	12
Carex sp. homogeneous vegetation	CA_HV	2	2	3	4	26
Pinguicula sp. combined vegetation	PI_CV	1	2	1	2	15

$$\rho = \frac{L_{\text{sam}}}{L_{\text{ref}}}\rho_{\text{ref}}, \tag{1}$$

where L_{sam} is the measured radiance from the sample plot and L_{ref} is the measured radiance from the white reference.

The sensor was positioned approximatively 1 m over the target with a 10° field of view. Consequently the ground spatial resolution is 0.18 m. The ASD was configured to collect 20 samples and automatically average in order to provide a single mean spectral measurement. Then a total of 7 to 53 field spectroradiometer measurements, *i.e.* spectral signatures, depending on vegetation type was taken.

Table 2. ASD FieldSpec Pro specifications.

	Spectral range	**Spectral resolution**	**Spectral sampling**
VNIR (Visible and Near InfraRed)	0.35 µm − 1.00 µm	3.00 nm at 0.70 µm	1.40 nm (0.35 µm − 1.05 µm)
SWIR (Short Wave InfraRed)	1.00 µm − 2.50 µm	10 nm at 1.40 µm 12 nm at 2.10 µm	2.00 nm (1.05 µm − 2.50 µm)

2.3. Data preprocessing

Some spectral bands (1350 nm to 1450 nm, 1810 nm to 1940 nm and 2400 nm to 2500 nm) have been removed due to a small signal-to-noise ratio resulting from strong atmospheric absorption mainly due to the presence of water vapour. More precisely, if the atmospheric transmittance value of the U.S. Standard profile was lower than 0.8 for a given wavelength, this wavelength was not taken into account in the analyse. Thus, each measured spectrum has been smoothed using a Savitzky-Golay filter [32] for reducing the noise. Figure 2 graphs the mean spectral reflectance of each vegetation type and the atmospheric transmittance. For the sake of clarity, standard deviation of each vegetation type is not printed on Figure 2 but can be seen in Appendix B.

3. Method description

The flowchart to evaluate the potential of hyperspectral data to discriminate and classify wetland vegetation types is given in Figure 3. More precisely, three classes of methods have been investigated and compared:

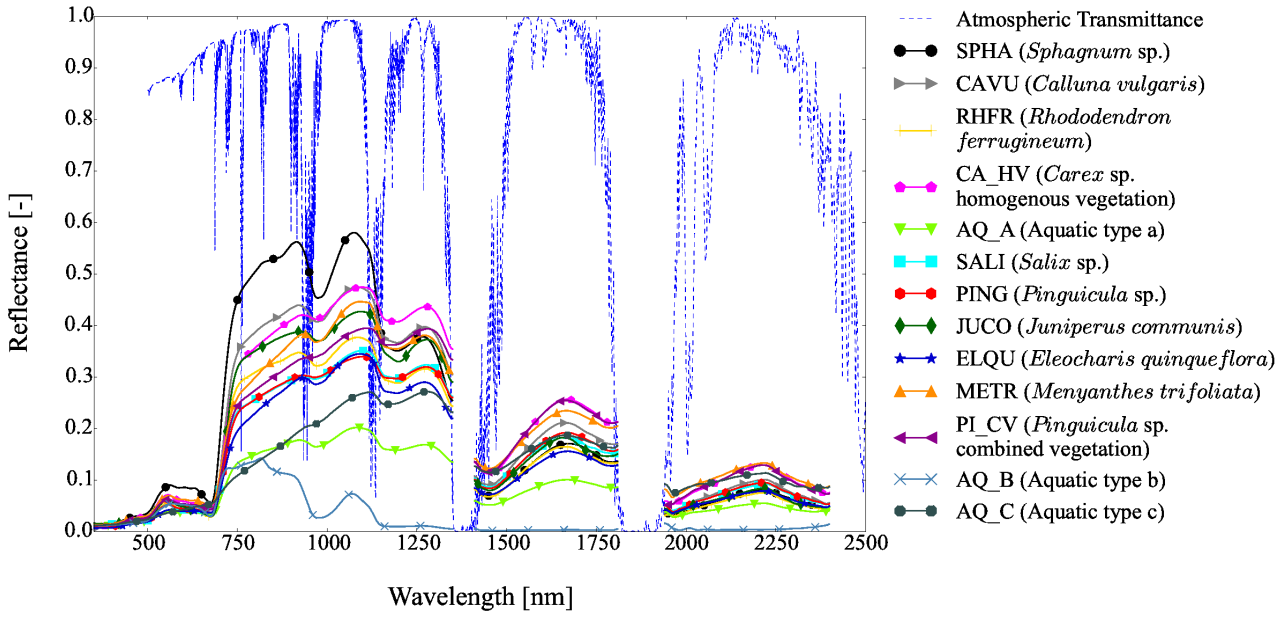


Figure 2. Mean spectral reflectances of the 13 vegetation types and the U.S. Standard atmospheric transmittance.

- similarity measures calculated on spectral reflectance,
- supervised classification based on "local" information (spectral vegetation indices),
- supervised classification based on "global" information (spectral ranges).

Indeed, spectral matching can be used to discriminate different vegetation types, because it is assumed that the spectral signatures of a given vegetation type must have similarities. To catch those similarities, several mathematical transformations – enhancing absorption features are applied on spectral signatures – (Section 3.1) and several similarity criteria – related to distances or spectral shapes or probabilistic behaviour – (Section 3.2) are investigated. Furthermore those similarity measures are applied on several spectral ranges which characterize specific biophysical properties (Section 3.5) and compared to a reference spectral database using relative spectral discriminatory probability (Section 3.3).

On the other hand as it may be difficult to have a spectral reference database, different supervised classifiers are used (Section 3.6). Besides, we assume that specific biophysical properties/components may help discriminating vegetation types. Biophysical components can be used in a local way considering spectral vegetation indices (Section 3.4) or in a global way considering spectral ranges and transformed spectral signatures as explained above.

To evaluate performance of similarity measures and supervised classification, the overall accuracy and F1-score are used (Section 3.7).

3.1. Transformed spectral signatures

As vegetation types are composed by a mix of various plant species that can be found in various vegetation types, different transformations are used (Table 3). Brightness-normalized spectral signature and second derivative are relatively insensible to variations in illumination intensity causes by changes in sun angle [33,34]. Other transformations (first derivative, second derivative, log transformation, Continuum Removal, Continuum Removed Derivative Reflectance (CRDR)) are linked to absorption features that may differ from one vegetation type to another, depending on the floristic composition.

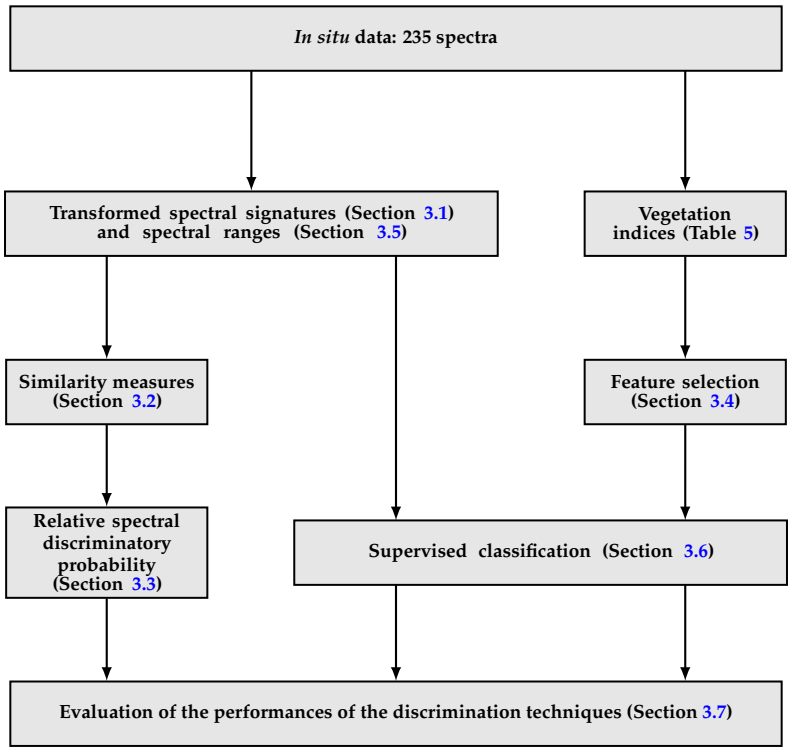


Figure 3. Flowchart showing the different methods used to classify the vegetation types.

Table 3. Transformed spectral signatures.

Transformation	Formulation	Reference
Brightness-normalized spectral signature	$\tilde{\rho}_{i,\lambda} = \frac{\rho_{i,\lambda}}{\left(\sum_{\lambda=1}^{L} \rho_{i,\lambda}^2\right)^{\frac{1}{2}}}$, $\forall \lambda \in [1, ..., L]$.	[33]
First derivative	$\left.\frac{\mathrm{d}\rho}{\mathrm{d}\lambda}\right\|_i \simeq \frac{\rho_{\lambda_j} - \rho_{\lambda_i}}{\Delta\lambda}$, where $\Delta\lambda$ is the separation between adjacent bands, $\Delta\lambda = \lambda_j - \lambda_i$ and $\lambda_j > \lambda_i$	[34]
Second derivative	$\left.\frac{\mathrm{d}^2\rho}{\mathrm{d}\lambda^2}\right\|_j = \frac{\mathrm{d}}{\mathrm{d}\lambda}\left(\frac{\mathrm{d}\rho}{\mathrm{d}\lambda}\right)\Big\|_j \simeq \frac{\rho_{\lambda_i} - 2\rho_{\lambda_j} + \rho_{\lambda_k}}{(\Delta\lambda)^2}$, where $\Delta\lambda = \lambda_k - \lambda_j = \lambda_j - \lambda_i, \lambda_k > \lambda_j > \lambda_i$.	[34]
log transformation or pseudo absorbance	$\tilde{\rho}_{i,\lambda} = \log\left(\frac{1}{\rho_{i,\lambda}}\right)$, $\forall \lambda \in [1, ..., L]$.	[35]
Continuum Removal	$CR_\lambda = \frac{\rho_\lambda}{C_\lambda}$, where C is a convex hull fitting over the top of the spectrum to connect local spectrum maxima.	[36,37]
Continuum removal derivative reflectance	$\left.\frac{\mathrm{d}CR_\lambda}{\mathrm{d}\lambda}\right\|_i$	[38]

L is the number of wavelengths.

3.2. Similarity measures

Let ρ_i be a spectral signature, $\rho_{i,\lambda}$ its reflectance at wavelength λ and $[1, ..., L]$ its spectral range. Several criteria have been used (Table 4). Some criteria characterize the difference between reflectance levels (like the distances) and other ones are related to the difference of the spectral shape (*e.g.* SAM) and other ones are related to probabilistic behaviour (*e.g.* SID, ...). Table 4 inventories main similarity measurement techniques described in the literature.

Table 4. Similarity measures.

Similarity measures	Formulation	Comments	Reference
Minkowski distance	$D_p(\rho_i,\rho_j) = \|\rho_i - \rho_j\|_p, = \left[\sum_{\lambda=1}^{L} \|\rho_{i,\lambda} - \rho_{j,\lambda}\|^p\right]^{1/p}.$	Spectral signatures are represented by vectors from $\mathbb{R}^L$. D_2 is the usual Euclidean distance ; D_1 is the Manhattan or City Block distance	D_2: [24,39,40] ; D_1: [41,42]
Canberra distance	$D_{\text{Canberra}}(\rho_i,\rho_j) = \sum_{\lambda=1}^{L} \frac{\|\rho_{i,\lambda} - \rho_{j,\lambda}\|}{\|\rho_{i,\lambda}\| + \|\rho_{j,\lambda}\|}.$	It is a weighted version of the Manhattan distance	[43]
Spectral Angle Mapper (SAM)	$\text{SAM}(\rho_i,\rho_j) = \cos^{-1}\left(\frac{\sum_{\lambda=1}^{L}\rho_{i,\lambda}\rho_{j,\lambda}}{\left(\sum_{\lambda=1}^{L}\rho_{i,\lambda}^2\right)^{1/2}\left(\sum_{\lambda=1}^{L}\rho_{j,\lambda}^2\right)^{1/2}}\right).$	Since the angle between two vectors is invariant with respect to the length of the vectors, this technique is relatively insensitive to illumination and albedo effects	[23,44]
Spectral Information Divergence (SID)	$SID(\rho_i,\rho_j) = D(\rho_i\|\|\rho_j) + D(\rho_j\|\|\rho_i),$ where $D(\rho_i\|\|\rho_j) = \sum_{\lambda=1}^{L} p_\lambda \log\left(\frac{p_\lambda}{q_\lambda}\right)$, where $p_\kappa = \frac{\rho_{i,\kappa}}{\sum_{\lambda=1}^{L}\rho_{i,\lambda}}, q_\kappa = \frac{\rho_{j,\kappa}}{\sum_{\lambda=1}^{L}\rho_{j,\lambda}}$	It calculates the probabilistic behaviour between spectral signatures	[45]
SAM-SID	$\text{SID-Tan}(\rho_i,\rho_j) = \text{SID}(\rho_i,\rho_j) \times \tan\left(\text{SAM}(\rho_i,\rho_j)\right),$ $\text{SID-Sin}(\rho_i,\rho_j) = \text{SID}(\rho_i,\rho_j) \times \sin\left(\text{SAM}(\rho_i,\rho_j)\right).$	It is a combination of probability and geometry spaces that improves discrimination ability	[46]
Spectral Correlation Measure (SCM)	$SCM(\rho_i,\rho_j) = \frac{L\sum_{\lambda=1}^{L}\rho_{i,\lambda}\rho_{j,\lambda} - \sum_{\lambda=1}^{L}\rho_{i,\lambda}\sum_{\lambda=1}^{L}\rho_{j,\lambda}}{\left[L\sum_{\lambda=1}^{L}\rho_{i,\lambda}^2\left(\sum_{\lambda=1}^{L}\rho_{i,\lambda}\right)^2\right]^{1/2}\left[L\sum_{\lambda=1}^{L}\rho_{j,\lambda}^2\left(\sum_{\lambda=1}^{L}\rho_{j,\lambda}\right)^2\right]^{1/2}}.$	It is calculated as the correlation coefficient of the pixel and their respective spectral signatures	[47]
Pearson Correlation Coefficient (PCC)	$PCC(\rho_i,\rho_j) = \frac{\sum_{\lambda=1}^{L}(\rho_{i,\lambda}-\mu_i)(\rho_{j,\lambda}-\mu_j)}{\left[\sum_{\lambda=1}^{L}(\rho_{i,\lambda}-\mu_i)^2\right]^{1/2}\left[\sum_{\lambda=1}^{L}(\rho_{j,\lambda}-\mu_j)^2\right]^{1/2}}$, where μ_i is the mean of ρ_i.		
Spectral Similarity Value (SSV)	$SSV(\rho_i,\rho_j) = \sqrt{D_2(\rho_i,\rho_j)^2 + \left(1 - [PCC(\rho_i,\rho_j)]^2\right)^2}.$	Low value of SSV means high similarity and *vice versa*	[48]
Spectral Correlation Angle (SCA)	$SCA(\rho_i,\rho_j) = \cos^{-1}\left(\frac{1+PCC(\rho_i,\rho_j)}{2}\right).$	It is an improvement of SAM derived from PCC that has been shown to be able to distinguish between positive and negative correlations and to yield better estimates in some experiments	[49,50]
Spectral Gradient Angle (SGA)	$SGA(\rho_i,\rho_j) = SAM\left(\|\nabla\rho_i\|,\|\nabla\rho_j\|\right), = \cos^{-1}\left(\frac{\sum_{\lambda=1}^{L-1}\|\rho_{i,\lambda+1}-\rho_{i,\lambda}\|\|\rho_{j,\lambda+1}-\rho_{j,\lambda}\|}{\left[\sum_{\lambda=1}^{L-1}(\rho_{i,\lambda+1}-\rho_{i,\lambda})^2\right]^{1/2}\left[\sum_{\lambda=1}^{L-1}(\rho_{j,\lambda+1}-\rho_{j,\lambda})^2\right]^{1/2}}\right).$	It is invariant to illumination conditions	[51]

3.3. Relative spectral discriminatory probability

To determine if a spectral signature belongs to a class, the method proposed by [45] is used. Let $\{\rho_j\}_{j=1}^{J}$ J spectral signatures in Δ an existing spectral reference database and τ be a target signature to be identified using Δ. Let $m(\cdot,\cdot)$ be a given hyperspectral measure, the spectral discriminatory probabilities of all ρ_j in Δ with respect to τ as is defined as follows:

$$p^m_{\tau,\Delta}(i) = \frac{m(\tau,\rho_i)}{\sum_{j=1}^{J} m(\tau,\rho_j)}, \text{ for } i = 1, 2, ..., J, \tag{2}$$

where $\sum_{j=1}^{J} m(\tau, \rho_j)$ is a normalization constant determined by τ and Δ. The resulting probability vector is defined as

$$\mathbf{p}^m_{\tau,\Delta} = \left(p^m_{\tau,\Delta}(1), p^m_{\tau,\Delta}(2), ..., p^m_{\tau,\Delta}(J)\right)^T. \quad (3)$$

Using Equation (3), the target signature can be identified by selecting the one with the smallest spectral discriminatory probability because τ and the selected one have the minimum spectral discrimination.

Spectral reference database

To build the spectral reference database, spectra of mean reflectance, spectra of median reflectance and median spectra are used. Spectra of mean reflectance is defined as the mean of reflectances for each wavelength λ:

$$\overline{\rho_\lambda} = \frac{1}{N}\sum_{i=1}^{N} \rho_{i,\lambda}, \ \forall \lambda \in [1, ..., L], \quad (4)$$

where N is the number of spectra for a plant species. Similarly, spectra of median reflectance is defined as the median of reflectances for each wavelength λ. Median spectra is defined as the "closest" spectrum of the median reflectance considering a vegetation type. In other words, giving a spectrum of median reflectance, the spectrum that minimize the Minkowski distance between them is considered as the median spectrum (Figure 4 shows differences between the median reflectances spectrum which is an theoretic spectral signature and the different median spectra which were investigated). As distances are not equivalent considering high-dimensional data, three Minkowski distances are investigated for this study: the Euclidean distance, the Canberra distance and the City Block or Manhattan distance (which are reminded in Section 3.1).

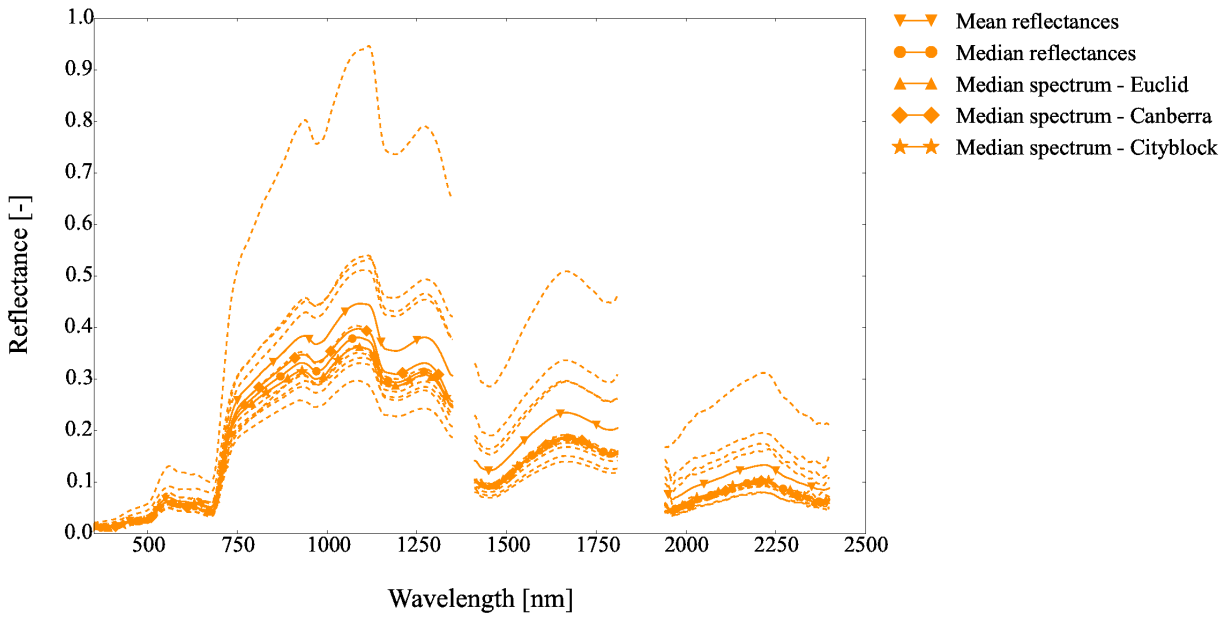


Figure 4. Median spectra, spectrum of mean reflectances, spectrum of median reflectances of *Eleocharis quinqueflora* (ELQU).

3.4. Feature selection of spectral indices

Spectral index description

Spectral indices are combinations of surface reflectance (or the derivated reflectance) at two or more wavelengths or narrow spectral bands. Lots of spectral indices can be found in literature (Table 5) to characterize some biochemical components of plant species such as chlorophyll, nitrogen, lignin, cellulose, water. Although these indices have never been selected in the literature to characterize wetlands plant species, we assume that some of them can still be useful to classify them.

Table 5. Spectral vegetation indices.

Index name	Formulation	Vegetation properties highlighted by the index	Reference
Boochs	D_{703}	Chlorophyll	[52]
Boochs2	D_{720}	Chlorophyll	
CAI (Cellulose Absorption Index)	$\frac{R_{2000}+R_{2200}}{2} - R_{2100}$	Cellulose, soil litter	[53]
CARI(Chlorophyll Absorption Ratio Index)	$\frac{R_{700}\sqrt{(670a+R_{670}+b)^2}}{R_{670}\sqrt{a^2+1}}$ where $a = \frac{R_{700}-R_{550}}{150}$; $b = R_{550} - 550a$	Chlorophyll	[54]
CI (Curvature Index)	$\frac{R_{675}R_{690}}{R_{683}^2}$	Chlorophyll	[55]
CCI (Canopy Chlorophyll Index)	$\frac{D_{720}}{D_{700}}$	Chlorophyll	[56]
CCCI (Canopy Chlorophyll Content Index)	$\frac{\text{NDRE}}{\text{NDVI}}$	Chlorophyll	[57]
Carter[695,420]	$\frac{R_{695}}{R_{420}}$	Stress	[58]
Carter[695,760]	$\frac{R_{695}}{R_{760}}$	Stress	
Carter[605,760]	$\frac{R_{605}}{R_{760}}$	Stress	
Carter[710,760]	$\frac{R_{710}}{R_{760}}$	Stress	
Carter[695,670]	$\frac{R_{695}}{R_{670}}$	Stress	
Carter2	R_{550}		
CaCoI[515,550] (Carotenoid Concentration Index)	$\frac{1}{R_{515}} - \frac{1}{R_{550}}$	Carotenoid	[59], [60]
CaCoI[515,700]	$\frac{1}{R_{515}} - \frac{1}{R_{700}}$	Carotenoid	
CaCoI2[770,510,700]	$R_{770}\left(\frac{1}{R_{510}} - \frac{1}{R_{700}}\right)$	Carotenoid	[59], [60]
CaCoI2[770,510,550]	$R_{770}\left(\frac{1}{R_{510}} - \frac{1}{R_{550}}\right)$	Carotenoid	
Datt[850]	$\frac{R_{850}-R_{710}}{R_{850}-R_{680}}$	Chlorophyll	[61]
Datt[780]	$\frac{R_{780}-R_{710}}{R_{780}-R_{680}}$	Chlorophyll	[61]
Datt2[850,710]	$\frac{R_{850}}{R_{710}}$	Chlorophyll	
Datt2[672,550]	$\frac{R_{672}}{R_{550}}$	Chlorophyll	
Datt_prime	$\frac{D_{754}}{D_{704}}$	Chlorophyll	
Datt3[672]	$\frac{R_{672}}{R_{550}R_{708}}$	Chlorophyll	[62]
Datt3[860]	$\frac{R_{860}}{R_{550}R_{708}}$	Chlorophyll	[62]
DCI	$\frac{D_{723}}{D_{703}}$		[63]

Table 5: continued from previous page.

DCNI (Double-peak Canopy Nitrogen Index)	$\frac{R_{720}-R_{700}}{(R_{700}-R_{670})(R_{720}-R_{670}+0.03)}$	Nitrogen	[64]
DD (Double Difference Index)	$(R_{749}-R_{720})-(R_{701}-R_{672})$	Chlorophyll	[65]
DDn (new Double Difference Index)	$2\Big(R_{710}-R_{(710-50)}-R_{(710+50)}\Big)$	Chlorophyll	[66]
DPI (Double Peak Index)	$\frac{D_{688}D_{710}}{D_{967}^2}$	Chlorophyll	[55]
dG	$\max_{i\in[\![500,580]\!]} D_i$	Chlorophyll, stress	
dRE	$\max_{i\in[\![680,750]\!]} D_i$	Chlorophyll, stress	[67]
D[730,706]	$\frac{D_{730}}{D_{706}}$	Chlorophyll	[55]
D[705,722]	$\frac{D_{705}}{D_{722}}$		
EVI (Enhanced Vegetation Index)	$2.5\frac{R_{800}-R_{670}}{R_{800}-6R_{670}-7.5R_{475}+1}$	Chlorophyll	[68]
EGFR (Edge-Green First derivative Ratio)	$\frac{\mathrm{dRE}}{\mathrm{dG}}$	Chlorophyll, nitrogen	[69]
EGFN (Edge-Green first Derivative Normalized difference)	$\frac{\mathrm{dRE}-\mathrm{dG}}{\mathrm{dRE}+\mathrm{dG}}$	Chlorophyll, nitrogen	
GEMI (Global Environment Monitoring Index)	$\eta(1-0.25\eta)-\frac{R_{660}-0.25}{1-R_{660}}$, where $\eta=2\frac{R_{830}^2-R_{660}^2+1.5R_{830}+0.5R_{660}}{R_{830}+R_{660}+0.5}$		[70]
GI (Greeness Index)	$\frac{R_{554}}{R_{677}}$	Chlorophyll	[71]
Gitelson	$\frac{1}{R_{700}}$	Chlorophyll	[72]
Gitelson2	$\frac{R_{750}-R_{800}}{R_{965}-R_{740}}-1$	Chlorophyll	[59]
GMI (Gitelson and Merzlyak Index)	$\frac{R_{750}}{R_{550}}$	Chlorophyll	[73]
Green NDVI	$\frac{R_{800}-R_{550}}{R_{800}+R_{550}}$	Chlorophyll	[74]
Maccioni	$\frac{R_{780}-R_{710}}{R_{780}-R_{680}}$	Chlorophyll	[75]
MARI (Modified Anthocyanin Reflectance Index)	$R_{800}\left(\frac{1}{R_{550}}-\frac{1}{R_{700}}\right)$	Anthocyanin	[76], [77]
MCARI[700,670] (Modified Chlorophyll Absorption Index)	$\Big((R_{700}-R_{670})-0.2(R_{700}-R_{550})\Big)\frac{R_{700}}{R_{670}}$	Chlorophyll, Leaf Area Index	[78]
MCARI[750,705]	$\Big((R_{750}-R_{705})-0.2(R_{750}-R_{550})\Big)\frac{R_{750}}{R_{705}}$	Chlorophyll	[79]
MCARI[700,670]/OSAVI[800,670]	$\frac{\Big((R_{700}-R_{670})-0.2(R_{700}-R_{550})\Big)\frac{R_{700}}{R_{670}}}{(1+0.16)\frac{R_{800}-R_{670}}{R_{800}+R_{670}+0.16}}$	Chlorophyll	[80]
MCARI[750,705]/OSAVI[750,705]	$\frac{\Big(R_{750}-R_{705})-0.2(R_{750}-R_{550})\Big)\frac{R_{750}}{R_{705}}}{(1+0.16)\frac{R_{750}-R_{705}}{R_{750}+R_{705}+0.16}}$	Chlorophyll	[79]
MCARI[750,705]/MTVI2[750]	$\frac{\mathrm{MCARI[750,705]}}{\mathrm{MTVI2[750]}}$	Nitrogen	[81]
MNDVI[800,680] (Modified NDVI)	$\frac{R_{800}-R_{680}}{R_{800}+R_{680}-2R_{445}}$	Chlorophyll	[82]
MNDVI[750,705]	$\frac{R_{750}-R_{705}}{R_{750}+R_{705}-2R_{445}}$	Chlorophyll	
MSAVI (Modified Soil Adjusted Vegetation Index)	$0.5\Big(2R_{800}+1-\sqrt{(2R_{800}+1)^2-8(R_{800}-R_{670})}\Big)$	Chlorophyll	[83]
MSI (Moisture Stress Index)	$\frac{R_{1599}}{R_{819}}$	Water stress	[84]
MSR[800,680] (modified Simple Ratio)	$\frac{R_{800}-R_{445}}{R_{680}-R_{445}}$	Chlorophyll	[82]
MSR[750,705]	$\frac{R_{750}-R_{445}}{R_{705}-R_{445}}$	Chlorophyll	
MSR2	$\frac{R_{750}}{R_{705}}-\frac{1}{\sqrt{\frac{R_{750}}{R_{705}}+1}}$	Chlorophyll, Leaf Area Index	[85]

Table 5: continued from previous page.

MTCI (MERIS [1] Terrestrial Chlorophyll Index)	$\frac{R_{754}-R_{709}}{R_{709}-R_{681}}$	Chlorophyll	[86]
MTVI[800] (Modified Triangular Vegetation Index)	$1.5\Big(1.2(R_{800}-R_{550})-2.5(R_{670}-R_{550})\Big)$	Leaf Area Index	[87]
MTVI[750]	$1.5\Big(1.2(R_{750}-R_{550})-2.5(R_{670}-R_{550})\Big)$	Leaf Area Index	[87]
MTVI2 [800]	$\frac{1.5\Big(1.2(R_{800}-R_{550})-2.5(R_{670}-R_{550})\Big)}{\sqrt{(2R_{800}+1)^2-(6R_{800}-5\sqrt{R_{670}})-0.5}}$	Leaf Area Index	[87]
MTVI2 [750]	$\frac{1.5\Big(1.2(R_{750}-R_{550})-2.5(R_{670}-R_{550})\Big)}{\sqrt{(2R_{750}+1)^2-(6R_{750}-5\sqrt{R_{670}})-0.5}}$		[87]
NDII (Normalized Difference Infrared Index)	$\frac{R_{850}-R_{1650}}{R_{850}+R_{1650}}$ $\frac{R_{819}-R_{1649}}{R_{819}+R_{1649}}$	Water status	[88]
NDLI (Normalized Difference Lignin Index)	$\frac{\log\left(\frac{1}{R_{1754}}\right)-\log\left(\frac{1}{R_{1680}}\right)}{\log\left(\frac{1}{R_{1754}}\right)+\log\left(\frac{1}{R_{1680}}\right)}$	Lignin	[35]
NDNI (Normalized Difference Nitrogen Index)	$\frac{\log\left(\frac{1}{R_{1510}}\right)-\log\left(\frac{1}{R_{1680}}\right)}{\log\left(\frac{1}{R_{1510}}\right)+\log\left(\frac{1}{R_{1680}}\right)}$	Nitrogen	[35]
NDRE (Normalized Difference Red Edge)	$\frac{R_{830}-R_{red}}{R_{830}+R_{red}}$, with $R_{red}=0.5(R_{670}+R_{780})$		[57]
NDVI[800,670] (Normalised Difference Vegetation Index)	$\frac{R_{800}-R_{670}}{R_{800}+R_{670}}$	Chlorophyll, Leaf Area Index	[89]
NDVI[750,705]	$\frac{R_{750}-R_{705}}{R_{750}+R_{705}}$	Chlorophyll	[73]
NDVI[682,553]	$\frac{R_{682}-R_{553}}{R_{682}+R_{553}}$	Chlorophyll	[90]
NDVI[573,440]	$\frac{R_{573}-R_{440}}{R_{573}+R_{440}}$	Nitrogen	[91]
NDWI[860,1240] (Normalized Difference Water Index)	$\frac{R_{860}-R_{1240}}{R_{860}+R_{1240}}$		
NDWI[860,1640]	$\frac{R_{860}-R_{1640}}{R_{860}+R_{1640}}$	Water status	[92]
NDWI[860,2130]	$\frac{R_{860}-R_{2130}}{R_{860}+R_{2130}}$		
NDWI[1100,1450]	$\frac{R_{1100}-R_{1450}}{R_{1100}+R_{1450}}$	Water stress	[93]
NDWI[1280,1450]	$\frac{R_{1280}-R_{1450}}{R_{1280}+R_{1450}}$	Water stress	[93]
NPCI (Normalised Pigment Chlorophyll Index)	$\frac{R_{680}-R_{430}}{R_{680}+R_{430}}$	(Total pigments) / chlorophyll	[94]
VI_opt (Vegetation Index optimal)	$(1+0.45)\frac{R_{800}^2+1}{R_{670}+0.45}$	Nitrogen	[95]
OSAVI[800,670] (Optimised Soil-Adjust Vegetation Index)	$(1+0.16)\frac{R_{800}-R_{670}}{R_{800}+R_{670}+0.16}$	Chlorophyll	[96]
OSAVI[750,705]	$(1+0.16)\frac{R_{750}-R_{705}}{R_{750}+R_{705}+0.16}$	Chlorophyll	[79]
PRI (Photochemical Reflectance Index)	$\frac{R_{531}-R_{570}}{R_{531}+R_{570}}$	Stress	[97]
RDVI (Renormalised Difference Vegetation Index)	$\frac{R_{800}-R_{670}}{\sqrt{R_{800}+R_{670}}}$	Chlorophyll, Leaf Area Index	[98]
REIP (Red-Edge Inflexion Point)	$\left\{i\middle\vert \max_{i\in[\![680,750]\!]} D_i\right\}$	Chlorophyll, Leaf Area Index	[67,99,100]
REMI (Red-Edge Model Index)	$\frac{R_{750}}{R_{720}}-1$	Chlorophyll	[101]

[1] MEdium Resolution Imaging Spectroradiometer

Table 5: continued from previous page.

REP_LE (Red-Edge Position Linear Extrapolation)	$-\frac{c_1 - c_2}{m_1 - m_2}$ where m_1 and c_1 represent the slope and the intercept of the far-red line and m_2 and c_2 represent the slope and the intercept of the NIR line	Nitrogen, chlorophyll	[102]
REP_LI (Red-Edge Position Linear Interpolation)	$700 + 40\frac{0.5(R_{800} + R_{780})}{R_{740} - R_{700}}$	Chlorophyll	[103]
RVI[810,660] (Ratio Vegetation Index)	$\frac{R_{810}}{R_{660}}$	Nitrogen	[104]
RVI[810,560]	$\frac{R_{810}}{R_{560}}$	Nitrogen	[105]
RVI[800,670]	$\frac{R_{800}}{R_{670}}$		
SIPI (Structure Insensitive Pigment Index)	$\frac{R_{800} - R_{445}}{R_{800} - R_{680}}$	Pigments / chlorophyll, stress	[106]
SPVI (Spectral Polygon Vegetation Index)	$0.4\left[3.7(R_{800} - R_{670}) - 1.2\sqrt{(R_{530} - R_{670})^2}\right]$	Chlorophyll × Leaf Area Index	[107]
SR[800,680] (Simple Ratio Index)	$\frac{R_{800}}{R_{680}}$	Chlorophyll	[108]
SR[750,700]	$\frac{R_{750}}{R_{700}}$		[73]
SR[752,690]	$\frac{R_{752}}{R_{690}}$		
SR[750,550]	$\frac{R_{750}}{R_{550}}$		
SR[700,670]	$\frac{R_{700}}{R_{670}}$	Chlorophyll	[109]
SR[675,700]	$\frac{R_{675}}{R_{700}}$	Chlorophyll	[110]
SR[750,710]	$\frac{R_{750}}{R_{710}}$	Chlorophyll	[111]
SR[440,690]	$\frac{R_{440}}{R_{690}}$	Stress	[112]
SRPI (Simple Ratio Pigment Index)	$\frac{R_{430}}{R_{680}}$	(Total pigments) / chlorophyll, stress	[106]
Sum_Dr[625,795]	$\sum_{i=625}^{795} D_i$	Chlorophyll	[113]
Sum_Dr[680,780]	$\sum_{i=680}^{780} D_i$	Chlorophyll, Leaf Area Index	[67]
TCARI[700,670] (Transformed Chlorophyll Absorption Ratio Index)	$3\left(R_{700} - R_{670} - 0.2(R_{700} - R_{550})\frac{R_{700}}{R_{670}}\right)$	Chlorophyll	[80]
TCARI[750,705]	$3\left(R_{750} - R_{705} - 0.2(R_{750} - R_{550})\frac{R_{750}}{R_{705}}\right)$	Chlorophyll	[79]
TCARI[700,670]/OSAVI[800,670]	$\frac{\text{TCARI}}{\text{OSAVI}}$	Chlorophyll	[80]
TCARI[750,705]/OSAVI[750,705]	$\frac{\text{TCARI2}}{\text{OSAVI2}}$	Chlorophyll	[79]
TVI (Triangular Vegetation Index)	$0.5(120(R_{750} - R_{550}) - 200(R_{670} - R_{550}))$	Leaf Area Index, Canopy chlorophyll density	[114]
Vogelmann	$\frac{R_{740}}{R_{720}}$	Chlorophyll	[115]
Vogelmann2	$\frac{R_{734} - R_{747}}{R_{715} + R_{726}}$	Chlorophyll	
Vogelmann3	$\frac{D_{715}}{D_{705}}$	Chlorophyll	
Maximum first derivatives of 8 different regions whithin the spectra	A_1D: 495–550 nm B_1D: 550–650 nm C_1D: 680–780 nm D_1D: 970–1090 nm E_1D: 1110–1205 nm F_1D: 1205–1285 nm H_1D: 1455–1640 nm	Pigments absorption, w., c., s., l absorption ; refer to Table 2 in [116] for a full description.	[116]

Table 5: continued from previous page.

Corresponding spectral positions of the maximum first derivatives	J_1D: 1925–2200 nm A_WP: 495–550 nm B_WP: 550–650 nm C_WP: 680–780 nm D_WP: 970–1090 nm E_WP: 1110–1205 nm F_WP: 1205–1285 nm H_WP: 1455–1640 nm J_WP: 1925–2200 nm	Pigments absorption, w., c., s., l. absorption ; refer to Table 2 in [116] for a full description.	[116]
WI (Water Index)	$\frac{R_{900}}{R_{970}}$	Water status	[117]
WI[1100,1450]	$\frac{R_{1100}}{R_{1450}}$	Water stress	[93]
WI[1280,1450]	$\frac{R_{1280}}{R_{1450}}$	Water stress	[93]
WI2	$\frac{1}{R_{1450}}$	Water stress	[93]

R_x represents reflectance at wavelength x nm.
D_x represents the derivative of the reflectance spectrum at wavelength x nm.
w., c., s., l = water, cellulose, starch, lignin

Classical feature selection method - the Kruskal-Wallis H-test

As some spectra per vegetation types were quite small (8 spectra for *Pinguicula* sp. (PING), 7 spectra for Aquatic type b (AQ_B)), usual ANOVA [118] test or Mann-Whitney U-test [119] can not be used. That is the reason why Kruskal-Wallis H-test [120], a non-parametric test is proposed. Moreover this test is adapted to not independent data and not normally distributed data. The H-test is used to test the hypothesis that there was no significant difference between the median spectral index value between pairs of plant species.

The null hypothesis for $N = 13$ vegetation types and $I = 129$ spectral vegetation indices per reflectance measurements is:

$$H_0 : \eta_n(i) = \eta_{n+1}(i), \tag{5}$$

where η_n is the median spectral index value for vegetation type number $n = 0, ..., N$, and $i = 1, ..., I$ the spectral index. The maximum frequency for this study is $\binom{13}{2} = \frac{13\times(13-1)}{2} = 78$. The hypothesis was therefore tested 78 times for all possible combinations of the 13 plant species at the adjusted Bonferroni significance level of $\alpha = \frac{0.05}{78} = 6.410^{-4}$.

Principle of the applied feature selection method

In order to discriminate the 78 pairs of vegetation types, the Hellinger distance, which is introduced further, is computed for each vegetation spectral index (Table 5). Then indices are ordered by frequency discrimination. A first subset of indices is composed of ones that can discriminate pairs of vegetation types and that are not redundant. If a pair of vegetation types is not discriminated, the Hellinger distance is computed for a pair of vegetation indices composed of the single most discriminating one and the other ones ordered by frequency distribution amongst previous selected. Then, a second subset of pairs of indices is composed by ordering those pairs of indices by frequency discrimination. To stop the process, a maximum number of subset is then defined. In our case, the maximum subset consists of not more than three indices. Indeed, longer is the tuple length, more difficult it is to explained why such combinations of indices or such biophysical components combination can discriminate such pairs. Finally, selected vegetation indices come from each subset and single spectral vegetation indices or spectral index combinations are retained.

For a better understanding of the feature selection method, an example is given. We consider 4 vegetation types named: V_1, V_2, V_3, V_4 and 5 spectral vegetation indices named: I_1, I_2, I_3, I_4, I_5. We suppose that no single

spectral vegetation index can discriminate neither V_1 from V_3 nor V_2 from V_4 nor V_3 from V_4. But different single indices can separate V_1 from V_2, V_1 from V_4 and V_2 from V_3. This is summarized in the following table:

	V_2	V_3	V_4
V_1	I_1, I_3	$\emptyset$	I_2, I_3
V_2	-	I_2, I_3	$\emptyset$
V_3	-	-	$\emptyset$

We obtain the first subset $S_1 = \{I_1, I_2, I_3\}$. To discriminate V_1 from V_3, V_2 from V_4 and V_3 from V_4, we are looking among the following combinations: $\{I_3 - I_2\}, \{I_3 - I_1\}, \{I_3 - I_4\}, \{I_3 - I_5\}$ because indices are ordered by frequency discrimination: $[I_3, I_2, I_1, I_4, I_5]$. We suppose that $\{I_3 - I_1\}$ can discriminate V_1 from V_3 and V_2 from V_4 but there is still no index that can discriminate V_3 from V_4. For the latter case, possible combinations are looking among $\{I_3 - I_1 - I_2\}, \{I_3 - I_1 - I_4\}, \{I_3 - I_1 - I_5\}$. Whatever a combination of spectral vegetation indices can be found to discriminate or not those plant species, the process will stop in our case.

The Bhattacharyya coefficient and the Hellinger distance

For two arbitrary discrete probability distributions $\mathbf{p}$ and $\mathbf{q}$, the amount of overlap between those distributions can be measured using the Bhattacharyya coefficient:

$$C(\mathbf{p}, \mathbf{q}) = \sum_{i=1}^{n} \sqrt{p_i q_i}, \tag{6}$$

where n is the partition number. To measure the similarity between two statistical distributions in remote sensing the Hellinger distance (also known as the Matusita distance) is commonly used. It is defined as:

$$H(\mathbf{p}, \mathbf{q}) = \sqrt{\frac{1}{2} \sum_{i=1}^{n} \left(\sqrt{p_i} - \sqrt{q_i}\right)^2}, \tag{7}$$

$$= \sqrt{1 - C(\mathbf{p}, \mathbf{q})}. \tag{8}$$

The Hellinger distance defined in Equation (8) has upper bound equal to 1, indicating the total separability of the class pairs characterized by their distribution. As a general rule adapted from [121],

- if $H(\mathbf{p}, \mathbf{q}) \geq 0.95$ then the classes can be separated,
- if $0.85 \leq H(\mathbf{p}, \mathbf{q}) < 0.95$ the separation is fairly good,
- if $H(\mathbf{p}, \mathbf{q}) < 0.85$ the separation is poor.

3.5. Spectral ranges

The transformed spectral signatures defined in Section 3.2 and the spectral ranges adapted from [31] (Table 6) were investigated:

- visible: 350 nm–750 nm,
- near infrared: 750 nm–1350 nm,
- shortwave infrared a: 1410 nm–1810 nm,
- shortwave infrared b: 1940 nm–2400 nm.

The shortwave infrared domain is split in 2 parts. The near infrared and the shortwave infrared are not continuous because of atmospheric water absorption.

3.6. Supervised classification

All the classifications are performed using Python scikit-learn package [129].

Table 6. The spectral reflectances of green vegetation on the four regions of electromagnetic spectrum from [31].

Wavelength range [nm]	Description	Spectral reflectance of vegetation	References
400-700	Visible	Low reflectance and transmittance due to chlorophyll and biologically active pigments (such as carotene) absorptions	[122,123]
680-750	Red-edge	The reflectance is strongly correlated with plant biochemical and biophysical parameters	[124,125]
700-1300	Near infrared	High reflectance and transmittance, very low absorption resulting from photon scattering at the air-cell interfaces within the leaf spongy mesophyll	[126,127]
1300-2500	Shortwave infrared	Lower reflectance than other spectral regions due to strong water absorption and minor absorption of biochemical contents such as lignin and carbon constituants	[126,128]

3.6.1. Random Forest (RF)

RF is an ensemble classifier that uses a set of Classification And Regression Trees (CARTs) to make a prediction [130]. The trees are created by drawing a subset of training samples through replacement (a bagging approach). In standard classification trees, each node is split using the best split among all variables. In RF, each node is split using the best predictor, among a user-defined number of features (*Mtry* that is usually set to the square root of the number of input variables [131]). By growing the forest up to a user-defined number of trees (*Ntree* that is usually set to 500 but different values such as 100, 1000 or 5000 have been investigated [131]), the algorithm creates trees that have high variance and low bias. The final classification decision is taken by averaging (using the arithmetic mean) the class assignment probabilities calculated by all produced trees.

For this study, $Mtry = 500$ and $Ntree \in [500, 1000, 2000, 5000]$.

3.6.2. Support Vector Machines (SVM)

SVM is a supervised non-parametric statistical learning technique therefore there is no assumption on the distribution of the data [132]. The main idea of SVM classification is to construct a hyperplane as a decision surface in a way that the margin of separation between two classes is maximized. To do this, the original feature space is mapped into a space with a higher dimensionality, where classes can be modelled to be linearly separable. This transformation is implicitly performed by applying kernel functions to the original data. The learning of the classifier is performed using a constrained optimization process that is associated with a complex cost function. For problems that involve identification of multiple classes, adjustments are made to the simple SVM binary classifier to operate as a multi-class classifier using methods such as one-against-all, one-against-others.

For this study, two kernels are retained: a linear kernel (SVM linear) and a Gaussian kernel (SVM RBF).

3.6.3. Regularized Logistic Regression (RLR)

RLR is a linear model based on logistic regression with an additional regularization term. This classifier has been successfully used with high dimensional data (gene selection in cancer classification [133], feature selection in remote sensing [28,29,134]).

For this study, the ℓ_1-norm and ℓ_2-norm regularization term are investigated.

3.6.4. Partial Least Squares-Discriminant analysis (PLS-DA)

PLS-DA is based upon the classical partial least square regression method for constructing predictive models [135]. The goal of PLS regression is to provide dimension reduction in an application where the response variable is related to the predictor variables. In the case of PLS-DA, the response variable (*i.e.* vegetation

types) is binary and expresses class membership [136,137]. This classifier has been successfully used with high dimensional data (gene selection [138], tree species discrimination [139]).

For this study the number of latent variables is fixed to the number of vegetation types - 1 [138]. This method is not applied on spectral vegetation indices selected but on spectral signatures and their transformations on spectral ranges because it is commonly used when the number of features is much bigger than the number of spectra.

3.7. Classification accuracy evaluation

To evaluate the classification accuracy of supervised classifiers, a 30 fold cross-validation is used and six training samples size were investigated: 50 %, 45 %, 40 %, 35 %, 30 % and 25 % of all spectra.

To evaluate the classifier precision overall accuracy and F1-score are used. Overall accuracy computes number of correct spectra over all spectra, whereas F1-score is given by:

$$\text{F1-score} = 2 \cdot \frac{\text{PA} \cdot \text{UA}}{\text{PA} + \text{UA}}, \tag{9}$$

where PA (Producer's Accuracy) is the fraction of retrieved classes that are relevant whereas UA (User's Accuracy) is the fraction of relevant classes that are retrieved.

4. Results and discussion

4.1. Similarity measures

Considering all transformed spectral signatures, spectral ranges and similarity measures, only the Canberra distance on [350 nm to 2500 nm] gives an overall accuracy higher than 50 % whatever the spectral reference database (Table 7). Indeed, the Canberra distance gives the higher overall accuracy because it is sensitive to a small change when both coordinates are closed to zero [140,141].

Because of the high variability of some vegetation types (Appendix B), spectral reference database built from median spectra, that are *real* spectra, gave worse results than spectral reference database built from median and mean spectra, that are *theoretical* spectra not representative of a *in situ* measured vegetation type (Table 7). There is a need to collect more spectral signatures to build a consistent spectral database.

As spectral signatures can be considered as high dimensional vectors, a specific distance is needed to compare them. It is well known that Euclidean distance is not good when comparing high dimension data [142]. Table 8 shows that the Canberra distance always outperforms other distances, including SAM, which is commonly used in remote sensing, when considering the whole spectral range (1823 wavelengths).

Table 7. Overall accuracy (%) for Canberra distance on [350–2500 nm].

	Median spectra			Median	Mean
	Canberra dist.	City Block dist.	Euclidean dist.	reflectance	reflectance
Spectral signature	53.62	52.34	51.91	57.02	50.64
Normalized spectral signature	51.91	52.34	50.64	55.74	57.87
log transformation of spectral signature	52.34	52.34	51.49	55.74	51.91
First Derivative	70.64	**70.21**	**67.23**	74.47	71.49
Second Derivative	**71.06**	68.51	64.68	**81.70**	**77.45**
Continuum removed Reflectance	51.06	50.64	51.06	54.04	52.77
Continuum Removed Derivative Reflectance	64.68	62.98	61.28	78.30	75.32

Using the Canberra distance, best results (overall accuracy higher than 60 %) are given with the second derivative, first derivative and CRDR (Table 7), that are closely related to absorption features rather than reflectance magnitude [38]. Indeed, vegetation types can be discriminated thanks to their biophysical

Table 8. Overall accuracy (%) for different distances on [350–2500 nm] considering Median reflectances as spectral reference database.

	Distance			SAM
	Euclid	Manhattan	Canberra	
Spectral signature	50.21	51.06	**57.02**	41.70
First Derivative	62.98	70.64	**74.47**	59.15
Second Derivative	65.96	74.04	**81.70**	63.83
CRDR	71.06	74.47	**78.30**	69.36

components which will be discussed in details in Section 4.2. Furthermore, Table 9 shows that the whole spectral range gives the best results. Although spectral ranges are related to specific biophysical components (Table 6), the whole spectral range is needed to discriminate the 13 vegetation types because some of them are sharing same plant species (Table A.1) and the spectral signatures are mixed. Worse results are obtained in [1940–2400 nm] whatever the transformed spectral signature. Table 9 show that worse results are obtained by the spectral signature whatever the spectral range. Indeed those transformations are related to absorption features as explained above, which confirm that transformed spectral signatures are more suitable to discriminate vegetation types than spectral signatures.

Table 9. Overall classification accuracy (%) for different spectral ranges considering Median reflectances as spectral reference database and Canberra distance.

	350–750 nm	750–1350 nm	1410–1810 nm	1940–2400 nm	350–2500 nm
Spectral signature	47.23	47.66	37.87	34.47	**57.02**
First Derivative	59.15	64.68	60.43	55.74	**74.47**
Second Derivative	72.34	69.79	72.34	53.19	**81.70**
CRDR	74.47	57.87	59.57	59.57	**78.30**

Considering classification accuracy for each vegetation type, Table 10 shows that best F1-score is obtained by *Sphagnum* sp. (SPHA) ($\simeq$ 98 %), *Juniperus communis* (JUCO) ($\simeq$ 97 %), Aquatic type b (AQ_B) ($\simeq$ 93 %) and *Salix* sp. (SALI) ($\simeq$ 92 %). Excepting JUCO, all of these vegetation types are well classified and their user's accuracy is higher than 85 %. Indeed these vegetation types are less mixed than others: Table A.1 shows that SPHA is mainly dominated by different kinds of *sphagnum*; AQ_B is dominated by *Utricularia* sp; JUCO is dominated by *Juniperus communis* and SALI is dominated by *Salix*. Only 3 other vegetation types have user's accuracy equal to 100 %: *Rhododendron ferrugineum* (RHFR), *Calluna vulgaris* (CAVU) and Aquatic type a (AQ_A). However, only around 57 % of spectral signatures are well identified for CAVU and AQ_A. This can be explained by the high variability of these sample plots. Contrary to SPHA, JUCO, AQ_B and SALI, there is not a single dominated plant species neither for CAVU nor for AQ_A (Table A.1). Worse F1-score is obtained by *Pinguicula* sp. (PING) ($\simeq$ 54 %) which is not dominated by only one plant species: this vegetation type is mainly dominated by *Eleocharis quinqueflora* (ELQU) (40 %), bare ground (15 %), *Molinia caerulea* ssp *caerulae* (10 %) and *Tomenthypnum nitens* (10 %). It can explain the difficulty to identify this vegetation type in particular rather than the low number of spectra: PING has 8 spectra whereas AQ_B has 7 spectra.

4.2. Supervised classification based on feature selection of spectral vegetation indices

Feature selection

The Kruskal-Wallis method (Section 3.4, p. 14) does not show any significant index (frequency discrimination $>$ 75 %) that allow discrimination between vegetation types (Figure 5, only the first 69 indices

Table 10. Confusion matrix of the classification based on Second derivative, Canberra Distance on [350–2500 nm] with Median reflectance as reference spectral database. The producer's and user's accuracies, the overall accuracy and the F1-score are also shown.

	SPHA	CAVU	RH_FR	CA_HV	AQ_A	SALI	PING	JUQO	ELQU	METR	PI_CV	AQ_B	AQ_C	Producer's accuracy (%)
SPHA	**22**	0	0	0	0	0	0	0	0	0	0	0	0	100.00
CAVU	0	**8**	0	2	0	0	1	0	0	0	3	0	0	57.14
RHFR	0	0	**11**	0	0	3	0	0	0	0	0	0	0	78.57
CA_HV	0	0	0	**22**	0	0	1	0	1	0	3	0	0	81.48
AQ_A	0	0	0	0	**30**	0	8	0	3	4	1	1	6	56.60
SALI	0	0	0	0	0	**17**	0	0	0	0	0	0	0	100.00
PING	0	0	0	1	0	0	**7**	0	0	0	0	0	0	87.50
JUCO	0	0	0	0	0	0	0	**18**	0	0	1	0	0	94.74
ELQU	0	0	0	1	0	0	0	0	**13**	1	0	0	0	86.67
METR	1	0	0	0	0	0	0	0	0	**11**	0	0	0	91.67
PI_CV	0	0	0	0	0	0	1	0	0	0	**14**	0	0	93.33
AQ_B	0	0	0	0	0	0	0	0	0	0	0	**7**	0	100.00
AQ_C	0	0	0	0	0	0	0	0	0	0	0	0	**12**	100.00
User's accuracy (%)	95.65	100.00	100.00	84.62	100.00	85.00	38.89	100.00	76.47	68.75	63.64	87.50	66.67	Overall accuracy: **81.70**
F1-score (%)	97.78	72.73	88.00	83.02	72.29	91.89	53.85	97.30	81.25	78.57	75.68	93.33	80.00	

are drawn). The best vegetation index (NDWI[860, 2130]) only allows us to discriminate 49 pairs of vegetation types, that may be explained by the plant species mixing within several vegetation types. The proposed method reduced the number of selected indices from 129 to 26 (Table 11). More precisely, on the first step of the method, only 17 single indices amongst 26 are needed to discriminate 59 pairs of vegetation types amongst 78. On the second step, these single indices must be completed by 7 additional spectral vegetation indices to discriminate 17 more pairs of vegetation types (Table 12 ; ∅ means either a pair of vegetation type can not be discriminated thanks to a pair of spectral vegetation indices built from single ones selected on the first step, either more than two vegetation indices are needed to discriminate a pair of vegetation type). On the last step, a single index is added to discriminate two vegetation types whereas a combination of previous selected indices allows us to discriminate another pair of vegetation type (Table 11). Finally several different – single or pair or triplet – vegetation indices allow us to discriminate pairs of vegetation types. However, none single spectral index allows us to discriminate all pairs of vegetation types nor the majority: *e.g.* the most discriminating single spectral index, the Water Index (WI), only discriminates around 45 % pairs of vegetation types (Table 11).

Table 13 shows that one single biophysical component can discriminate most of vegetation types except *Carex* sp. homogeneous vegetation (CA_HV). More precisely, three kinds of vegetation types (*Sphagnum* sp. (SPHA), Aquatic type b (AQ_B) and Aquatic type c (AQ_C)) are separated thanks to a single biophysical component. However, some biophysical components are more discriminant than others according to vegetation types: *e.g.* the chlorophyll is more discriminant than the water content for AQ_C whereas the water content is the only discriminant biophysical component for AQ_B ; the water content, the chlorophyll and water, cellulose, starch, lignin (w., c., s., l.) equally discriminate SPHA from all other vegetation types.

Only two indices related to water content are needed to separate AQ_B from all other vegetation types: WI and NDWI[860,1240] (Table 13) because AQ_B vegetation type is mainly composed of *Utricularia* sp. and water (Table A.1). The AQ_B spectral signatures are lower than the spectral reflectance values of the other vegetation types and the water absorption band at 900 nm and 970 nm are highlighted.

The chlorophyll is the main biophysical component (86.33 %) able to discriminate AQ_C from all other vegetation types, except with Aquatic type a (AQ_A) and AQ_B differentiated by considering additional water indices (MSI and NDWI[860,1240]). Indeed, dry matter can be seen on spectral signatures (Figure 7): AQ_B has the lowest slope on the spectral range [705–730 nm] whereas other vegetation types (except AQ_A and AQ_B) have higher values because they still contain chlorophyll. However, as AQ_B and AQ_C have low values of Boochs2 index, they can be discriminated thanks to a water index (right side of Figure 8 shows that those vegetation types can be clearly separated ; indeed, those vegetation types have different shapes and values that characterize each type).

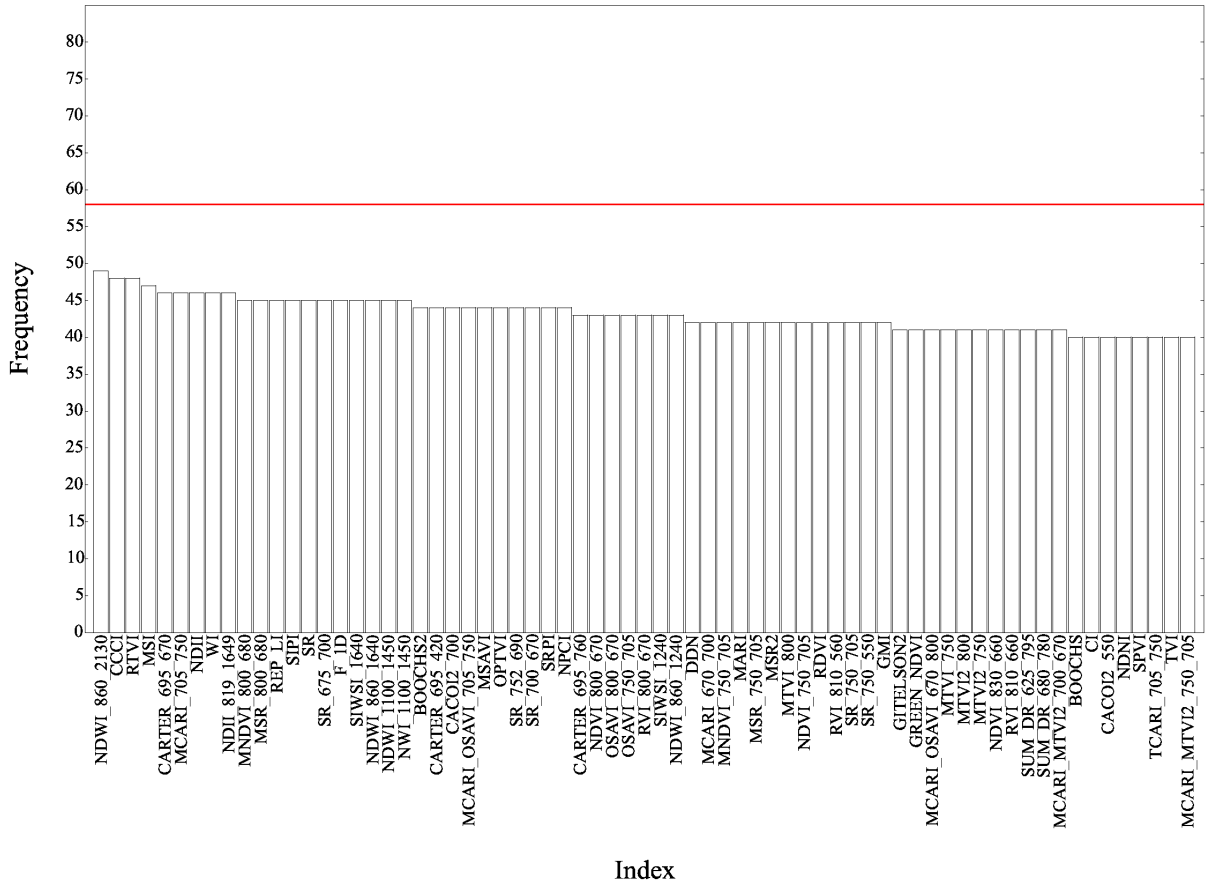


Figure 5. Frequency distribution of the Kruskal-Wallis test for the 129 spectral indices for paired species across the 13 vegetation types. The horizontal red line stands for 75 % of all 78 possible combinations of the 13 vegetation types.

Table 11. Single selected indices from the Hellinger distance and their occurrences.

Biophysical component	Index name	No. of all occurrences	No. of single occurrences	No. of occurrences within pair	No. of occurrences within triple
Chlorophyll	CCCI*	35	24	10	1
	GMI*	33	25	8	0
	DPI**	33	16	17	0
	NDVI[750,705]**	32	25	7	0
	BOOCHS2*	32	24	8	0
	SR[700,670]*	31	25	6	0
	OSAVI[800,670]*	31	20	8	3
	DDN**	26	18	8	0
	MNDVI[800,680]*	23	18	5	0
	GITELSON***	13	5	5	3
Water	WI*	40	33	6	1
	MSI*	39	31	8	0
	NDWI[860,1240]*	38	31	7	0
	NDII**	38	28	9	1
	NDWI[860,2130]*	35	24	11	0
	NDWI[1100,1450]**	32	22	10	0
Stress	CARTER[695,670]*	36	26	9	1
	CARTER[695,420]**	36	16	20	0
Pigment	MARI*	75	13	62	0
	PRI*	35	9	26	0
Nitrogen	NDNI*	37	18	19	0
	MCARI/MTVI2[750,705]**	30	22	8	0
(Total pigments)/chlorophyll	NPCI*	31	18	13	0
	SRPI*	29	16	13	0
Water, cellulose, starch, lignin	F_1D*	89	27	62	2
	F_WP*	20	15	5	0

x*: index selected on first step.
x**: index selected on second step.
x***: index selected on third step.

Table 12. Single spectral index or pairs of spectral indices retained to discriminate vegetation types pairs.

	CAVU	RHFR	CA_HV	AQ_A	SALI
SPHA	F_WP	F_WP	WI	OSAVI[800,670]	F_WP
CAVU	-	∅	∅	F_1D	∅
RHFR	NPCI-F_1D	-	∅	∅	∅
CA_HV	MARI-WI	CARTER[695, 670]-MCARI/MTVI2[750, 705]	-	∅	∅
AQ_A	-	F_1D-WI	NDNI-NDWI[1100,1450]	-	∅
SALI	CARTER[695, 420]-NDII	CARTER[695, 670]-BOOCHS2	SRPI-NDVI[750,705]	F_1D-MSI	-
PING	-	-	NDNI-WI	DDN-NDWI[860,2130]	-
JUCO	-	F_1D-WI	-	-	-
ELQU	-	-	MARI-WI	MARI-MSI	-
METR	-	-	CCCI-NDWI[860,1240]	-	-
PI_CV	-	-	-	-	-
AQ_B	-	-	-	-	-

∅ means either a pair of vegetation type can not be discriminated thanks to a pair of spectral vegetation indices built from single ones selected on the first step, either more than two vegetation indices are needed to discriminate a pair of vegetation type.

	PING	JUCO	ELQU	METR	PI_CV	AQ_B	AQ_C
SPHA	MSI	F_WP	CCCI	CCCI	WI	WI	OSAVI[800,670]
CAVU	GMI	MARI	CCCI	GMI	GMI	WI	SR[700, 670]
RHFR	MNDVI[800, 680]	∅	SRPI	CCCI	WI	WI	MNDVI[800, 680]
CA_HV	∅	F_WP	∅	∅	∅	WI	CCCI
AQ_A	∅	F_1D	∅	∅	NDNI	WI	MSI
SALI	NPCI	F_WP	NPCI	NPCI	NPCI	WI	MNDVI[800,680]
PING	-	NDWI[860, 2130]	PRI	∅	BOOCHS2	NDWI[860,1240]	BOOCHS2
JUCO	-	-	F_WP	F_WP	F_WP	WI	MNDVI[800,680]
ELQU	-	-	-	MARI	CARTER[695,420]	NDWI[860,1240]	BOOCHS2
METR	DPI-F_1D	-	-	-	∅	NDWI[860,1240]	BOOCHS2
PI_CV	-	-	-	PRI-WI	-	NDWI[860,1240]	OSAVI[800,670]
AQ_B	-	-	-	-	-	-	NDWI[860,1240]

Table 13. Single main discriminating biophysical components for each vegetation type and their occurrences (%).

Biophysical components	SPHA	CAVU	RHFR	CA_HV	AQ_A	SALI	PING	JUCO	ELQU	METR	PI_CV	AQ_B	AQ_C
Water	33.33	8.33	16.67	16.67	16.67	8.33	25.00	16.67	8.33	8.33	25.00	100.00	16.67
Chlorophyll	33.33	41.67	25.00	8.33	8.33	8.33	33.33	8.33	25.00	33.33	25.00	0.00	83.33
Stress	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	8.33	0.00	8.33	0.00	0.00
Nitrogen	0.00	0.00	0.00	0.00	8.33	0.00	0.00	0.00	0.00	0.00	8.33	0.00	0.00
Pigment	0.00	8.33	0.00	0.00	0.00	0.00	8.33	8.33	16.67	8.33	0.00	0.00	0.00
(Total pigments)/chlorophyll	0.00	0.00	8.33	0.00	0.00	33.33	8.33	0.00	16.67	8.33	8.33	0.00	0.00
W., c., s., l.	33.33	16.67	8.33	8.33	16.67	16.67	0.00	58.33	8.33	8.33	8.33	0.00	0.00
Total	100.00	75.00	58.33	33.33	50.00	66.67	75.00	91.67	83.33	66.67	83.33	100.00	100.00

W., c., s., l. = Water, cellulose, starch, lignin.

Table 14. Pairs of main discriminating biophysical components for each vegetation type and their occurrences (%).

Biophysical components	CAVU	RHFR	CA_HV	AQ_A	SALI	PING	JUCO	ELQU	METR	PI_CV
Water - chlorophyll	0.00	0.00	8.33	8.33	0.00	8.33	0.00	0.00	8.33	0.00
Water - stress	8.33	0.00	0.00	0.00	8.33	0.00	0.00	0.00	0.00	0.00
Water - nitrogen	0.00	0.00	16.67	8.33	0.00	8.33	0.00	0.00	0.00	0.00
Water - pigment	8.33	0.00	16.67	8.33	0.00	0.00	0.00	16.67	8.33	8.33
Water - w., c., s., l.	0.00	16.67	0.00	16.67	8.33	0.00	8.33	0.00	0.00	0.00
Chlorophyll - stress	0.00	8.33	0.00	0.00	8.33	0.00	0.00	0.00	0.00	0.00
Chlorophyll - (total pigments)/chlorophyll	0.00	0.00	8.33	0.00	8.33	0.00	0.00	0.00	0.00	0.00
Chlorophyll - w., c., s., l.	0.00	0.00	0.00	0.00	0.00	8.33	0.00	0.00	8.33	0.00
Stress - nitrogen	0.00	8.33	8.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00
(Total pigments)/chlorophyll - w., c., s., l.	8.33	8.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

W., c., s., l. = Water, cellulose, starch, lignin.

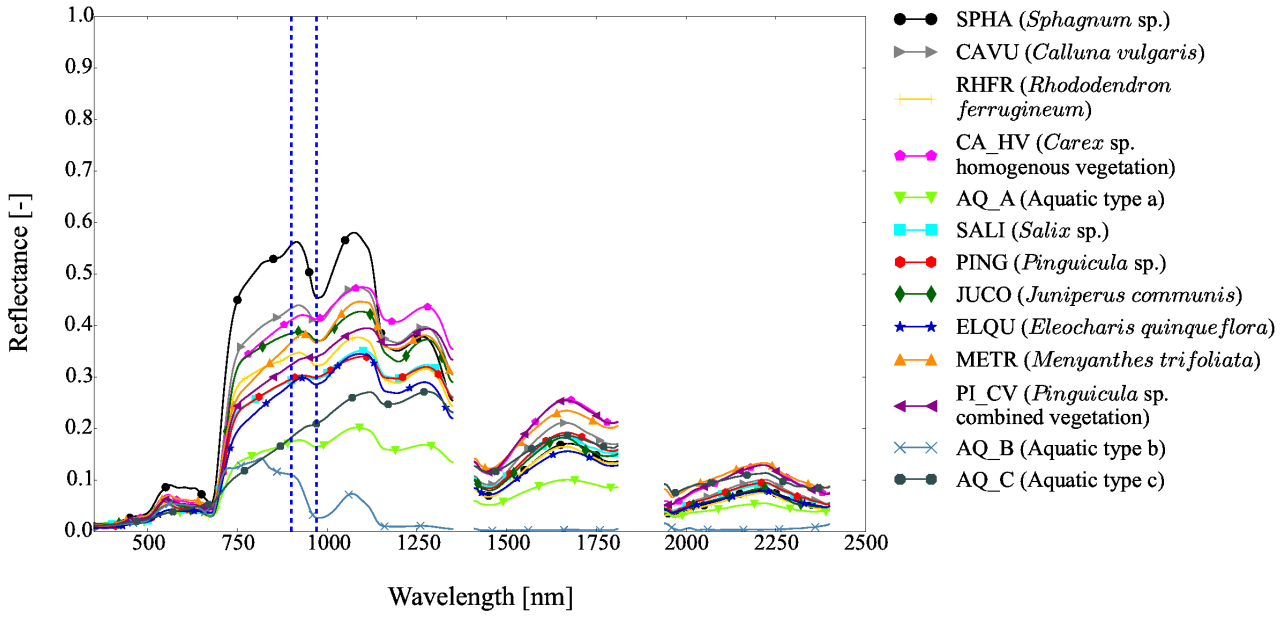


Figure 6. Mean spectral reflectance of the 13 vegetation types. Dashed lines represent the wavelengths used by WI.

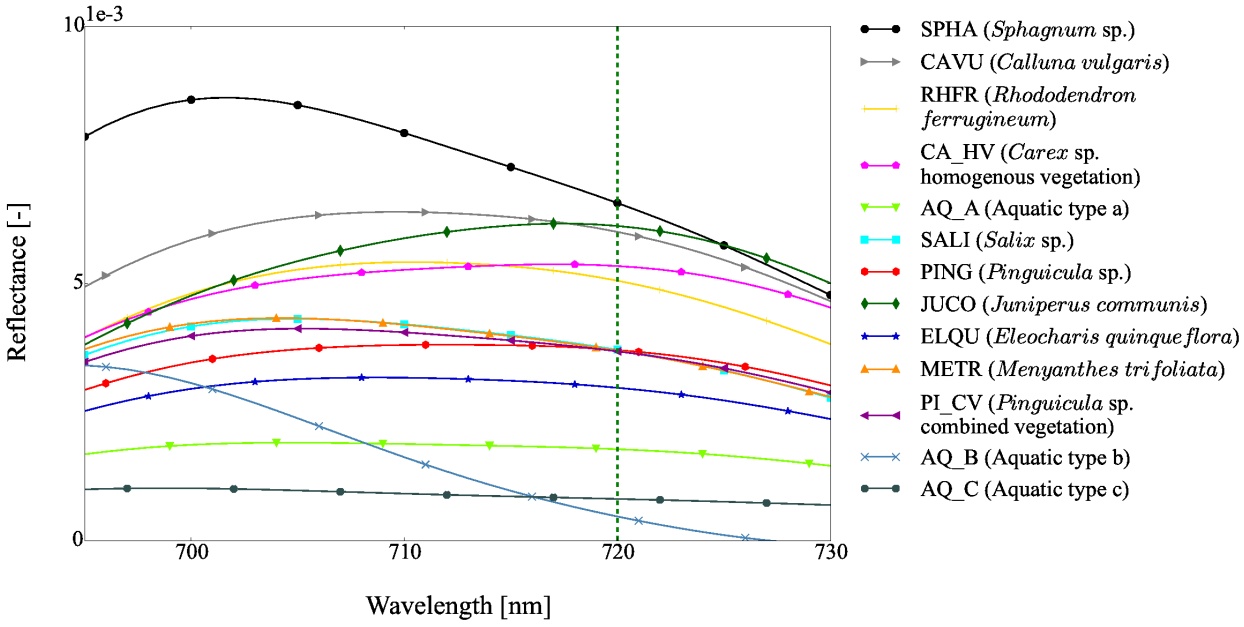


Figure 7. Mean first derivative spectral signatures of the 13 vegetation types on [695–730 nm]. The green dashed line represents the wavelength used by the Boochs2 index.

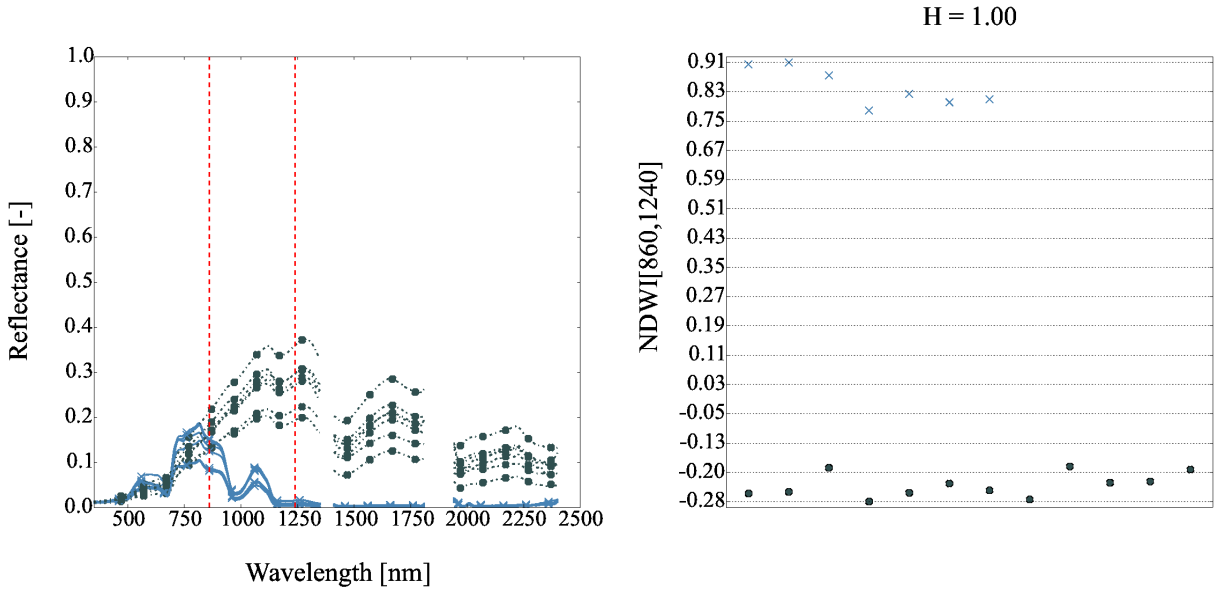


Figure 8. Left: spectral signatures of AQ_B (blue) and AQ_C (dark slate gray). Red dashed lines are the wavelengths used by the NDWI[860,1240] index. Right: NDWI[860,1240] values for each vegetation type, H is the Hellinger distance.

In some case, there is no single biophysical component allowing us to discriminate vegetation types: *e.g.* both water content (33.33 %), chlorophyll (33.33 %) and w., c., s., l. (33.33 %) are needed to distinguish SPHA from all other vegetation types. More precisely, biophysical components related to water (WI, MSI) are discriminating SPHA from CA_HV, *Pinguicula* sp. (PING), *Pinguicula* sp. combined vegetation (PI_CV) and AQ_B ; biophysical components related to chlorophyll (CCCI, OSAVI[800,670]) are differentiating SPHA from AQ_A, AQ_C, *Eleocharis quinqueflora* (ELQU) and *Menyanthes trifoliata* (METR) ; biophysical components related to w., c., s., l. (F_WP) are separating SPHA from *Calluna vulgaris* (CAVU), *Rhododendron ferrugineum* (RHFR), *Salix* sp. (SALI) and *Juniperus communis* (JUCO) (Table 13). Unlike an index related to water content (Figure 9), an index related to the chlorophyll will discriminate SPHA from AQ_A. Indeed, the right side of Figure 9 shows that some AQ_A plant species can not be distinguished from SPHA because it is a dry moss and the left side of Figure 9 shows that SPHA and non discerned AQ_A have the same spectral signature shape. The right side of Figure 10 shows that these two vegetation species can clearly be separated despite the class variability of AQ_A. A complex biophysical component such as F_WP will differentiate SPHA from CAVU (left side of Figure 11) shows that different spectral shapes between those vegetation types can be exploited on the [1220–1280 nm] domain. The right side of Figure 10 shows that the wavelengths corresponding to the maximum of the first derivatives can clearly discern these two vegetation types even if these vegetation types can be mixed.

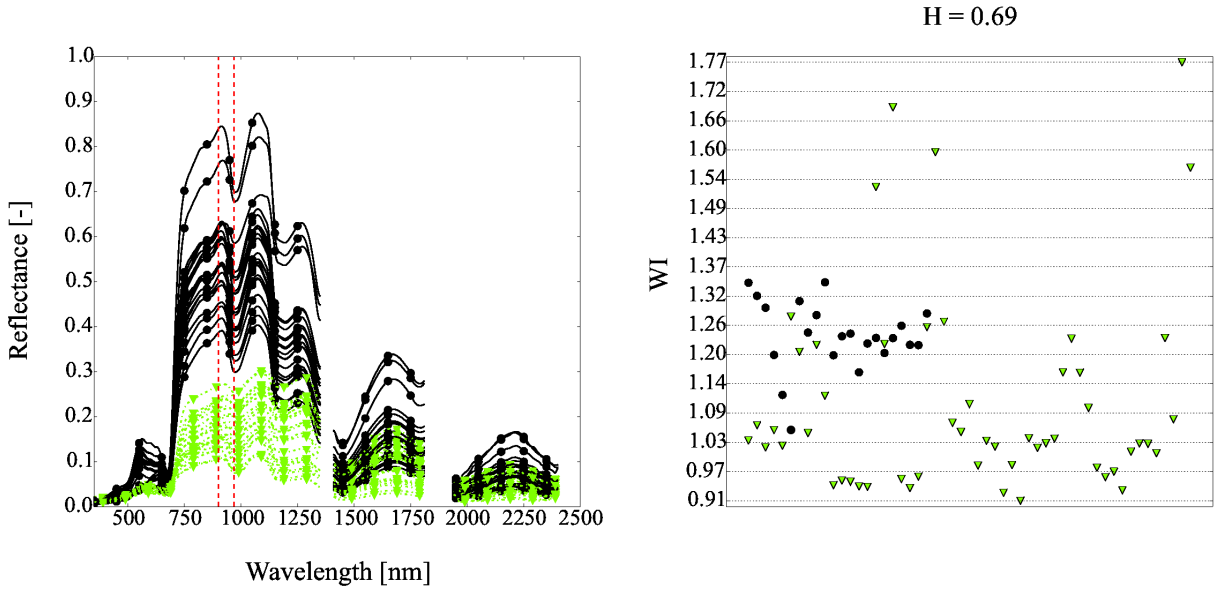


Figure 9. Left: spectral signatures of SPHA (black) and AQ_A (green). Red dashed lines are WI wavelengths. Right: WI values for each vegetation type, H is the Hellinger distance.

In most case, a single biophysical component is sufficient to class a vegetation type from the others (except for CA_HV), but a pair of biophysical components is needed to discriminate more specifically some vegetation types (Table 12), apart from some particular cases where a pair of biophysical components is needed CA_HV (Figure 12). Indeed, CAVU and SALI are differentiated with the stress index (CARTER[695, 420]) and the water index (NDII).

Among the 78 combinations of pair of vegetation types, only two require three indices to be separated: CA_HV *vs* PING and AQ_A *vs* METR. Indeed, because of its within class variability (Table A.1), only 33.33 % of single biophysical component can discriminate CA_HV from all other vegetation types (Table 13). Besides, as mentioned in Section 4.1, none of the main plant species of PING represents more than 50 % of this vegetation type. The advent of a third index only improves significantly their discrimination (Figure 13).

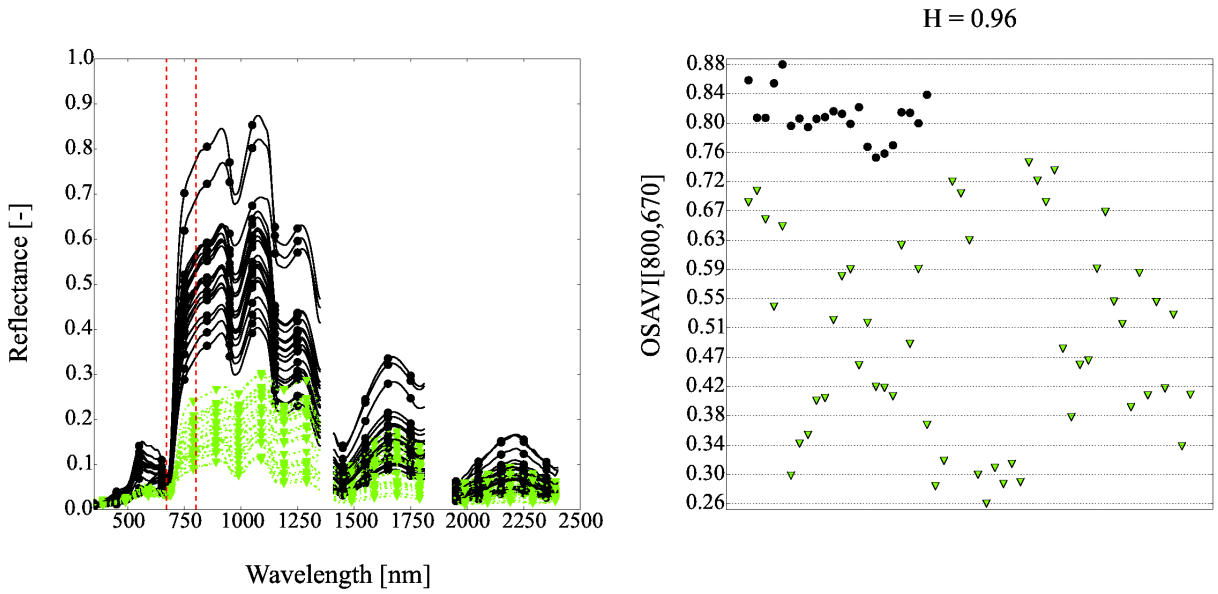


Figure 10. Left: spectral signatures of SPHA (black) and AQ_A (green). Red dashed lines are OSAVI[800,670] wavelengths. Right: OSAVI[800,670] values for each vegetation type, H is the Hellinger distance.

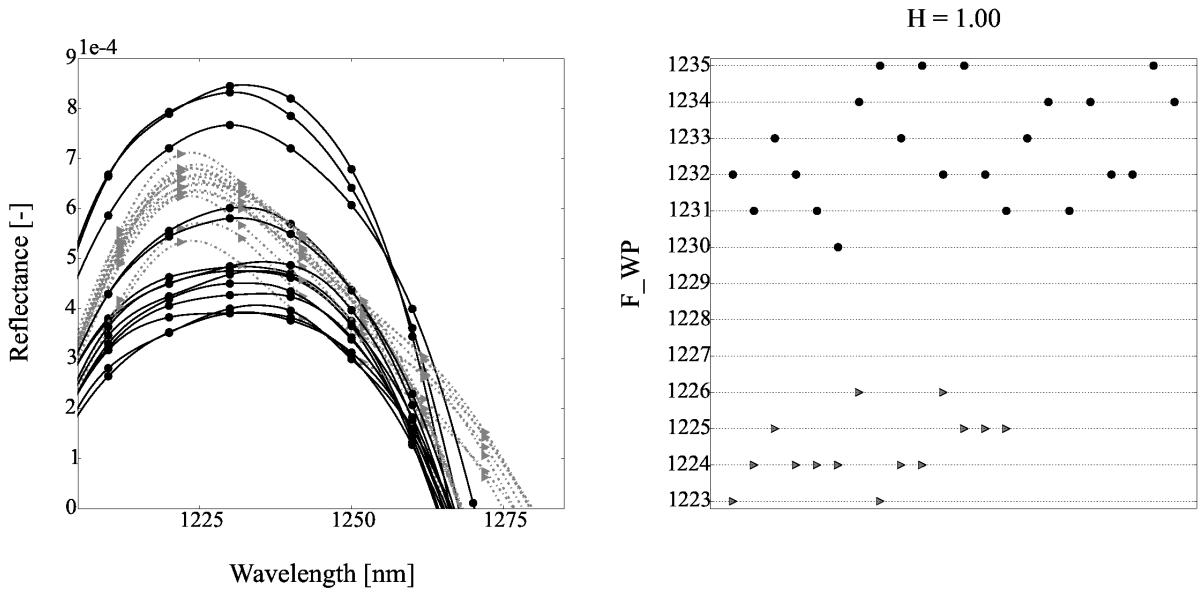


Figure 11. Left: spectral signatures of SPHA (black) and CAVU (gray). Right: F_WP values for each vegetation type, H is the Hellinger distance.

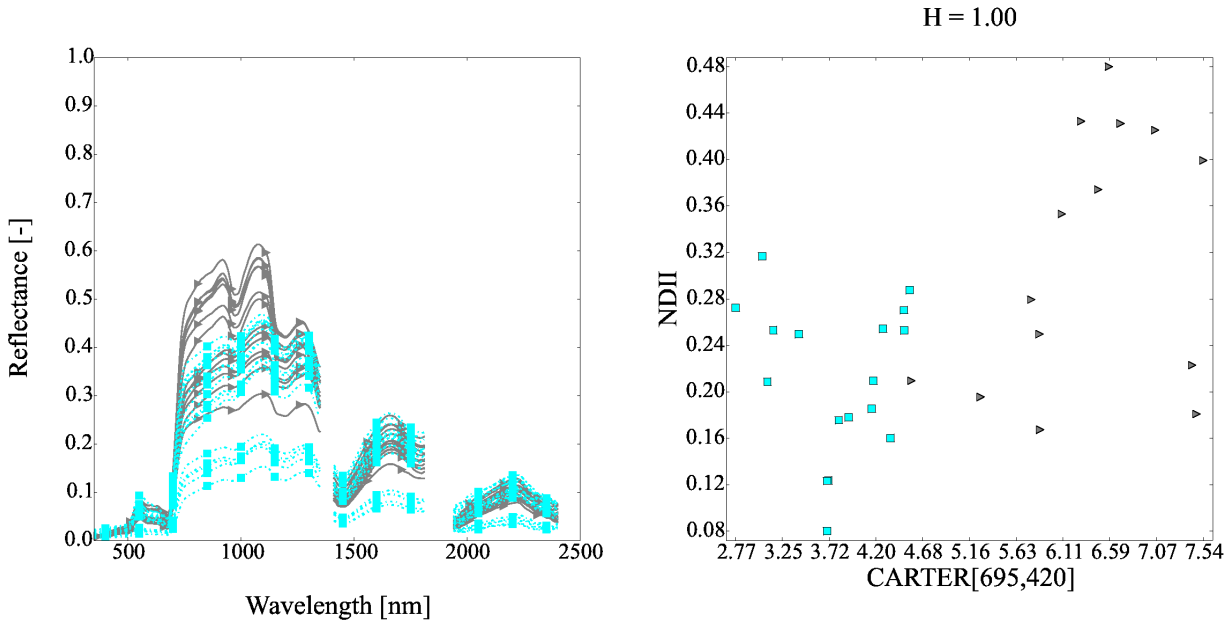


Figure 12. Left: spectral signatures of CAVU (gray) and SALI (cyan). Right: map of CARTER[695,420] and NDII values for each vegetation type, H is the Hellinger distance.

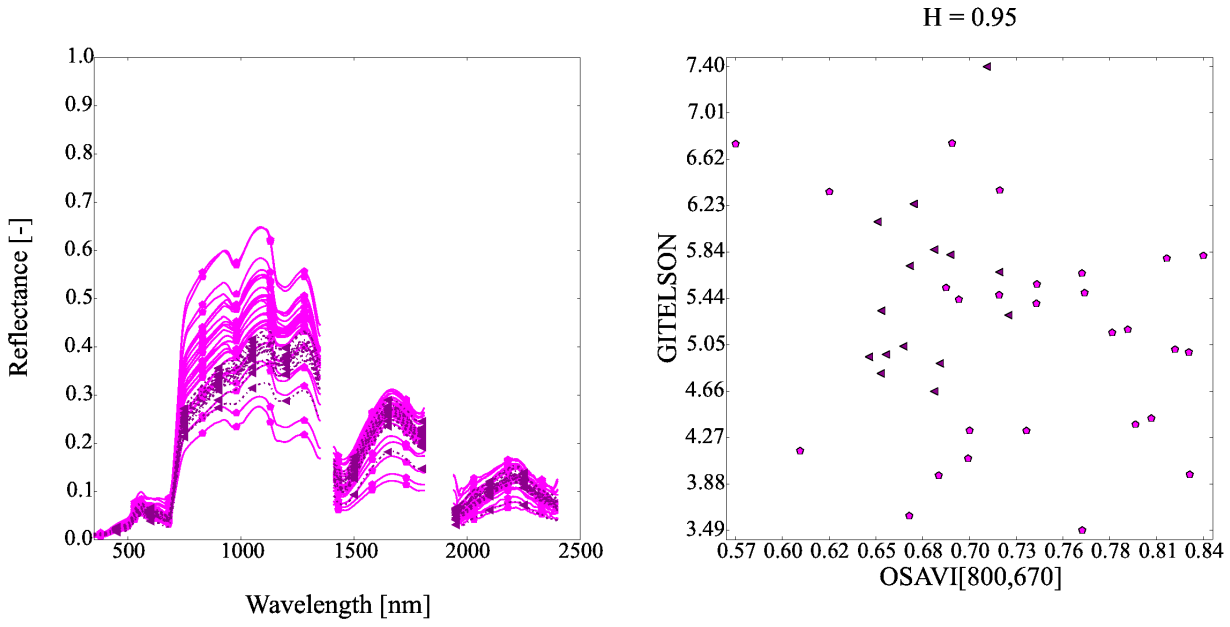


Figure 13. Left: spectral signatures of CA_HV (pink) and PI_CV (magenta). Right: map of OSAVI[800,670] and GITELSON values for each vegetation type, H is the Hellinger distance value.

399 Supervised classification

400 The 26 indices selected WITH the Hellinger distance enables overall classification accuracy scores ranging
401 from 72.90 % to 85.20 % depending on the training size, whereas when considering all indices overall accuracy
402 scores range from 66.70 % to 82.80 % (Table 15). Moreover, these selected indices are robust because no significant
403 difference between classifiers score (except for RF) regardless of the training size is noted (Figure 14). As
404 expected, worst results are given by the Kruskal-Wallis method (to compare performance of the two features
405 selection methods, 26 first indices given by Kruskal-Wallis method have been selected).

406 RLR gives better results than SVM and RF (Table 15, Figure 14) except when the size of training set equals
407 50 % for the Hellinger distance. That may be explained by the possible confusion between some vegetation
408 types due to their plant species composition. Indeed, SVM aims to find the best hyperplane that can separate
409 data, whereas RLR aims to find a probability (according to a logistic function) to separate them.

410 Considering RLR-$\ell 2$ some vegetation types are not easily discriminated whatever the indices. Table 16 and
411 Table 17 show that PING has the lowest F1-score (20.99 % and 33.13 % respectively) which can be explained by
412 the mixed composition of this habitat (Table B) and not the low number of spectra. Indeed, AQ_B has about the
413 same number of spectra: 7 spectra whereas 8 spectral measurements have been collected for PING. Yet it has a
414 F1-score = 91.95 % considering all indices and F1-score = 91.66 % considering indices selected by the Hellinger
415 distance that can be explained by its composition dominated by *Utricularia* sp.

416 Focusing on shrubs, JUCO has the best performances (F1-score = 94.83 %) whereas SALI and RHFR are
417 often confounded. Table 17 shows that on average 2.53 spectra of RHFR ($\simeq$ 20.02 %) are classified as SALI and
418 on average 2.30 spectra of SALI ($\simeq$ 19.15 %) are classified as RHFR. Indeed, as JUCO has a higher foliage density,
419 the overall spatial signature is less sensitive to the ground influence and as a result JUCO spectral reflectance is
420 close to a pure endmember (Appendix B). In the latter case, the spectral measurements are composed of soil
421 and more affected by mixed signatures. Another pair of vegetation types is hardly discriminated: PI_CV and
422 CA_HV. Table 17 shows that on average 4.93 spectra of CA_HV ($\simeq$ 25 %) are classified as PI_CV which may be
423 explained by the plant species they have in common: *Carex* (50 %–100 % depending on the location) and *Molinia*
424 *caerulea* ssp. *caerulae* (40 %–70 %) (Appendix B).

Table 15. Vegetation types identification (overall accuracy ($\pm$ standard deviation) in %) with indices.

Training size	Classifier	Overall accuracy ($\pm$ Standard deviation) (%)		
		All indices	Kruskal-Wallis	Hellinger distance
50 %	SVM linear	79.17 ($\pm$ 3.51)	75.45 ($\pm$ 3.95)	83.31 ($\pm$ 3.95)
	SVM RBF	77.63 ($\pm$ 2.82)	75.45 ($\pm$ 3.65)	**83.55**($\pm$**3.65**)
	RLR-ℓ_1	**80.58**($\pm$**3.05**)	**78.37**($\pm$**3.54**)	82.84 ($\pm$ 3.54)
	RLR-ℓ_2	80.55 ($\pm$ 3.33)	78.07 ($\pm$ 3.48)	83.22 ($\pm$ 3.48)
	RF	78.71 ($\pm$ 3.34)	71.05 ($\pm$ 3.56)	81.60 ($\pm$ 3.56)
45 %	SVM linear	78.44 ($\pm$ 3.09)	74.82 ($\pm$ 3.86)	82.46 ($\pm$ 3.86)
	SVM RBF	76.59 ($\pm$ 4.39)	74.49 ($\pm$ 4.53)	83.21 ($\pm$ 4.53)
	RLR-ℓ_1	**80.26**($\pm$**4.25**)	77.26 ($\pm$ 4.16)	**83.51**($\pm$**4.16**)
	RLR-ℓ_2	79.85 ($\pm$ 3.36)	**77.64**($\pm$**3.80**)	83.13 ($\pm$ 3.80)
	RF	77.26 ($\pm$ 4.14)	70.33 ($\pm$ 3.04)	80.26 ($\pm$ 3.04)
40 %	SVM linear	76.95 ($\pm$ 3.59)	73.33 ($\pm$ 3.48)	81.89 ($\pm$ 3.48)
	SVM RBF	76.28 ($\pm$ 3.27)	73.43 ($\pm$ 3.84)	81.68 ($\pm$ 3.84)
	RLR-ℓ_1	79.69 ($\pm$ 3.43)	77.72 ($\pm$ 3.62)	**83.19**($\pm$**3.62**)
	RLR-ℓ_2	**79.74**($\pm$**2.47**)	**78.25**($\pm$**3.34**)	82.97 ($\pm$ 3.34)
	RF	76.86 ($\pm$ 3.41)	70.34 ($\pm$ 3.96)	80.96 ($\pm$ 3.96)
35 %	SVM linear	76.02 ($\pm$ 3.35)	70.41 ($\pm$ 3.57)	80.02 ($\pm$ 3.57)
	SVM RBF	73.44 ($\pm$ 4.38)	71.02 ($\pm$ 4.17)	79.20 ($\pm$ 4.17)
	RLR-ℓ_1	74.98 ($\pm$ 2.74)	74.87 ($\pm$ 3.78)	80.89 ($\pm$ 3.78)
	RLR-ℓ_2	**77.25**($\pm$**2.80**)	**75.06**($\pm$**2.76**)	**81.04**($\pm$**2.76**)
	RF	75.32 ($\pm$ 3.32)	67.79 ($\pm$ 3.55)	79.37 ($\pm$ 3.55)
30 %	SVM linear	73.62 ($\pm$ 3.84)	70.53 ($\pm$ 3.18)	78.34 ($\pm$ 3.18)
	SVM RBF	72.71 ($\pm$ 2.82)	69.68 ($\pm$ 4.33)	79.13 ($\pm$ 4.33)
	RLR-ℓ_1	74.08 ($\pm$ 4.03)	**73.66**($\pm$**3.23**)	79.25 ($\pm$ 3.23)
	RLR-ℓ_2	**75.74**($\pm$**3.99**)	73.39 ($\pm$ 3.33)	**80.36**($\pm$**3.33**)
	RF	72.53 ($\pm$ 2.60)	66.00 ($\pm$ 2.74)	77.17 ($\pm$ 2.74)
25 %	SVM linear	71.37 ($\pm$ 3.18)	68.38 ($\pm$ 3.44)	75.91 ($\pm$ 3.44)
	SVM RBF	69.85 ($\pm$ 3.54)	67.63 ($\pm$ 2.67)	75.76 ($\pm$ 2.67)
	RLR-ℓ_1	69.42 ($\pm$ 4.06)	70.90 ($\pm$ 3.34)	76.35 ($\pm$ 3.34)
	RLR-ℓ_2	**73.31**($\pm$**3.34**)	**71.22**($\pm$**3.72**)	**77.21**($\pm$**3.72**)
	RF	70.79 ($\pm$ 2.95)	65.10 ($\pm$ 3.31)	75.05 ($\pm$ 3.31)

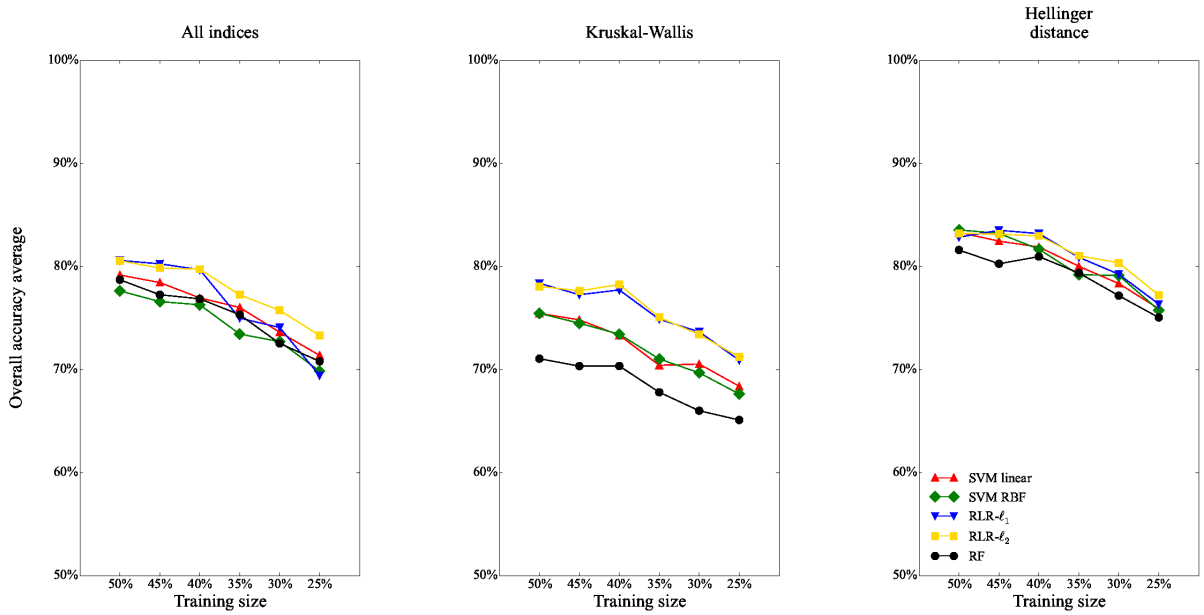


Figure 14. Vegetation types identification accuracies (overall accuracy) with indices.

Table 16. Confusion matrix of the classification based on RLR$-\ell_2$ with all indices and training size = 25 %. The producer's and user's accuracies and the overall accuracy average (OAA) are also shown.

	SPHA	CAVU	RH_FR	CA_HV	AQ_A	SALI	PING	JUCO	ELQU	METR	PI_CV	AQ_B	AQ_C	Producer's accuracy (%)
SPHA	**15.20**	0.73	0.43	0.33	0.00	0.17	0.00	0.00	0.07	0.03	0.03	0.00	0.00	89.46
CAVU	2.30	**6.20**	0.67	0.83	0.00	0.07	0.30	0.20	0.17	0.10	0.17	0.00	0.00	56.31
RHFR	1.13	0.77	**4.20**	0.00	0.07	1.67	0.70	1.57	0.50	0.17	0.23	0.00	0.00	38.15
CA_HV	0.00	0.17	0.00	**12.17**	1.03	0.00	0.53	0.07	0.57	0.57	4.90	0.00	0.00	60.82
AQ_A	0.00	0.00	0.07	0.47	**33.40**	0.20	0.83	0.00	0.80	1.60	1.00	0.17	1.47	83.48
SALI	0.00	0.30	1.00	0.13	1.33	**8.57**	0.23	0.00	0.30	0.40	0.70	0.00	0.03	65.97
PING	0.00	0.23	0.23	1.57	1.13	0.00	**1.10**	0.00	0.60	0.27	0.83	0.00	0.03	18.36
JUCO	0.07	0.00	0.10	0.00	0.13	0.00	0.10	**13.40**	0.00	0.00	0.20	0.00	0.00	95.71
ELQU	0.00	0.00	0.00	0.00	0.00	0.00	0.07	0.00	**10.93**	0.00	0.00	0.00	0.00	**99.36**
METR	0.07	0.00	0.00	1.17	1.40	0.00	0.23	0.00	0.63	**4.43**	1.03	0.00	0.03	49.28
PI_CV	0.00	0.00	0.07	1.83	0.40	0.03	0.37	0.00	0.03	0.10	**8.03**	0.00	0.13	73.07
AQ_B	0.23	0.00	0.00	0.00	0.30	0.00	0.00	0.00	0.00	0.00	0.07	**4.40**	0.00	88.00
AQ_C	0.00	0.07	0.00	0.10	0.67	0.03	0.03	0.00	0.00	0.07	0.30	0.00	**7.73**	85.89
User's accuracy (%)	80.00	73.20	62.04	65.43	83.79	79.80	24.50	87.93	74.86	57.24	45.91	**96.28**	82.06	OAA: **73.31**
F1-score (%)	84.47	63.66	47.24	63.04	83.64	72.23	20.99	91.66	85.39	52.96	56.39	**91.95**	83.93	

4.3. Supervised classification according to the spectral ranges

Only the best results are presented, obtained with the four spectral ranges ([350–750 nm], [750–1350 nm], [350–1350 nm], [350–2500 nm]) and the spectral signature as reference and the three transformed spectral signatures (second derivative, first derivative, Continuum Removed Derivative Reflectance).

Table 18 to Table 21 show the best results obtained with RLR-ℓ_2 on [350–1350 nm] whatever the transformed spectral signatures.

Considering wavelengths used by selected indices (Section 4.2), most of them use spectral bands located on [350–1350 nm] either: 50 % are located in visible range and 32.35 % in near-infrared range. Indeed, in this spectral range all the biophysical components discriminating the peatland vegetation types can be taken into account. That is confirmed by Figure 15 which shows that the best results are given by [350–1350 nm] considering the training size = 25 % regardless the transformed spectral signatures and the the classifier, except

Table 17. Confusion matrix of the classification based on RLR$-\ell_2$ with indices selected by the Hellinger distance and training size = 25 %. The producer's and user's accuracies and the overall accuracy average (OAA) are also shown.

	SPHA	CAVU	RH_FR	CA_HV	AQ_A	SALI	PING	JUCO	ELQU	METR	PI_CV	AQ_B	AQ_C	Producer's accuracy (%)
SPHA	**15.40**	0.90	0.13	0.47	0.00	0.00	0.00	0.00	0.03	0.07	0.00	0.00	0.00	90.59
CAVU	0.90	**8.03**	0.67	0.47	0.00	0.03	0.70	0.00	0.03	0.03	0.13	0.00	0.00	73.07
RHFR	0.47	0.30	**6.70**	0.03	0.00	2.53	0.43	0.20	0.13	0.20	0.00	0.00	0.00	60.96
CA_HV	0.00	0.17	0.20	**11.93**	0.77	0.00	0.77	0.03	0.57	0.63	4.93	0.00	0.00	59.65
AQ_A	0.00	0.00	0.23	0.40	**33.40**	0.43	1.50	0.03	0.43	1.63	1.33	0.00	0.60	83.54
SALI	0.00	0.00	2.30	0.00	0.87	**7.77**	0.80	0.07	0.03	0.40	0.60	0.00	0.17	59.72
PING	0.00	0.27	0.17	1.67	0.37	0.00	**2.20**	0.00	0.17	0.40	0.73	0.00	0.03	36.61
JUCO	0.00	0.03	0.20	0.07	0.10	0.17	0.07	**12.93**	0.00	0.07	0.37	0.00	0.00	92.29
ELQU	0.00	0.00	0.00	0.07	0.00	0.00	0.33	0.00	**10.60**	0.00	0.00	0.00	0.00	**96.36**
METR	0.00	0.00	0.03	0.87	0.73	0.00	0.07	0.00	0.03	**6.23**	1.03	0.00	0.00	69.30
PI_CV	0.00	0.00	0.10	1.23	0.17	0.07	0.23	0.00	0.00	0.37	**8.83**	0.00	0.00	80.27
AQ_B	0.03	0.00	0.47	0.00	0.10	0.00	0.07	0.00	0.00	0.10	0.00	**4.23**	0.00	84.60
AQ_C	0.00	0.00	0.00	0.03	0.47	0.00	0.10	0.00	0.00	0.00	0.00	0.00	**8.40**	93.33
User's accuracy (%)	91.67	82.78	59.82	69.20	90.32	70.64	30.26	97.51	88.19	61.50	49.19	**100.00**	91.30	OAA: **77.21**
F1-score (%)	91.12	77.62	60.39	64.07	86.80	64.72	33.13	**94.83**	92.09	65.17	61.00	91.66	92.31	

Table 18. Vegetation types identification accuracies (overall accuracy (± standard deviation) in %) on [350–750 nm].

Training size	Classifier	Overall accuracy (± Standard deviation) (%)			
		Spectral signature	Second derivative	First derivative	Continuum Removed Derivative Reflectance
50 %	SVM linear	80.99 (± 6.61)	86.94 (± 5.21)	85.95 (± 3.81)	**88.26**(±**2.53**)
	SVM RBF	67.44 (± 4.69)	78.35 (± 2.74)	81.32 (± 2.13)	86.94 (± 3.11)
	RLR-ℓ_1	86.45 (± 3.57)	86.94 (± 4.10)	89.75 (± 2.48)	86.94 (± 1.76)
	RLR-ℓ_2	**88.10**(±**3.64**)	**88.43**(±**2.02**)	**90.91**(±**2.86**)	87.44 (± 1.84)
	RF	62.98 (± 3.52)	84.79 (± 4.92)	73.88 (± 2.84)	86.45 (± 4.07)
	PLS-DA	75.21 (± 3.88)	71.90 (± 4.99)	73.72 (± 3.52)	75.04 (± 3.28)
45 %	SVM linear	81.38 (± 4.80)	85.85 (± 1.79)	84.62 (± 1.54)	**87.69**(±**1.88**)
	SVM RBF	64.15 (± 2.41)	73.54 (± 4.71)	76.92 (± 2.06)	86.00 (± 1.02)
	RLR-ℓ_1	83.85 (± 4.01)	84.00 (± 2.64)	85.85 (± 4.63)	86.00 (± 1.57)
	RLR-ℓ_2	**85.85**(±**2.78**)	**86.92**(±**2.01**)	**87.08**(±**2.64**)	85.69 (± 1.66)
	RF	59.85 (± 3.35)	82.31 (± 4.43)	72.46 (± 3.13)	85.23 (± 3.13)
	PLS-DA	75.38 (± 2.18)	72.62 (± 2.86)	72.15 (± 1.23)	71.08 (± 2.60)
40 %	SVM linear	75.97 (± 4.31)	83.60 (± 3.23)	**84.89**(±**2.69**)	**87.77**(±**2.77**)
	SVM RBF	62.45 (± 3.07)	73.09 (± 4.50)	72.52 (± 4.69)	83.45 (± 2.41)
	RLR-ℓ_1	80.72 (± 2.06)	82.16 (± 1.47)	83.88 (± 2.83)	82.73 (± 1.11)
	RLR-ℓ_2	**84.46**(±**3.48**)	**85.18**(±**3.17**)	84.60 (± 3.85)	84.32 (± 1.79)
	RF	56.69 (± 1.95)	80.29 (± 4.50)	70.36 (± 3.17)	83.74 (± 2.93)
	PLS-DA	76.69 (± 2.75)	72.52 (± 1.79)	72.81 (± 1.32)	70.22 (± 1.62)
35 %	SVM linear	69.74 (± 7.38)	**80.52**(±**5.15**)	80.00 (± 3.22)	**83.77**(±**2.63**)
	SVM RBF	56.23 (± 3.09)	68.05 (± 4.01)	68.31 (± 3.86)	80.39 (± 2.07)
	RLR-ℓ_1	77.92 (± 4.11)	77.79 (± 3.37)	80.00 (± 4.78)	79.74 (± 3.35)
	RLR-ℓ_2	**82.08**(±**2.80**)	78.96 (± 3.55)	**82.47**(±**3.36**)	81.69 (± 2.07)
	RF	53.25 (± 3.05)	77.27 (± 3.15)	67.27 (± 2.12)	80.52 (± 2.17)
	PLS-DA	75.45 (± 3.42)	69.48 (± 2.63)	70.52 (± 2.12)	68.70 (± 1.71)
30 %	SVM linear	70.42 (± 3.08)	**79.52**(±**5.22**)	79.64 (± 1.78)	**84.48**(±**1.82**)
	SVM RBF	55.39 (± 5.74)	67.03 (± 4.17)	68.61 (± 3.48)	80.73 (± 1.50)
	RLR-ℓ_1	78.30 (± 2.08)	74.91 (± 7.86)	77.94 (± 3.77)	78.79 (± 6.37)
	RLR-ℓ_2	**80.85**(±**2.98**)	77.33 (± 9.20)	**81.94**(±**3.42**)	81.70 (± 4.01)
	RF	54.30 (± 1.86)	76.97 (± 4.58)	68.00 (± 0.97)	79.88 (± 3.33)
	PLS-DA	72.00 (± 3.54)	69.09 (± 4.58)	68.73 (± 3.20)	68.48 (± 4.85)
25 %	SVM linear	65.65 (± 4.57)	**74.69**(±**2.46**)	74.46 (± 2.33)	**80.45**(±**2.49**)
	SVM RBF	52.54 (± 5.26)	60.45 (± 5.24)	63.28 (± 4.33)	78.42 (± 3.36)
	RLR-ℓ_1	75.59 (± 2.49)	71.98 (± 3.33)	75.25 (± 4.25)	75.25 (± 4.92)
	RLR-ℓ_2	**77.74**(±**3.81**)	72.99 (± 6.61)	**79.77**(±**3.79**)	77.63 (± 2.52)
	RF	52.66 (± 4.40)	73.79 (± 1.41)	65.42 (± 1.69)	77.40 (± 2.34)
	PLS-DA	71.53 (± 0.92)	69.72 (± 3.96)	70.40 (± 2.44)	70.40 (± 4.18)

Table 19. Vegetation types identification accuracies (overall accuracy (± standard deviation) in %) on [750–1350 nm].

Training size	Classifier	Overall accuracy (± Standard deviation) (%)			
		Spectral signature	Second derivative	First derivative	Continuum Removed Derivative Reflectance
50 %	SVM linear	83.31 (± 1.10)	89.09 (± 2.05)	90.91 (± 1.38)	84.13 (± 2.42)
	SVM RBF	57.69 (± 4.03)	79.34 (± 4.37)	87.60 (± 2.34)	78.68 (± 2.93)
	RLR-ℓ_1	**90.41**(±**2.00**)	88.76 (± 2.19)	89.92 (± 1.42)	87.44 (± 2.42)
	RLR-ℓ_2	86.28 (± 3.25)	**91.07**(±**1.42**)	**94.88**(±**1.10**)	**90.91**(±**2.45**)
	RF	53.88 (± 2.05)	86.28 (± 1.70)	79.83 (± 1.44)	80.66 (± 1.53)
	PLS-DA	77.52 (± 2.30)	73.72 (± 1.91)	77.69 (± 2.96)	70.74 (± 2.84)
45 %	SVM linear	78.15 (± 5.43)	84.15 (± 1.86)	86.31 (± 4.17)	82.77 (± 3.85)
	SVM RBF	59.54 (± 2.21)	72.77 (± 3.82)	82.77 (± 4.20)	75.85 (± 2.31)
	RLR-ℓ_1	**86.46**(±**3.46**)	85.38 (± 3.67)	87.69 (± 2.43)	82.92 (± 1.78)
	RLR-ℓ_2	85.23 (± 3.49)	**85.69**(±**2.86**)	**90.46**(±**2.46**)	**85.85**(±**1.58**)
	RF	53.54 (± 1.79)	80.15 (± 2.73)	76.77 (± 3.87)	77.54 (± 2.20)
	PLS-DA	73.54 (± 3.97)	70.46 (± 2.31)	74.15 (± 3.56)	68.15 (± 3.53)
40 %	SVM linear	77.70 (± 5.46)	80.72 (± 3.98)	83.88 (± 3.82)	80.43 (± 6.11)
	SVM RBF	58.85 (± 2.20)	69.64 (± 4.20)	80.29 (± 3.04)	72.95 (± 1.62)
	RLR-ℓ_1	**85.32**(±**3.88**)	84.46 (± 3.60)	88.06 (± 3.24)	81.29 (± 2.91)
	RLR-ℓ_2	82.88 (± 2.25)	**86.19**(±**2.38**)	**89.64**(±**3.39**)	**82.73**(±**3.83**)
	RF	53.24 (± 2.61)	77.99 (± 2.75)	74.96 (± 3.29)	73.96 (± 3.48)
	PLS-DA	72.09 (± 1.54)	72.09 (± 2.89)	74.96 (± 3.07)	68.35 (± 3.61)
35 %	SVM linear	72.86 (± 4.33)	78.44 (± 4.81)	80.65 (± 4.47)	75.84 (± 2.83)
	SVM RBF	55.06 (± 2.03)	67.14 (± 4.69)	76.23 (± 3.50)	66.88 (± 2.87)
	RLR-ℓ_1	**80.39**(±**3.71**)	79.22 (± 3.60)	84.55 (± 2.89)	73.90 (± 3.27)
	RLR-ℓ_2	78.57 (± 3.46)	**82.86**(±**5.61**)	**87.27**(±**3.22**)	**78.57**(±**4.19**)
	RF	52.99 (± 2.08)	73.64 (± 2.89)	73.51 (± 3.00)	69.61 (± 3.14)
	PLS-DA	70.65 (± 2.80)	70.52 (± 2.92)	72.47 (± 3.66)	66.23 (± 2.82)
30 %	SVM linear	74.18 (± 1.70)	80.48 (± 3.37)	81.58 (± 2.83)	75.39 (± 2.53)
	SVM RBF	55.27 (± 2.93)	70.06 (± 3.81)	76.24 (± 4.72)	67.39 (± 7.39)
	RLR-ℓ_1	**80.97**(±**1.19**)	79.88 (± 2.61)	84.73 (± 3.05)	76.12 (± 1.61)
	RLR-ℓ_2	80.00 (± 3.49)	**83.88**(±**3.38**)	**87.64**(±**3.31**)	**78.79**(±**2.06**)
	RF	52.00 (± 1.69)	74.42 (± 2.58)	73.21 (± 2.61)	70.55 (± 2.35)
	PLS-DA	72.36 (± 3.69)	70.06 (± 4.35)	73.45 (± 3.31)	64.48 (± 0.82)
25 %	SVM linear	67.80 (± 3.52)	75.48 (± 2.59)	78.19 (± 1.37)	73.11 (± 0.68)
	SVM RBF	53.11 (± 2.20)	60.90 (± 3.90)	69.94 (± 3.63)	66.78 (± 2.98)
	RLR-ℓ_1	75.14 (± 3.31)	77.29 (± 2.93)	80.90 (± 2.46)	72.77 (± 1.65)
	RLR-ℓ_2	**76.84**(±**2.88**)	**78.87**(±**3.46**)	**83.05**(±**4.55**)	**76.95**(±**2.66**)
	RF	48.59 (± 4.14)	71.64 (± 3.87)	73.11 (± 2.04)	69.83 (± 2.36)
	PLS-DA	70.62 (± 2.70)	69.83 (± 0.68)	72.09 (± 2.28)	63.95 (± 3.12)

Table 20. Vegetation types identification accuracies (overall accuracy (± standard deviation) in %) in [350–1350 nm].

Training size	Classifier	Overall accuracy (± Standard deviation) (%)			
		Spectral signature	Second derivative	First derivative	Continuum Removed Derivative Reflectance
50 %	SVM linear	83.47 (± 2.77)	93.22 (± 0.96)	92.40 (± 1.42)	91.57 (± 2.24)
	SVM RBF	69.75 (± 2.98)	55.04 (± 4.10)	76.20 (± 4.66)	78.02 (± 1.53)
	RLR-ℓ_1	89.26 (± 1.65)	92.73 (± 1.69)	94.05 (± 2.63)	90.41 (± 1.34)
	RLR-ℓ_2	**91.07**(±**3.56**)	**94.05**(±**1.32**)	**96.36**(±**2.00**)	**94.05**(±**1.76**)
	RF	69.75 (± 2.80)	90.25 (± 1.91)	85.45 (± 1.44)	89.26 (± 2.45)
	PLS-DA	78.51 (± 2.45)	80.83 (± 2.05)	81.49 (± 2.80)	79.17 (± 2.24)
45 %	SVM linear	80.15 (± 4.02)	87.38 (± 2.15)	88.62 (± 3.05)	91.54 (± 1.61)
	SVM RBF	65.69 (± 3.91)	49.38 (± 3.87)	67.54 (± 4.70)	72.77 (± 2.31)
	RLR-ℓ_1	86.31 (± 3.49)	90.46 (± 1.43)	90.15 (± 3.01)	88.62 (± 0.58)
	RLR-ℓ_2	**90.15**(±**3.35**)	**92.15**(±**2.09**)	**92.77**(±**1.73**)	**91.85**(±**2.21**)
	RF	65.54 (± 3.99)	85.85 (± 3.25)	81.54 (± 3.08)	86.31 (± 4.28)
	PLS-DA	78.15 (± 1.79)	79.85 (± 3.17)	79.69 (± 2.04)	76.92 (± 1.54)
40 %	SVM linear	77.55 (± 3.71)	86.76 (± 1.62)	88.49 (± 3.44)	**89.93**(±**4.07**)
	SVM RBF	63.31 (± 3.37)	50.79 (± 3.60)	66.76 (± 5.62)	69.35 (± 3.24)
	RLR-ℓ_1	83.17 (± 1.91)	88.06 (± 1.33)	89.64 (± 1.33)	85.04 (± 3.26)
	RLR-ℓ_2	**87.48**(±**2.79**)	**91.22**(±**0.95**)	**91.80**(±**1.41**)	89.64 (± 1.96)
	RF	64.60 (± 2.51)	84.46 (± 3.17)	80.86 (± 2.64)	85.32 (± 4.70)
	PLS-DA	77.99 (± 1.68)	80.00 (± 2.00)	79.42 (± 1.33)	76.40 (± 1.24)
35 %	SVM linear	68.05 (± 5.02)	83.90 (± 3.77)	84.16 (± 2.68)	85.58 (± 2.74)
	SVM RBF	59.61 (± 3.06)	44.03 (± 3.37)	63.12 (± 4.81)	64.03 (± 3.69)
	RLR-ℓ_1	80.52 (± 2.25)	85.71 (± 2.79)	85.32 (± 2.04)	80.52 (± 5.08)
	RLR-ℓ_2	**84.68**(±**2.83**)	**85.97**(±**3.71**)	**89.09**(±**1.99**)	**87.27**(±**3.73**)
	RF	63.25 (± 2.42)	80.26 (± 3.33)	77.92 (± 1.74)	82.21 (± 3.35)
	PLS-DA	75.58 (± 1.86)	76.36 (± 2.65)	79.61 (± 1.95)	75.19 (± 1.04)
30 %	SVM linear	72.61 (± 1.93)	84.61 (± 3.22)	85.58 (± 1.97)	83.76 (± 4.10)
	SVM RBF	60.24 (± 2.62)	42.42 (± 3.36)	62.79 (± 7.09)	65.21 (± 3.08)
	RLR-ℓ_1	80.48 (± 2.11)	82.55 (± 4.01)	85.58 (± 2.95)	83.03 (± 4.29)
	RLR-ℓ_2	**84.12**(±**4.12**)	**87.39**(±**4.76**)	**89.70**(±**4.22**)	**86.30**(±**4.48**)
	RF	65.21 (± 3.31)	79.52 (± 4.22)	77.21 (± 1.98)	81.58 (± 3.08)
	PLS-DA	76.24 (± 3.37)	76.85 (± 4.99)	77.58 (± 4.20)	74.79 (± 3.27)
25 %	SVM linear	70.28 (± 2.44)	80.90 (± 2.16)	83.73 (± 2.75)	82.94 (± 2.59)
	SVM RBF	51.64 (± 1.54)	39.89 (± 1.91)	52.54 (± 2.84)	61.58 (± 2.34)
	RLR-ℓ_1	77.40 (± 1.96)	**82.15**(±**3.64**)	**83.95**(±**1.70**)	79.66 (± 2.02)
	RLR-ℓ_2	**81.47**(±**1.10**)	80.79 (± 4.42)	83.16 (± 6.33)	**83.84**(±**3.17**)
	RF	62.03 (± 3.86)	76.16 (± 3.20)	76.84 (± 1.86)	80.45 (± 3.67)
	PLS-DA	75.93 (± 2.74)	74.58 (± 2.88)	78.76 (± 2.28)	72.66 (± 2.49)

Table 21. Vegetation types identification accuracies (overall accuracy (± standard deviation) in %) on [350–2500 nm].

Training size	Classifier	Overall accuracy (± Standard deviation) (%)			
		Spectral signature	Second derivative	First derivative	Continuum Removed Derivative Reflectance
50 %	SVM linear	83.47 (± 2.34)	85.29 (± 4.10)	**87.44**(±**1.21**)	**91.90**(±**1.76**)
	SVM RBF	61.98 (± 4.31)	19.34 (± 5.95)	22.81 (± 0.40)	25.12 (± 0.84)
	RLR-ℓ_1	91.07 (± 2.30)	82.31 (± 3.16)	83.80 (± 3.07)	88.26 (± 1.60)
	RLR-ℓ_2	**91.57**(±**1.42**)	81.49 (± 2.37)	82.81 (± 2.05)	84.79 (± 2.37)
	RF	71.24 (± 2.63)	**89.92**(±**1.98**)	84.96 (± 2.42)	90.58 (± 0.40)
	PLS-DA	75.04 (± 2.05)	78.35 (± 4.91)	75.70 (± 2.98)	79.83 (± 0.84)
45 %	SVM linear	79.08 (± 1.32)	79.38 (± 1.57)	**82.31**(±**1.61**)	**90.62**(±**1.78**)
	SVM RBF	55.38 (± 6.10)	22.31 (± 0.00)	22.46 (± 0.31)	24.15 (± 1.58)
	RLR-ℓ_1	85.23 (± 2.25)	79.69 (± 2.86)	81.08 (± 2.56)	84.77 (± 2.89)
	RLR-ℓ_2	**86.00**(±**2.73**)	79.23 (± 2.33)	79.54 (± 2.36)	77.69 (± 3.61)
	RF	69.08 (± 4.42)	**85.08**(±**2.46**)	80.92 (± 1.32)	87.69 (± 2.96)
	PLS-DA	73.08 (± 3.34)	75.23 (± 4.31)	72.00 (± 3.29)	77.69 (± 1.88)
40 %	SVM linear	76.12 (± 0.84)	79.42 (± 0.86)	**82.30**(±**2.35**)	**88.06**(±**1.68**)
	SVM RBF	53.24 (± 3.61)	23.02 (± 0.00)	23.45 (± 0.58)	25.18 (± 1.02)
	RLR-ℓ_1	83.88 (± 3.69)	79.28 (± 1.79)	79.86 (± 3.83)	82.59 (± 3.98)
	RLR-ℓ_2	**84.75**(±**2.86**)	81.01 (± 3.11)	79.57 (± 2.35)	79.28 (± 3.57)
	RF	65.90 (± 3.48)	**84.17**(±**3.34**)	79.28 (± 2.67)	86.04 (± 2.60)
	PLS-DA	73.67 (± 1.85)	74.39 (± 2.07)	71.94 (± 3.75)	76.55 (± 4.31)
35 %	SVM linear	69.74 (± 1.13)	77.27 (± 1.09)	**79.87**(±**1.79**)	**84.42**(±**4.35**)
	SVM RBF	49.87 (± 3.64)	20.00 (± 5.45)	20.13 (± 5.53)	22.21 (± 4.69)
	RLR-ℓ_1	82.47 (± 3.74)	74.42 (± 2.38)	76.23 (± 2.04)	78.05 (± 1.26)
	RLR-ℓ_2	**83.64**(±**3.19**)	77.27 (± 2.87)	77.14 (± 1.99)	74.94 (± 2.80)
	RF	64.03 (± 3.01)	**79.35**(±**2.83**)	77.27 (± 1.23)	82.47 (± 2.82)
	PLS-DA	71.95 (± 2.19)	72.34 (± 2.27)	70.65 (± 3.57)	74.42 (± 3.20)
30 %	SVM linear	69.94 (± 3.90)	77.33 (± 1.82)	**79.64**(±**2.59**)	**84.36**(±**5.88**)
	SVM RBF	48.85 (± 4.05)	22.42 (± 0.00)	22.42 (± 0.00)	24.12 (± 0.89)
	RLR-ℓ_1	79.39 (± 2.24)	71.27 (± 3.29)	76.36 (± 3.27)	78.06 (± 5.44)
	RLR-ℓ_2	**83.27**(±**3.48**)	75.88 (± 4.64)	75.52 (± 3.03)	75.15 (± 4.11)
	RF	65.21 (± 3.83)	**78.06**(±**2.22**)	77.21 (± 2.67)	80.00 (± 4.25)
	PLS-DA	70.18 (± 2.80)	71.27 (± 3.61)	68.85 (± 4.67)	73.45 (± 2.58)
25 %	SVM linear	65.31 (± 4.24)	74.24 (± 1.54)	**77.51**(±**1.49**)	**83.05**(±**3.29**)
	SVM RBF	43.05 (± 1.31)	22.60 (± 0.00)	22.60 (± 0.00)	24.07 (± 0.58)
	RLR-ℓ_1	74.92 (± 1.70)	67.46 (± 3.44)	71.64 (± 2.35)	75.03 (± 5.27)
	RLR-ℓ_2	**80.23**(±**0.80**)	73.79 (± 3.57)	74.35 (± 2.19)	70.73 (± 1.84)
	RF	62.49 (± 4.15)	**74.58**(±**2.14**)	76.61 (± 2.22)	79.10 (± 2.95)
	PLS-DA	70.17 (± 1.40)	70.96 (± 4.00)	70.06 (± 3.24)	72.43 (± 2.64)

for RF applied on the spectral signature. In this case, considering the whole spectral range improves the result by 1 % compared with [350–1350 nm].

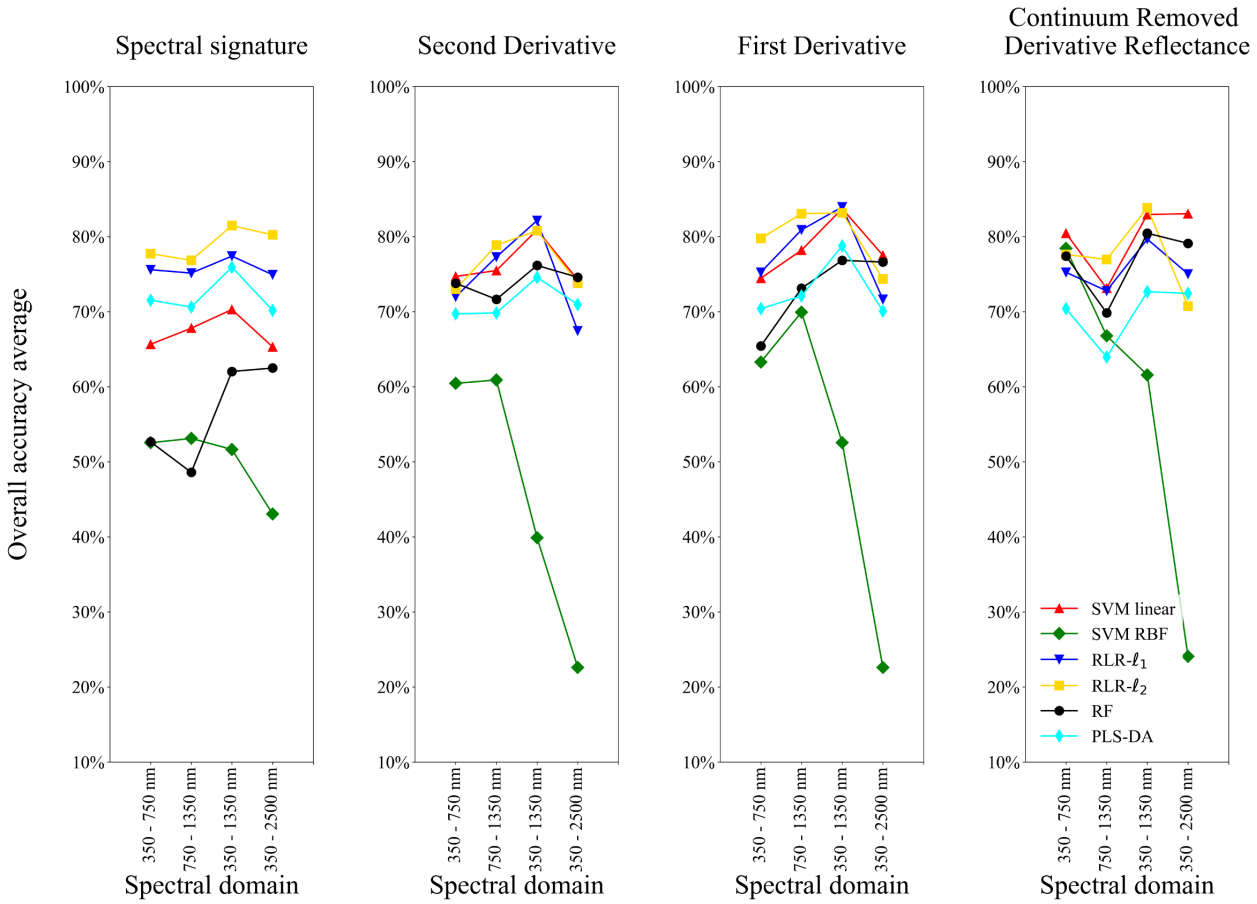


Figure 15. Vegetation type identification accuracies with the training size = 25 %.

Considering RLR-ℓ_2 in [350–1350 nm], Table 22 shows that the best overall accuracies are given by first derivative, second derivative and CRDR. First and second derivatives overall accuracies are very close (difference lower than 1 %). However, those transformations are sensitive to noise. But CRDR delivered better results than spectral signatures and similar performances than the first and second derivatives (difference is lower than 4 %). As mentioned in Section 4.1, those transformations are closely related to absorption features rather than reflectance magnitude [38], and are helpful to discriminate peatland vegetation types which are clearly characterized by different biophysical components as mentioned in Section 4.2.

Considering RLR, ℓ_1 regularization, which controls the selection or the removal of variables, always underperforms ℓ_2-regularization, which handles with collinear variables [16]. Because of mixed plant species, it is difficult to remove variables that are not involved in the classification of all the vegetation types. Although, SVM and RF are popular classifiers in remote sensing community, they are outclassed by RLR in [350 nm to 1350 nm] which is the spectral range where results are the best (Figure 16). Results given by SVM RBF are lower than those obtained with RLR and can be explained by the difficulty to find adapted parameters considering this high dimensionality problem. However, it is interesting to note that results from SVM linear are close to RLR ones considering first derivative, second derivative and CRDR. Further investigations should be conducted to better understand the link between those classifiers and improve the choice of the parameters. Figure 16

Table 22. Vegetation types identification accuracies (overall accuracy (± standard deviation) in %) on [350–1350nm] for RLR-ℓ_2.

Training size	Overall accuracy (± Standard deviation) (%)					
	Spectral signature	Second derivative	First derivative	Continuum Removal	Continuum Removed Derivative Reflectance	log transformation
50 %	91.07 (± 3.56)	94.05 (± 1.32)	**96.36**(±**2.00**)	89.59 (± 1.93)	94.05 (± 1.76)	93.72 (± 2.13)
45 %	90.31 (± 3.39)	92.15 (± 2.09)	**92.77**(±**1.73**)	87.85 (± 2.59)	91.85 (± 2.21)	89.69 (± 4.03)
40 %	87.48 (± 2.79)	91.22 (± 0.95)	**91.80**(±**1.41**)	83.31 (± 3.79)	89.64 (± 1.96)	88.35 (± 2.15)
35 %	84.68 (± 2.83)	85.97 (± 3.71)	**89.09**(±**1.99**)	81.56 (± 3.45)	87.27 (± 3.73)	86.23 (± 3.45)
30 %	84.24 (± 4.07)	87.39 (± 4.76)	**89.70**(±**4.22**)	82.79 (± 4.09)	86.30 (± 4.48)	84.36 (± 4.22)
25 %	81.47 (± 1.10)	80.79 (± 4.42)	83.16 (± 6.33)	80.45 (± 2.62)	**83.84**(±**3.17**)	82.15 (± 2.13)

shows that PLS-DA is the least sensitive classifier to training size regardless transformed spectral signatures in [350–1350 nm].

Table 23 shows that *Pinguicula* sp. (PING) has the lowest F1-score (66.67 % and 56.00 % respectively) as well as for the spectral vegetation indices (Section 4.2). Besides, this vegetation type is hardly discriminated from the other ones (Producer's accuracy (PA) = 53.33 %) and some *Pinguicula* sp. combined vegetation (PI_CV) spectra are classified as PING). However, it should be kept in mind that PING has a small number of spectra. Considering Aquatic type b (AQ_B) which has about the same number of spectra (7 spectra against 8 for PING), User's Accuracy (UA) = 60.98 % and some Aquatic type a (AQ_A) spectra are predicted as AQ_B ones. These poor UA results compared to one obtained by spectral vegetation indices can not be explained by the spectral domain. Indeed, the best spectra vegetation index (NDWI[860, 1240]) that discriminate AQ_A from AQ_B has both wavelengths in [350–1350 nm]. However, this result may be qualified by PA. Indeed, on [350–1350 nm] domain, UA = 100.00 % whereas UA = 84.60 % for spectral vegetation indices. Nevertheless, using a continuous spectral domain can lead to worse results for other vegetation types such as *Sphagnum* sp. (SPHA), *Calluna vulgaris* (CAVU), AQ_A: F1-score is always better considering the same classifier (RLR-ℓ_2) applied on spectral vegetation indices selected by the Hellinger distance (SPHA: 91.12 % *vs* 82.80 %; CAVU: 77.62 % *vs* 71.43 %; AQ_A: 86.80 % *vs* 82.81 %). Considering SPHA, if PA = 90.59 % for spectral vegetation indices or for [350–1350 nm], the latter predicts more SPHA than observed (UA = 76.24 %) and is more confused with CAVU. This can be explained by plot 7 which is mainly composed of *Calluna vulgaris* (20 %), *Carex rostrata* (25 %), *Molinia caerulea* ssp. *caerulae* (20 %) and *Sphagnum palustre* (20 %) (Appendix B).

In our case, reducing feature space by selecting most discriminant wavelengths (using PCA or MNF) has not been implemented, whereas it can be an interesting track to explore to see if it improves results for RLR-ℓ_2. *Juniperus communis* (JUCO), *Eleocharis quinqueflora* (ELQU) and Aquatic type c (AQ_C) have about the same F1-score considering spectral vegetation indices or [350–1350 nm]: less than 2 % difference. However, they have better PA on the continuous spectral range (PA = 100.00 % for JUCO; 95.56 % for AQ_C) which means that this spectral range contains discriminant wavelengths able to catch characteristic of those vegetation types.

Rhododendron ferrugineum (RHFR), *Carex* sp. homogeneous vegetation (CA_HV), *Salix* sp. (SALI) and *Menyanthes trifoliata* (METR) have better results considering [350–1350 nm]. This can be explained by the fact that the spectral vegetation indices used have not been built for that kind of vegetation types. Further investigations can be lead to find specific indices that can discriminate those vegetation types from other ones.

Table 23. Confusion matrix of the RLR-ℓ_2 classification using CRDR on [350–1350 nm] (training size = 25 %). The producer's and user's accuracies, the overall accuracy and the F1-score are also shown.

	SPHA	CAVU	RH_FR	CA_HV	AQ_A	SALI	PING	JUQO	ELQU	METR	PI_CV	AQ_B	AQ_C	Producer's accuracy (%)
SPHA	**15.40**	1.40	0.00	0.20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	90.59
CAVU	3.20	**7.00**	0.00	0.00	0.00	0.00	0.20	0.00	0.20	0.00	0.40	0.00	0.00	63.64
RHFR	1.40	0.00	**8.20**	0.20	0.00	0.40	0.20	0.60	0.00	0.00	0.00	0.00	0.00	74.55
CA_HV	0.00	0.00	0.00	**16.00**	1.40	0.20	0.00	0.20	0.00	0.20	2.00	0.00	0.00	80.00
AQ_A	0.20	0.00	0.00	1.80	**31.80**	0.00	0.00	0.20	0.40	1.40	0.20	3.20	0.80	79.50
SALI	0.00	0.00	0.20	0.40	0.20	**11.80**	0.00	0.40	0.00	0.00	0.00	0.00	0.00	90.77
PING	0.00	0.20	0.00	0.40	0.40	0.00	**3.20**	0.00	0.40	0.00	1.40	0.00	0.00	53.33
JUCO	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**14.00**	0.00	0.00	0.00	0.00	0.00	**100.00**
ELQU	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**11.00**	0.00	0.00	0.00	0.00	**100.00**
METR	0.00	0.00	0.00	0.00	2.60	0.00	0.00	0.00	0.60	**5.80**	0.00	0.00	0.00	64.44
PI_CV	0.00	0.00	0.00	0.40	0.00	0.00	0.00	0.00	0.00	0.00	**10.60**	0.00	0.00	96.36
AQ_B	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**5.00**	0.00	**100.00**
AQ_C	0.00	0.00	0.00	0.00	0.40	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**8.60**	95.56
User's accuracy (%)	76.24	81.40	**97.62**	82.47	86.41	95.16	88.89	90.91	87.30	78.38	72.60	60.98	91.49	OAA: **83.84**
F1-score (%)	82.80	71.43	84.54	81.22	82.81	92.91	66.67	**95.24**	93.22	70.73	82.81	75.76	93.48	

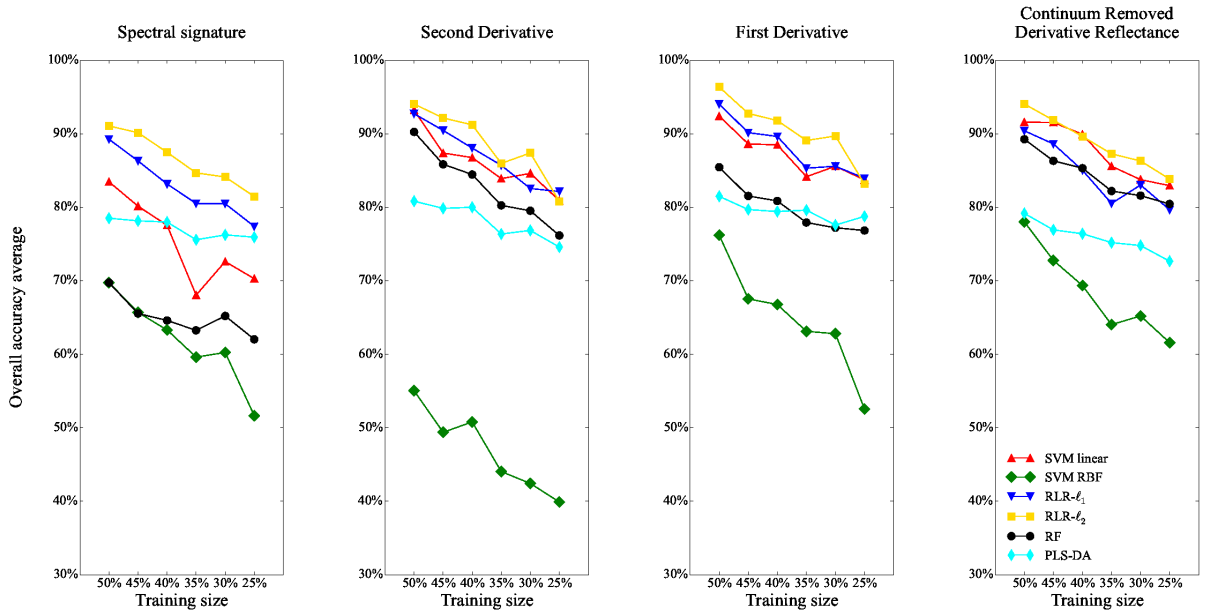


Figure 16. Vegetation type identification accuracies on [350–1350 nm].

5. Conclusions and Perspectives

This study aimed at inventorying and evaluating the performance of discrimination techniques for peatland habitats based on *in situ* hyperspectral measurements with a high spectral resolution and high signal-to-noise ratio. To evaluate the potential of hyperspectral data to separate and classify those habitats, three classes of methods were investigated and compared:

- similarity measures calculated on spectral reflectance,
- supervised classification based on "local" information (spectral vegetation indices),
- supervised classification based on "global" information (spectral ranges).

This study demonstrated that peatland vegetation types could be discriminated using the Canberra distance on the whole spectral range [350–2500 nm]. This distance is sensitive to a small change when both coordinates approach zero which is the case of reflectance especially in the visible ranges and in the SWIR (Figure 2). Further investigations should be conducted to see if combinations of spectral range can improve overall accuracy or if the lack of spectral signatures in the reference database (which is a weakness of this method) may explain why the whole spectral range is needed to compare spectra in that case. Besides, it is of importance to collect more spectral signatures from peatland vegetation types to build a spectral reference database of peatland vegetation types that can catch more spectral variability.

Although, there is no spectral vegetation indices built to discriminate peatland vegetation types, this study showed that some indices could be selected using the Hellinger distance. Although those indices have not been built to discriminate peatland vegetation types, they were able to classify them because they focus on biochemical properties such as chlorophyll, nitrogen, water stress, ... Further investigations have to be done to see the impact of spectral bandwidth around the wavelength of selected indices instead of working with one particular wavelength. For instance there are lots of indices that catch the same biochemical property but wavelengths of interest change because they focus on specific plant species (*e.g.* for the chlorophyll, SR[700, 670] is built for field corn, whereas SR[675, 700] is built for soy beans leaves; contrary to SR[675, 700], SR[700,670] has been selected with the Hellinger distance).

Contrary to similarity measures which had best results considering the whole spectral range, supervised classification on specific spectral range as defined by [31] achieved best overall accuracy considering [350–1350 nm] domain. This is in agreement with the spectral vegetation indices: only 4 indices (NDWI[860, 1240], NDWI[860, 2130], NDWI[1110, 1450], MSI) over the 26 selected have a discriminant wavelength which is not in this spectral range. More precisely, the discriminant wavelength is located in the SWIR and all concerned vegetation indices are linked to the water status. Further investigations should be conducted on the extraction or the reduction of features of this spectral range to understand why this domain gave sometimes worse results than spectral vegetation indices depending on the vegetation type.

Among the three methods, the best results are obtained considering a specific spectral domain [350–1350 nm with RLR regardless the transformed spectral signatures and the size of the training size (overall accuracy ranges from 81.47 % to 96.36 %). However, it should be of interest to apply feature reduction methods usually applied on remote sensing (such as PCA or MNF) to see it results are improved or specific spectral wavelength can be selected.

To our knowledge, although not popular in remote sensing for classifying (but already used for feature selection), RLR classifier achieves best overall classification accuracy whether applied to the spectral vegetation indices selected by the Hellinger distance (77.21 %) on the [350–1350 nm] domain (83.84 %) considering training size = 25 %.

Furthermore, this study showed that CRDR gave encouraging results event if it is slightly below those obtained by the first derivative and the second derivative considering RLR classifier.

Considering the habitats, some vegetation types were more easily separated. For instance, JUCO had the best F1-score with the spectral vegetation indices selected by the Hellinger distance (94.83 %) or on the [350–1350 nm] (95.24 %) with RLR and the training size = 25 %. In some case this specific spectral domain gave better results (F1-score = 92.21 % whereas with spectral vegetation indices F1-score = 64.72 % for SALI) while in

other case, the spectral vegetation indices gave better results (F1-score = 91.12 % whereas F1-score = 82.80 % for SPHA). As mentioned earlier, reducing feature space have to be investigated to see if a particular feature space exists that can discriminate and classify all vegetation types or if we need to consider either spectral vegetation indices or a specific spectral domain depending on the vegetation type to classify.

Although all the results strongly depended on the current dataset, this study illustrated promising methods for classifying peatland vegetation types using *in situ* hyperspectral measurements. The next step concerns the application or adaptation of those methods to airborne hyperspectral imageries with high spatial resolution acquired on September 2014 (simultaneously with *in situ* measurements). With the objective of evaluating the benefits of airborne or spaceborne sensors with a lower spectral resolution a lower signal-to-noise ratio, these conclusions may change. For that purpose, some indices (involving wavelengths lower than 480 nm) will not be used because of the camera spectral range sensitivity and some transformed spectral signatures such as second derivative will neither be used because of signal-to-noise ratio. Similarly, the first derivative transformation is very sensitive to the noise coming from the instrument but also from the atmosphere correction and thus can lead to degrade its performance...

Additional imageries acquired in October 2012 and July 2013 would allow us to test these methods with spectral signatures extracted from the ancillary dataset. Multi-temporal analysis could also be conducted to discriminate vegetation types thanks to the phenological changes. This step would be of interest to evaluate the robustness of spectral measurements, spectral vegetation indices and classifiers selected previously from *in situ* hyperspectral measurements to airborne data.

Acknowledgments: The authors would like to thank Rosa Oltra-Carrió and Olivier Vaudelin for their help with field measurements and acknowledge the LabEx DRIIHM and the Observatoire Hommes-Milieux (OHM-CNRS) Haut-Vicdessos for funding and supporting the study.

Appendix A. Composition of vegetation types

Table A.1. Presence (+) and actual cover percentage of plant species collected on Bernadouze peatbog (Ariège, France) by Florence MAZIER & Nicolas DE MUNIK (09/04/2014 & 09/11/2014).

Plant Species / Plots	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Code	SPHA	SPHA	SPHA	SPHA	SPHA	CAVU	CAVU	ELQU	ELQU	PING	METR	JUCO	JUCO	RHFR	RHFR	SALI	SALI	SALI	AQ_A	AQ_A	AQ_A	AQ_A	AQ_A	AQ_A
Alchemilla glabra																								
Anthoxanthum odoratum	2	2	2	1	+																			
Apiaceae																								
Bare ground							1	5	4	15														
Briza media								2		+														
Calluna vulgaris		2		5	15	70	25			+														
Caltha palustris																5	10			2				
Campylium stellatum								35																
Cardamine pratensis				+							+													
Carex demissa																								
Carex echinata	5	2	2	+	2				+	+	5								1					
Carex flava								+		+														
Carex nigra	5	2	2		2				10	5														
Carex panicea				+				+		+	5			1										
Carex paniculata																	50							
Carex rostrata																5			35	70			40	10
Carex sp						2	25																	
Circaea lutetiana													4											
Cirsium palustre													2											
Dactylorhiza masculata	2				+		+							+	1									
Drepanocladus revolvens											30												+	
Drosera rotundifolia				+	+	1					+													
Dryopteraceae													+											
Eleocharis quinqueflora								60	40	40												70		
Epikeros pyrenaeus				+		+								+										
Equisetum sp	1	+	+				+									5		5		1	30		+	+
Eriophorum angustifolium				5		10																		
Festuca rubra				3																				
Galium palustre																	+	2						
Galium saxatile											1		2			5				+				
Gentiana ciliata										+														
Hylocomium brevirostre																	+							
Hypnum cupressiforme													2											
Juncus alpinus										+														
Juncus bulbosus																								
Juncus sp																+								
Juniperus communis						5						95	80											
Lathyrus montanus						5						+												
Leotodon hispidus																								
Lotus sp				+	2																			
Luzula sp																								
Lychnis floscuculi											4					+	+							
Mentha arvensis																3				+				
Menyanthes trifoliata											10					25	5	10		10				4
Molinia caerulea ssp. caerulae	15	25	30	15	20	10	20	15	5	10	30		5	25					5					
Narthecium ossifragum							2																	
Parnassia palustris					1	4	+	1		2	3					1	+							
Pedicularis sylvatica				1	+																			
Pilosella lactucella					+					1														
Pinguicula sp									1															
Pinguicula vulgaris								+		5														
Plagiomnium elatum																+								
Plantago lanceolata																								
Polytrichum sp		2																						
Potentilla erecta	5	5	5	5	10	5	6			5		2	+	5	5									
Potentilla sp																	+							
Prunella vulgaris					+					2														
Ranunculus acris													+			1								
Rhododendron ferrugineum														80	40									
Salix atrocinerea																90	90	100						
Scorpidium sp																								
Selaginella selaginoides									+	1									1					
Sphagnum capillifolium	10	5	5	70		25																		
Sphagnum cuspidatum																			25					
Sphagnum palustre	90	75	65	10	80	20	20				8			50	80									
Sphagnum papillosum		15	25																					
Succisa pratensis							+																	
Tofieldia calyculata										+														
Tomenthypnum nitens							3		30	10														
Trichophorum cespitosum	+																							
Trifolium arvense																								
Trifolium pratense					1						1													
Utricularia sp.																						5		
Vaccinium myrtillus										+		3		30	15									
Vicia sepium														+										
Viola palustris	2													+										
Viola sp					+								5				+							
Water																			50	25	70	30	60	90

Plant Species / Plots	25	26	27	28	29	30	31
Code	AQ_B	AQ_C	CA_HV	CA_HV	CA_HV	CA_HV	PI_CV
Alchemilla glabra			2	+	3		
Anthoxanthum odoratum							
Apiaceae					+		
Bare ground		40					
Briza media			5	5			
Calluna vulgaris							
Caltha palustris		1					
Campylium stellatum							
Cardamine pratensis							
Carex demissa							
Carex echinata			2	2			
Carex flava							
Carex nigra			5	2			
Carex panicea							
Carex paniculata						100	
Carex rostrata							
Carex sp		2			60		50
Circaea lutetiana							
Cirsium palustre						5	
Dactylorhiza masculata			+				
Drepanocladus revolvens							
Drosera rotundifolia							
Dryopteraceae							
Eleocharis quinqueflora							
Epikeros pyrenaeus							
Equisetum sp		+	+				+
Eriophorum angustifolium							
Festuca rubra			1	+			10
Galium palustre							
Galium saxatile		+	1				
Gentiana ciliata							
Hylocomium brevirostre							
Hypnum cupressiforme							
Juncus alpinus							
Juncus bulbosus			1				
Juncus sp							
Juniperus communis							
Lathyrus montanus					+		
Leotodon hispidus							
Lotus sp							
Luzula sp					+		
Lychnis floscuculi		1					
Mentha arvensis						2	
Menyanthes trifoliata		50					
Molinia caerulea ssp. caerulae		4	60	70	40		50
Narthecium ossifragum							+
Parnassia palustris		2	2	2		+	1
Pedicularis sylvatica			1				
Pilosella lactucella			+	2	+		+
Pinguicula sp							3
Pinguicula vulgaris							
Plagiomnium elatum							
Plantago lanceolata					2		
Polytrichum sp							
Potentilla erecta		3	2		1		2
Potentilla sp							
Prunella vulgaris			4	5	1		1
Ranunculus acris			1	+	2		+
Rhododendron ferrugineum							
Salix atrocinerea							
Scorpidium sp						25	
Selaginella selaginoides							
Sphagnum capillifolium							
Sphagnum cuspidatum							
Sphagnum palustre							
Sphagnum papillosum							
Succisa pratensis			4				
Tofieldia calyculata							
Tomenthypnum nitens					1		
Trichophorum cespitosum							
Trifolium arvense			1				
Trifolium pratense			4	5	2		1
Utricularia sp.	80						
Vaccinium myrtillus							
Vicia sepium							
Viola palustris							
Viola sp		+					
Water	20						

Appendix B. Data from vegetation types

Sphagnum sp. (SPHA)

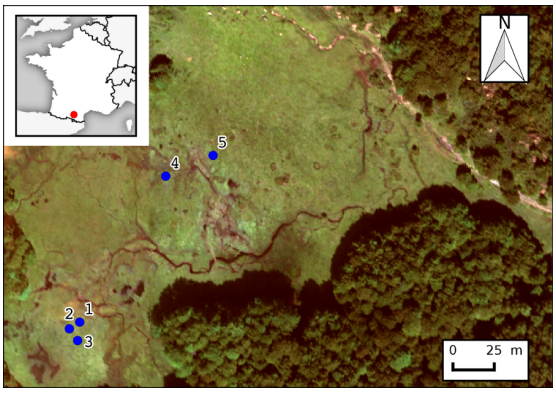


Figure B.1. Location of the *in situ* spectroradiometer measurements for the plots of *Sphagnum* sp. (SPHA).

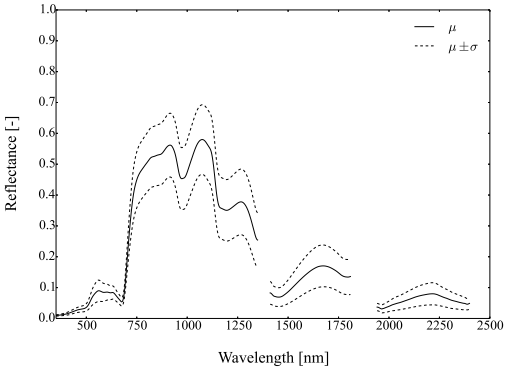
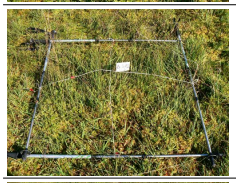


Figure B.2. Mean reflectance (μ) and standard deviation (σ) of *Sphagnum* sp. (SPHA).

Picture	Plot	Longitude (DD)	Latitude (DD)	Altitude (m)	No. of spectra
	1	1.423156	42.802105	1343.715	4
	2	1.423080	42.802068	1344.046	4
	3	1.423143	42.802005	1344.004	4
	4	1.423771	42.802907	1344.747	7
	5	1.424118	42.803025	1346.327	3

556 *Calluna vulgaris* (CAVU)

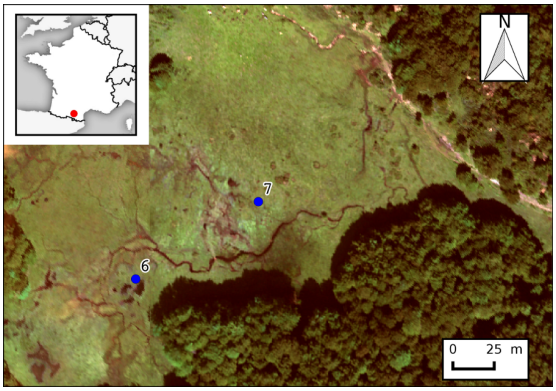


Figure B.3. Location of the *in situ* spectroradiometer measurements for the plots of *Calluna vulgaris* (CAVU).

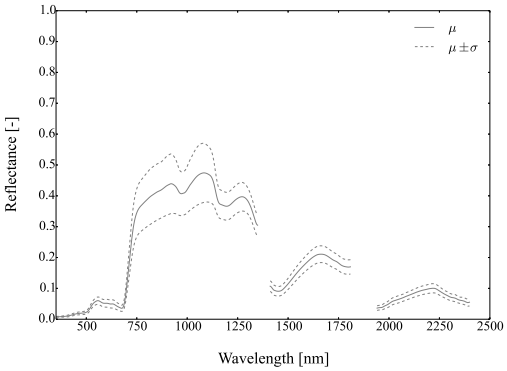
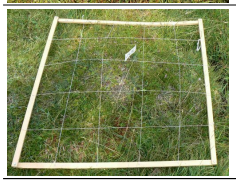


Figure B.4. Mean reflectance (μ) and standard deviation (σ) of *Calluna vulgaris* (CAVU).

Picture	Plot	Longitude (DD)	Latitude (DD)	Altitude (m)	No. of spectra
	6	1.423564	42.80234	1343.762	7
	7	1.42446	42.802773	1343.636	7

557 *Eleocharis quinqueflora* (ELQU)

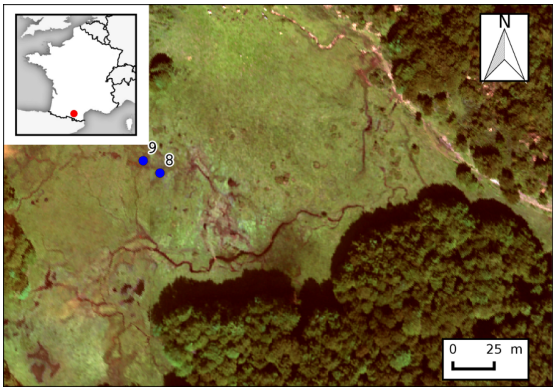


Figure B.5. Location of the *in situ* spectroradiometer measurements for the plots of *Eleocharis quinqueflora* (ELQU).

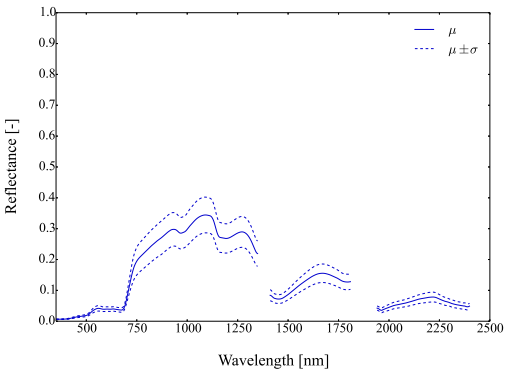
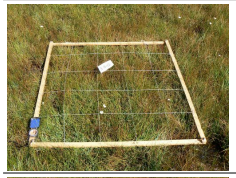
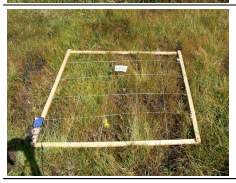


Figure B.6. Mean reflectance (μ) and standard deviation (σ) of *Eleocharis quinqueflora* (ELQU).

Picture	Plot	Longitude (DD)	Latitude (DD)	Altitude (m)	No. of spectra
	8	1.423728	42.802918	1344.617	3
	9	1.423602	42.802983	1344.650	12

558 *Pinguicula* sp. (PING)

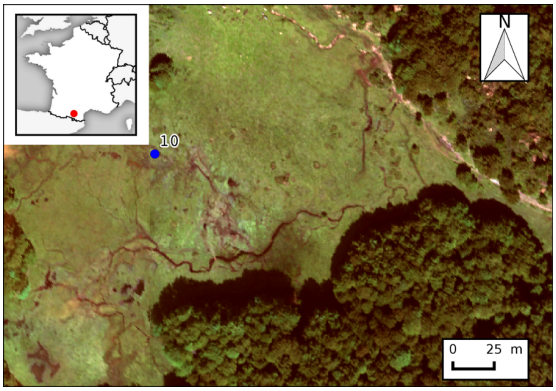


Figure B.7. Location of the *in situ* spectroradiometer measurements for the plots of *Pinguicula* sp. (PING).

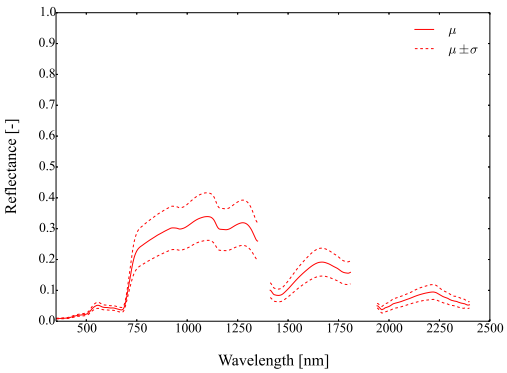


Figure B.8. Mean reflectance (μ) and standard deviation (σ) of *Pinguicula* sp. (PING).

Picture	Plot	Longitude (DD)	Latitude (DD)	Altitude (m)	No. of spectra
	10	1.423687	42.803021	1345.138	8

559 *Menyanthes trifoliata* (METR)

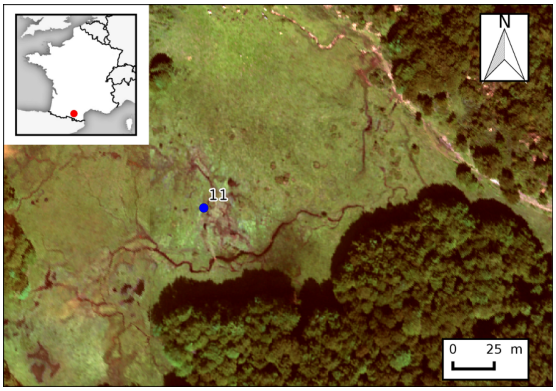


Figure B.9. Location of the *in situ* spectroradiometer measurements for the plots of *Menyanthes trifoliata* (METR).

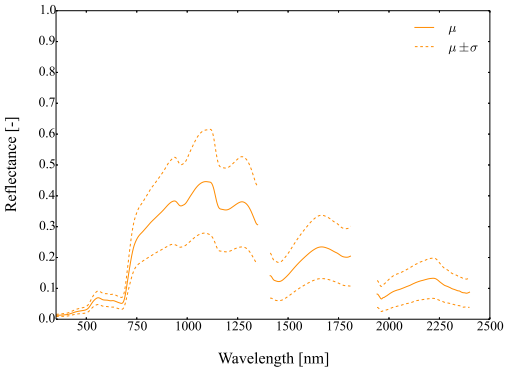


Figure B.10. Mean reflectance (μ) and standard deviation (σ) of *Menyanthes trifoliata* (METR).

Picture	Plot	Longitude (DD)	Latitude (DD)	Altitude (m)	No. of spectra
	11	1.424057	42.802733	1343.781	12

560 *Juniperus communis* (JUCO)

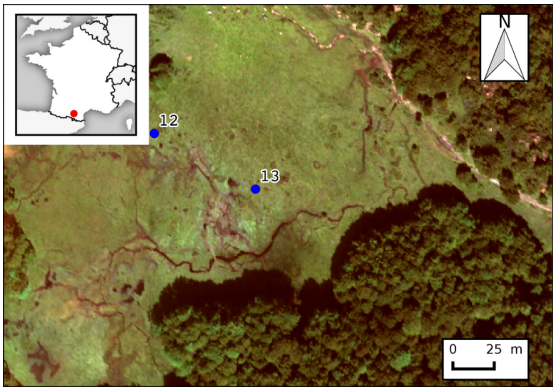


Figure B.11. Location of the *in situ* spectroradiometer measurements for the plots of *Juniperus communis* (JUCO).

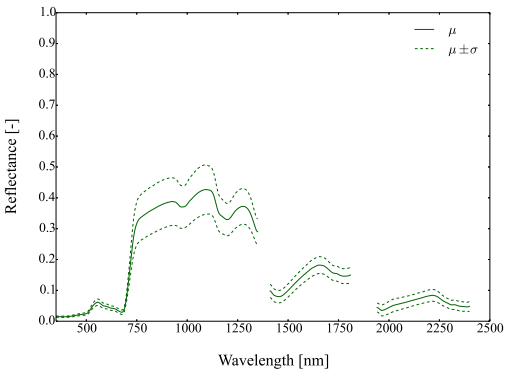


Figure B.12. Mean reflectance (μ) and standard deviation (σ) of *Juniperus communis* (JUCO).

Picture	Plot	Longitude (DD)	Latitude (DD)	Altitude (m)	No. of spectra
	12	1.42368	42.803132	1345.667	12
	13	1.424437	42.802841	1344.217	7

561 *Rhododendron ferrugineum* (RHFR)

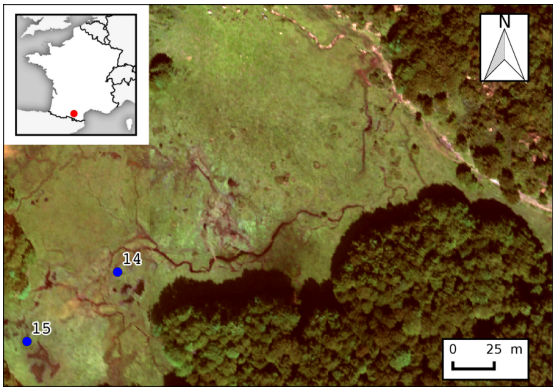


Figure B.13. Location of the *in situ* spectroradiometer measurements for the plots of *Rhododendron ferrugineum* (RHFR).

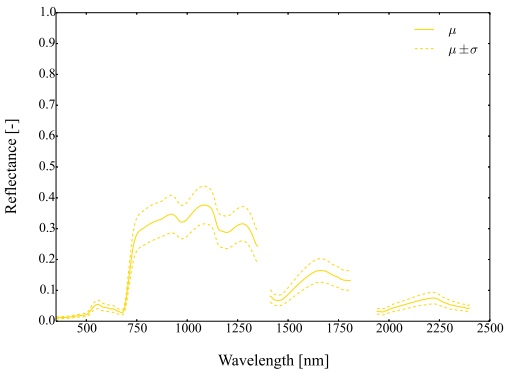
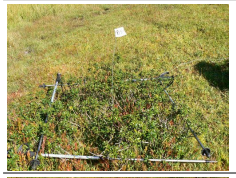


Figure B.14. Mean reflectance (μ) and standard deviation (σ) of *Rhododendron ferrugineum* (RHFR).

Picture	Plot	Longitude (DD)	Latitude (DD)	Altitude (m)	No. of spectra
	14	1.423429	42.802376	1343.301	7
	15	1.422769	42.801989	1344.606	7

562 *Salix* sp. (SALI)

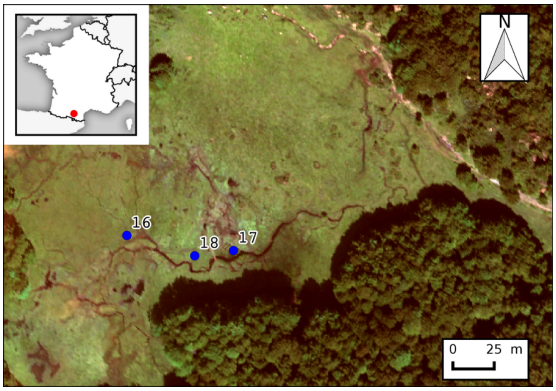


Figure B.15. Location of the *in situ* spectroradiometer measurements for the plots of *Salix* sp. (SALI).

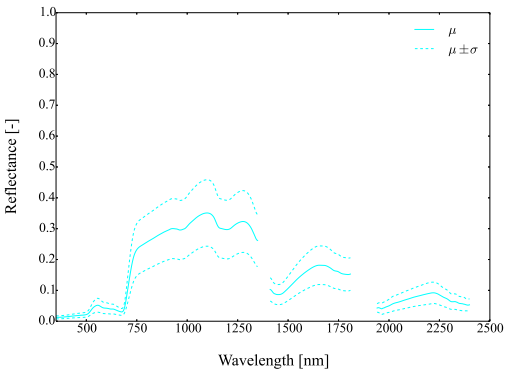


Figure B.16. Mean reflectance (μ) and standard deviation (σ) of *Salix* sp. (SALI).

Picture	Plot	Longitude (DD)	Latitude (DD)	Altitude (m)	No. of spectra
	16	1.423492	42.802575	1343.198	9
	17	1.424283	42.802505	1343.082	4
	18	1.423997	42.802472	1343.025	4

563 Aquatic type a (AQ_A)

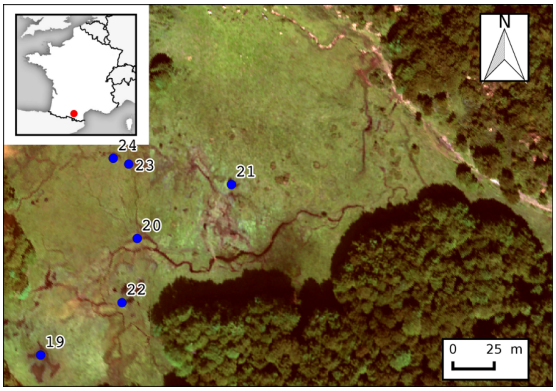


Figure B.17. Location of the *in situ* spectroradiometer measurements for the plots of Aquatic type a (AQ_A).

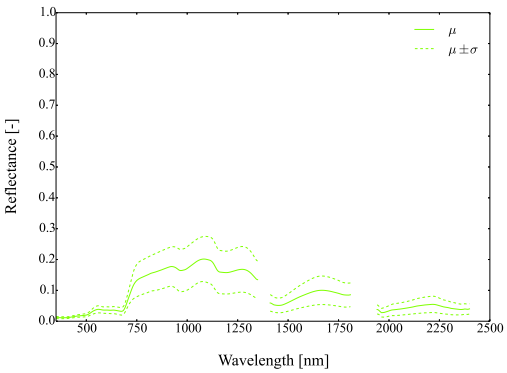
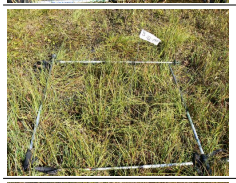
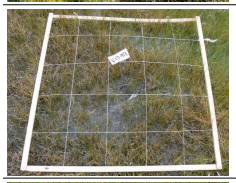


Figure B.18. Mean reflectance (μ) and standard deviation (σ) of Aquatic type a (AQ_A).

Picture	Plot	Longitude (DD)	Latitude (DD)	Altitude (m)	No. of spectra
	19	1.422872	42.801917	1344.375	7
	20	1.423569	42.80256	1343.070	12
	21	1.424258	42.802863	1344.285	6
	22	1.423466	42.80221	1343.305	4
	23	1.423495	42.802963	1344.493	12
	24	1.42338	42.802993	1344.632	12

564 Aquatic type b (AQ_B)

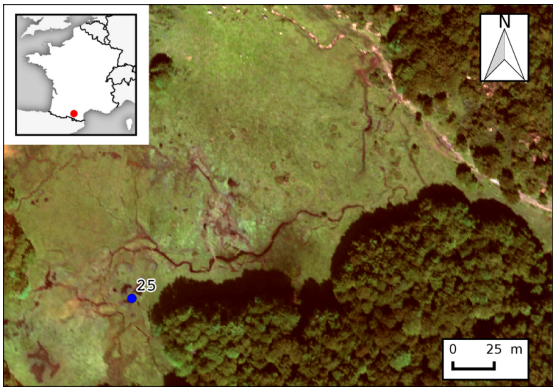


Figure B.19. Location of the *in situ* spectroradiometer measurements for the plots of Aquatic type b (AQ_B).

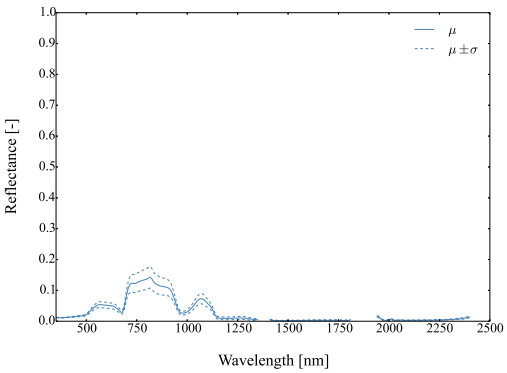


Figure B.20. Mean reflectance (μ) and standard deviation (σ) of Aquatic type b (AQ_B).

Picture	Plot	Longitude (DD)	Latitude (DD)	Altitude (m)	No. of spectra
	25	1.423539	42.802234	1343.04	7

565 Aquatic type c (AQ_C)

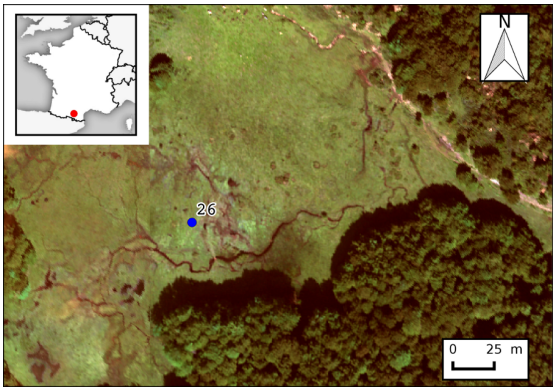


Figure B.21. Location of the *in situ* spectroradiometer measurements for the plots of Aquatic type c (AQ_C).

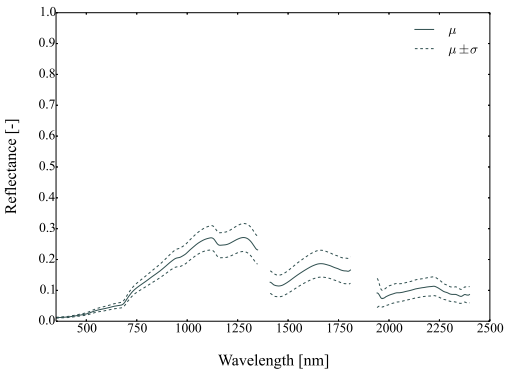


Figure B.22. Mean reflectance (μ) and standard deviation (σ) of Aquatic type c (AQ_C).

Picture	Plot	Longitude (DD)	Latitude (DD)	Altitude (m)	No. of spectra
	26	1.423972	42.802653	1343.362	12

566 *Carex* sp. homogeneous vegetation (CA_HV)

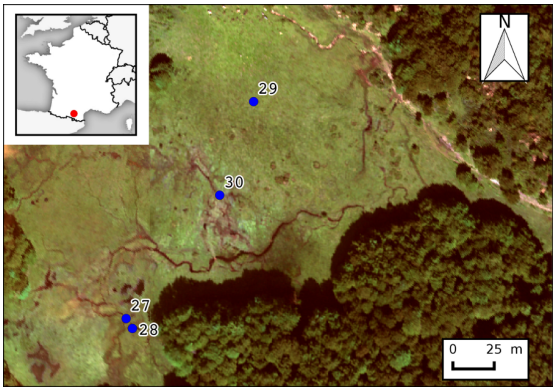


Figure B.23. Location of the *in situ* spectroradiometer measurements for the plots of *Carex* sp. homogeneous vegetation (CA_HV).

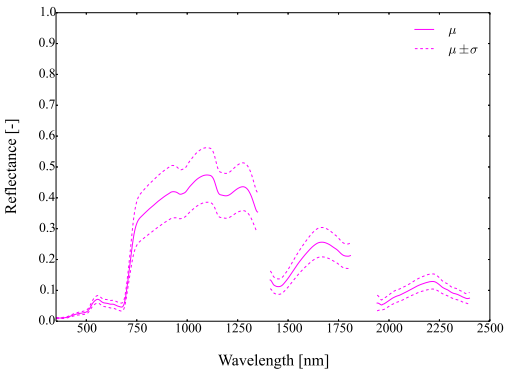
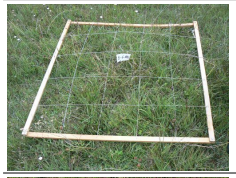


Figure B.24. Mean reflectance (μ) and standard deviation (σ) of *Carex* sp. homogeneous vegetation (CA_HV).

Picture	Plot	Longitude (DD)	Latitude (DD)	Altitude (m)	No. of spectra
	27	1.423499	42.802124	1343.533	4
	28	1.423547	42.802071	1344.568	4
	29	1.42441	42.803316	1351.678	9
	30	1.424173	42.802804	1344.481	10

567 *Pinguicula* sp. combined vegetation (PI_CV)

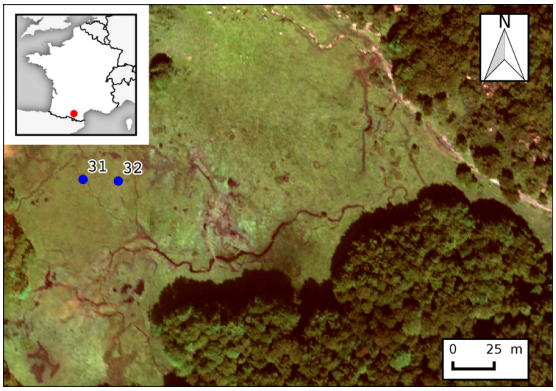


Figure B.25. Location of the *in situ* spectroradiometer measurements for the plots of *Pinguicula* sp. combined vegetation (PI_CV).

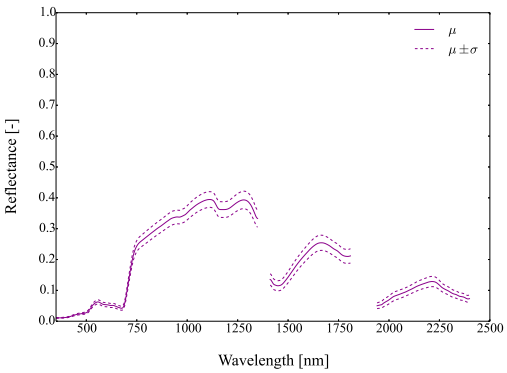
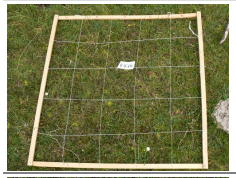


Figure B.26. Mean reflectance (μ) and standard deviation (σ) of *Pinguicula* sp. combined vegetation (PI_CV).

Picture	Plot	Longitude (DD)	Latitude (DD)	Altitude (m)	No. of spectra
	31	1.42316	42.802875	1344.344	12
	32	1.423421	42.80287	1344.247	3

568

References

1. Gorham, E. Northern peatlands: role in the carbon cycle and probable responses to climatic warming. *Ecological Applications* **1991**, *1*, 182–195.
2. Yu, Z.; Loisel, J.; Brosseau, D.P.; Beilman, D.W.; Hunt, S.J. Global peatland dynamics since the Last Glacial Maximum. *Geophysical Research Letters* **2010**, *37*.
3. Rydin, H.; Jeglum, J.K. *The Biology of Peatlands*, second ed.; Oxford University Press, 2013.
4. Kent, M.; Coker, P. *Vegetation Description and Analysis: A Practical Approach.*; Belhaven Press, 1992.
5. Schmidt, K.; Skidmore, A. Spectral discrimination of vegetation types in a coastal wetland. *Remote Sensing of Environment* **2003**, *85*, 92–108.
6. Adam, E.; Mutanga, O. Spectral discrimination of papyrus vegetation (*Cyperus papyrus* L.) in swamp wetlands using field spectrometry. *ISPRS Journal of Photogrammetry and Remote Sensing* **2009**, *64*, 612–620.
7. Seher, J.S.; Tueller, P.T. Color aerial photos for marshland. *Photogrammetric Engineering* **1973**, *9*, 489–499.
8. Adam, E.; Mutanga, O.; Rugege, D. Multispectral and hyperspectral remote sensing for identification and mapping of wetland vegetation: a review. *Wetlands Ecology and Management* **2010**, *18*, 281–296.
9. Guyot, G. Optical properties of vegetation canopies. *Applications of Remote Sensing in Agriculture* **1990**, pp. 19–43.
10. Yuan, L.; Zhang, L. Identification of the spectral characteristics of submerged plant *Vallisneria spiralis*. *Acta Ecologica Sinica* **2006**, *26*, 1005–1010.
11. Hestir, E.L.; Khanna, S.; Andrew, M.E.; Santos, M.J.; Viers, J.H.; Greenberg, J.A.; Rajapakse, S.S.; Ustin, S.L. Identification of invasive vegetation using hyperspectral remote sensing in the California Delta ecosystem. *Remote Sensing of Environment* **2008**, *112*, 4034–4047.
12. Torbick, N.; Becker, B.; Qi, J.; Lusch, D. Characterizing field-level hyperspectral measurements for identifying wetland invasive plant species. In *Invasive Species: Detection, Impact and Control*; Nova Science Publishers, 2009; pp. 97–115.
13. Hamada, Y.; Stow, D.A.; Coulter, L.L.; Jafolla, J.C.; Hendricks, L.W. Detecting Tamarisk species (*Tamarix* spp.) in riparian habitats of Southern California using high spatial resolution hyperspectral imagery. *Remote Sensing of Environment* **2007**, *109*, 237–248.
14. Vaiphasa, C.; Skidmore, A.K.; de Boer, W.F.; Vaiphasa, T. A hyperspectral band selector for plant species discrimination. *ISPRS Journal of Photogrammetry and Remote Sensing* **2007**, *62*, 225–235.
15. Jia, M.; Zhang, Y.; Wang, Z.; Song, K.; Ren, C. Mapping the distribution of mangrove species in the Core Zone of Mai Po Marshes Nature Reserve, Hong Kong, using hyperspectral data and high-resolution data. *International Journal of Applied Earth Observation and Geoinformation* **2014**, *33*, 226–231.
16. Prospere, K.; McLaren, K.; Wilson, B. Plant species discrimination in a tropical wetland using in situ hyperspectral data. *Remote Sensing* **2014**, *6*, 8494–8523.
17. Krankina, O.; Pflugmacher, D.; Friedl, M.; Cohen, W.; Nelson, P.; Baccini, A. Meeting the challenge of mapping peatlands with remotely sensed data. *Biogeosciences* **2008**, *5*, 1809–1820.
18. Hubert-Moy, L.; Clément, B.; Lennon, M.; Houet, T.; Lefeuvre, E. Etude de zones humides de fond de vallées à partir d'images hyperspectrales CASI: Application à un bassin versant de la région de Pleine-Fougères (Bretagne, France). *Photo-Interprétation* **2003**, *39*, 33–43.
19. Thomas, V.; Treitz, P.; Jelinski, D.; Miller, J.; Lafleur, P.; McCaughey, J.H. Image classification of a northern peatland complex using spectral and plant community data. *Remote Sensing of Environment* **2003**, *84*, 83–99.
20. Knoth, C.; Klein, B.; Prinz, T.; Kleinebecker, T. Unmanned aerial vehicles as innovative remote sensing platforms for high-resolution infrared imagery to support restoration monitoring in cut-over bogs. *Applied Vegetation Science* **2013**, *16*, 509–517.
21. Yagoub, H.; Belbachir, A.H.; Benabadji, N. Detection and mapping vegetation cover based on the Spectral Angle Mapper algorithm using NOAA AVHRR data. *Advances in Space Research* **2014**, *53*, 1686–1693.
22. Bahri, E.M.; Haboudane, D.; Bannari, A.; Bonn, F.; Chillasse, L. Essai de cartographie des espèces forestières dominantes dans le moyen atlas (Maroc) à l'aide des données Aster. *Revue Télédétection* **2007**, *7*, 283–301.
23. Sobhan, I. Species discrimination from a hyperspectral perspective. PhD thesis, International Institute for Geo-information Science & Earth Observation, University of Twente, the Netherlands (ITC), 2007.
24. Clark, M.L.; Roberts, D.A.; Clark, D.B. Hyperspectral discrimination of tropical rain forest tree species at leaf to crown scales. *Remote Sensing of Environment* **2005**, *96*, 375 – 398.

25. Lawrence, R.L.; Wood, S.D.; Sheley, R.L. Mapping invasive plants using hyperspectral imagery and Breiman Cutler classifications (RandomForest). *Remote Sensing of Environment* **2006**, *100*, 356–362.
26. Dalponte, M.; Ørka, H.O.; Gobakken, T.; Gianelle, D.; Næsset, E. Tree species classification in boreal forests with hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing* **2013**, *51*, 2632–2645.
27. Vyas, D.; Krishnayya, N.; Manjunath, K.; Ray, S.; Panigrahy, S. Evaluation of classifiers for processing Hyperion (EO-1) data of tropical vegetation. *International Journal of Applied Earth Observation and Geoinformation* **2011**, *13*, 228–235.
28. Pant, P.; Heikkinen, V.; Korpela, I.; Hauta-Kasari, M.; Tokola, T. Logistic regression-based spectral band selection for tree species classification: effects of spatial scale and balance in training samples. *IEEE Geoscience and Remote Sensing Letters* **2014**, *11*, 1604–1608.
29. Pal, M. Multinomial logistic regression-based feature selection for hyperspectral data. *International Journal of Applied Earth Observation and Geoinformation* **2012**, *14*, 214–220.
30. Merot, P.; Hubert-Moy, L.; Gascuel-Odoux, C.; Clement, B.; Durand, P.; Baudry, J.; Thenail, C. A Method for Improving the Management of Controversial Wetland. *Environmental Management* **2006**, *37*, 258–270.
31. Asner, G.P. Biophysical and biochemical sources of variability in canopy reflectance. *Remote Sensing of Environment* **1998**, *64*, 234–253.
32. Savitzky, A.; Golay, M.J. Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry* **1964**, *36*, 1627–1639.
33. Feilhauer, H.; Asner, G.P.; Martin, R.E.; Schmidtlein, S. Brightness-normalized partial least squares regression for hyperspectral data. *Journal of Quantitative Spectroscopy and Radiative Transfer* **2010**, *111*, 1947–1957.
34. Tsai, F.; Philpot, W. Derivative analysis of hyperspectral data. *Remote Sensing of Environment* **1998**, *66*, 41–51.
35. Serrano, L.; Peñuelas, J.; Ustin, S.L. Remote sensing of nitrogen and lignin in Mediterranean vegetation from AVIRIS data: Decomposing biochemical from structural signals. *Remote Sensing of Environment* **2002**, *81*, 355–364.
36. Clark, R.N.; Roush, T.L. Reflectance spectroscopy: Quantitative analysis techniques for remote sensing applications. *Journal of Geophysical Research: Solid Earth* **1984**, *89*, 6329–6340.
37. Kokaly, R.F.; Clark, R.N. Spectroscopic determination of leaf biochemistry using band-depth analysis of absorption features and stepwise multiple linear regression. *Remote Sensing of the Environment* **1999**, *67*, 267–287.
38. Mutanga, O.; Skidmore, A.K.; Prins, H. Predicting in situ pasture quality in the Kruger National Park, South Africa, using continuum-removed absorption features. *Remote Sensing of Environment* **2004**, *89*, 393–408.
39. Hu, B.; Lévesque, J.; Ardouin, J.P. Vegetation Species Identification Using Hyperspectral Imagery. IEEE International Geoscience and Remote Sensing Symposium, 2008 (IGARSS 2008); IEEE, , 2008; Vol. 2, pp. II299–II302.
40. Ghiyamat, A.; Shafri, H.Z.M.; Mahdiraji, G.A.; Shariff, A.R.M.; Mansor, S. Hyperspectral discrimination of tree species with different classifications using single-and multiple-endmember. *International Journal of Applied Earth Observation and Geoinformation* **2013**, *23*, 177–191.
41. Chang, C.I.; Ren, H. An experiment-based quantitative and comparative analysis of target detection and image classification algorithms for hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing* **2000**, *38*, 1044–1063.
42. Chauhan, H.; Mohan, B.K. Effectiveness of spectral similarity measures to develop precise crop spectra for hyperspectral data analysis. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* **2014**, *2*, 83–90.
43. Lance, G.N.; Williams, W.T. Computer programs for hierarchical polythetic classification ("similarity analyses"). *The Computer Journal* **1966**, *9*, 60–64.
44. Kruse, F.; Lefkoff, A.; Boardman, J.; Heidebrecht, K.; Shapiro, A.; Barloon, P.; Goetz, A. The spectral image processing system (SIPS)—interactive visualization and analysis of imaging spectrometer data. *Remote Sensing of Environment* **1993**, *44*, 145–163.
45. Chang, C.I. An information-theoretic approach to spectral variability, similarity, and discrimination for hyperspectral image analysis. *IEEE Transactions on Information Theory* **2000**, *46*, 1927–1932.
46. Du, Y.; Chang, C.I.; Ren, H.; Chang, C.C.; Jensen, J.O.; D'Amico, F.M. New hyperspectral discrimination measure for spectral characterization. *Optical Engineering* **2004**, *43*, 1777–1786.
47. van der Meer, F.; Bakker, W. Cross correlogram spectral matching: application to surface mineralogical mapping by using AVIRIS data from Cuprite, Nevada. *Remote Sensing of Environment* **1997**, *61*, 371–382.

48. Farifteh, J.; Van Der Meer, F.; Carranza, E. Similarity measures for spectral discrimination of salt-affected soils. *International Journal of Remote Sensing* **2007**, *28*, 5273–5293.
49. de Carvalho Jr, O.A.; Meneses, P.R. Spectral correlation mapper (SCM): an improvement on the spectral angle mapper (SAM). Summaries of the Ninth JPL Airborne Earth Science Workshop; Jet Propulsion Laboratory, National Aeronautics and Space Administration: Pasadena, CA, U.S.A, 2000; Vol. 9, *JPL Publication*.
50. Robila, S. An analysis of spectral metrics for hyperspectral image processing. IEEE International Geoscience and Remote Sensing Symposium, 2004 (IGARSS'04); IEEE: Anchorage, Alaska, U.S.A, 2004; Vol. 5, pp. 3233–3236.
51. Angelopoulou, E.; Lee, S.W.; Bajcsy, R. Spectral Gradient: A Material Descriptor Invariant to Geometry and Incident Illumination. Proceedings of the Seventh International Conference on Computer Vision (ICCV); IEEE Computer Society Press: Kerkyra, Greece, 1999; Vol. 2, pp. 861–867.
52. Boochs, F.; Kupfer, G.; Dockter, K.; Kühbauch, W. Shape of the red edge as vitality indicator for plants. *Remote Sensing* **1990**, *11*, 1741–1753.
53. Nagler, P.; Daughtry, C.; Goward, S. Plant litter and soil reflectance. *Remote Sensing of Environment* **2000**, *71*, 207–215.
54. Kim, M.S.; Daughtry, C.S.; Chappelle, E.; McMurtrey, J.; Walthall, C.L. The use of high spectral resolution bands for estimating absorbed photosynthetically active radiation (A par). Proceedings of 6th International Symposium on Physical Measurements and Signatures in Remote Sensing; CNES Editions: Toulouse, France, 1994; pp. 299–306.
55. Zarco-Tejada, P.J.; Pushnik, J.; Dobrowski, S.; Ustin, S. Steady-state chlorophyll a fluorescence detection from canopy derivative reflectance and double-peak red-edge effects. *Remote Sensing of Environment* **2003**, *84*, 283–294.
56. Sims, D.A.; Luo, H.; Hastings, S.; Oechel, W.C.; Rahman, A.F.; Gamon, J.A. Parallel adjustments in vegetation greenness and ecosystem CO_2 exchange in response to drought in a Southern California chaparral ecosystem. *Remote Sensing of Environment* **2006**, *103*, 289–303.
57. Barnes, E.; Clarke, T.; Richards, S.; Colaizzi, P.; Haberland, J.; Kostrzewski, M.; Waller, P.; Choi, C.; Riley, E.; Thompson, T.; others. Coincident detection of crop water stress, nitrogen status and canopy density using ground-based multispectral data. Proceedings of the 5th International Conference on Precision Agriculture; American Society of Agronomy : Crop Science Society of America : Soil Science Society of America: Bloomington, Minnesota, U.S.A, 2000; pp. 1–15.
58. Carter, G.A.; Miller, R.L. Early detection of plant stress by digital imaging within narrow stress-sensitive wavebands. *Remote Sensing of Environment* **1994**, *50*, 295–302.
59. Gitelson, A.A.; Gritz, Y.; Merzlyak, M.N. Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. *Journal of Plant Physiology* **2003**, *160*, 271–282.
60. Gitelson, A.A.; Keydan, G.P.; Merzlyak, M.N. Three-band model for noninvasive estimation of chlorophyll, carotenoids, and anthocyanin contents in higher plant leaves. *Geophysical Research Letters* **2006**, *33*, L11402.
61. Datt, B. A new reflectance index for remote sensing of chlorophyll content in higher plants: tests using Eucalyptus leaves. *Journal of Plant Physiology* **1999**, *154*, 30–36.
62. Datt, B. Remote sensing of chlorophyll a, chlorophyll b, chlorophyll a+b, and total carotenoid content in eucalyptus leaves. *Remote Sensing of Environment* **1998**, *66*, 111–121.
63. Zarco-Tejada, P.J.; Miller, J.R.; Mohammed, G.; Noland, T.; Sampson, P. Vegetation Stress Detection through Chlorophyll+ Estimation and Fluorescence Effects on Hyperspectral Imagery. *Journal of Environmental Quality* **2002**, *31*, 1433–1441.
64. Chen, P.; Haboudane, D.; Tremblay, N.; Wang, J.; Vigneault, P.; Li, B. New spectral indicator assessing the efficiency of crop nitrogen treatment in corn and wheat. *Remote Sensing of Environment* **2010**, *114*, 1987–1997.
65. le Maire, G.; François, C.; Dufrêne, E. Towards universal broad leaf chlorophyll indices using PROSPECT simulated database and hyperspectral reflectance measurements. *Remote Sensing of Environment* **2004**, *89*, 1–28.
66. le Maire, G.; François, C.; Soudani, K.; Berveiller, D.; Pontailler, J.Y.; Bréda, N.; Genet, H.; Davi, H.; Dufrêne, E. Calibration and validation of hyperspectral indices for the estimation of broadleaved forest leaf chlorophyll content, leaf mass per area, leaf area index and leaf canopy biomass. *Remote Sensing of Environment* **2008**, *112*, 3846–3864.
67. Filella, I.; Peñuelas, J. The red edge position and shape as indicators of plant chlorophyll content, biomass and hydric status. *International Journal of Remote Sensing* **1994**, *15*, 1459–1470.
68. Huete, A.; Liu, H.; Batchily, K.; Van Leeuwen, W. A comparison of vegetation indices over a global set of TM images for EOS-MODIS. *Remote Sensing of Environment* **1997**, *59*, 440–451.

69. Peñuelas, J.; Gamon, J.; Fredeen, A.; Merino, J.; Field, C. Reflectance indices associated with physiological changes in nitrogen-and water-limited sunflower leaves. *Remote Sensing of Environment* **1994**, *48*, 135–146.
70. Pinty, B.; Verstraete, M. GEMI: a non-linear index to monitor global vegetation from satellites. *Vegetatio* **1992**, *101*, 15–20.
71. Smith, R.; Adams, J.; Stephens, D.; Hick, P. Forecasting wheat yield in a Mediterranean-type environment from the NOAA satellite. *Crop and Pasture Science* **1995**, *46*, 113–125.
72. Gitelson, A.A.; Buschmann, C.; Lichtenthaler, H.K. The chlorophyll fluorescence ratio F 735/F 700 as an accurate measure of the chlorophyll content in plants. *Remote Sensing of Environment* **1999**, *69*, 296–302.
73. Gitelson, A.A.; Merzlyak, M.N. Remote estimation of chlorophyll content in higher plant leaves. *International Journal of Remote Sensing* **1997**, *18*, 2691–2697.
74. Gitelson, A.A.; Kaufman, Y.J.; Merzlyak, M.N. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sensing of Environment* **1996**, *58*, 289–298.
75. Maccioni, A.; Agati, G.; Mazzinghi, P. New vegetation indices for remote measurement of chlorophylls based on leaf directional reflectance spectra. *Journal of Photochemistry and Photobiology B: Biology* **2001**, *61*, 52–61.
76. Gitelson, A.A.; Merzlyak, M.N.; Chivkunova, O.B. Optical Properties and Nondestructive Estimation of Anthocyanin Content in Plant Leaves. *Photochemistry and Photobiology* **2001**, *74*, 38–45.
77. Gitelson, A.A.; Chivkunova, O.B.; Merzlyak, M.N. Nondestructive estimation of anthocyanins and chlorophylls in anthocyanic leaves. *American Journal of Botany* **2009**, *96*, 1861–1868.
78. Daughtry, C.; Walthall, C.; Kim, M.; De Colstoun, E.B.; McMurtrey, J. Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. *Remote Sensing of Environment* **2000**, *74*, 229–239.
79. Wu, C.; Niu, Z.; Tang, Q.; Huang, W. Estimating chlorophyll content from hyperspectral vegetation indices: Modeling and validation. *Agricultural and Forest Meteorology* **2008**, *148*, 1230–1241.
80. Haboudane, D.; Miller, J.R.; Tremblay, N.; Zarco-Tejada, P.J.; Dextraze, L. Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. *Remote Sensing of Environment* **2002**, *81*, 416–426.
81. Eitel, J.; Long, D.; Gessler, P.; Smith, A. Using in-situ measurements to evaluate the new RapidEye^{TM} satellite series for prediction of wheat nitrogen status. *International Journal of Remote Sensing* **2007**, *28*, 4183–4190.
82. Sims, D.A.; Gamon, J.A. Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. *Remote Sensing of Environment* **2002**, *81*, 337–354.
83. Qi, J.; Chehbouni, A.; Huete, A.; Kerr, Y.; Sorooshian, S. A modified soil adjusted vegetation index. *Remote Sensing of Environment* **1994**, *48*, 119–126.
84. Hunt, E.R.; Rock, B.N. Detection of changes in leaf water content using near-and middle-infrared reflectances. *Remote Sensing of Environment* **1989**, *30*, 43–54.
85. Chen, J.M. Evaluation of vegetation indices and a modified simple ratio for boreal applications. *Canadian Journal of Remote Sensing* **1996**, *22*, 229–242.
86. Dash, J.; Curran, P. The MERIS terrestrial chlorophyll index. *International Journal of Remote Sensing* **2004**, *25*, 5403–5413.
87. Haboudane, D.; Miller, J.R.; Pattey, E.; Zarco-Tejada, P.J.; Strachan, I.B. Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture. *Remote Sensing of Environment* **2004**, *90*, 337–352.
88. Hardisky, M.A.; Klemas, V.; Smart, R.M. The influence of soil salinity, growth form, and leaf moisture on the spectral radiance of *Spartina alterniflora* canopies. *Photogrammetric Engineering and Remote Sensing* **1983**, *49*, 77–83.
89. Tucker, C.J. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment* **1979**, *8*, 127–150.
90. Gandia, S.; Fernández, G.; García, J.; Moreno, J. Retrieval of vegetation biophysical variables from CHRIS/PROBA data in the SPARC campaign. Proceedings of the 2nd ESA CHRIS/Proba Workshop; ESA Publications Division: Frascati, Italy, 2004; Vol. 578, pp. 40–48.
91. Hansen, P.; Schjoerring, J. Reflectance measurement of canopy biomass and nitrogen status in wheat crops using normalized difference vegetation indices and partial least squares regression. *Remote Sensing of Environment* **2003**, *86*, 542–553.
92. Gao, B.C. NDWI — A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment* **1996**, *58*, 257–266.

93. Uto, K.; Kosugi, Y. Hyperspectral manipulation for the water stress evaluation of plants. *Contemporary Materials* **2012**, *1*, 18–25.
94. Peñuelas, J.; Gamon, J.A.; Griffin, K.L.; Field, C.B. Assessing community type, plant biomass, pigment composition, and photosynthetic efficiency of aquatic vegetation from spectral reflectance. *Remote Sensing of Environment* **1993**, *46*, 110–118.
95. Reyniers, M.; Walvoort, D.J.; De Baardemaaker, J. A linear model to predict with a multi-spectral radiometer the amount of nitrogen in winter wheat. *International Journal of Remote Sensing* **2006**, *27*, 4159–4179.
96. Rondeaux, G.; Steven, M.; Baret, F. Optimization of soil-adjusted vegetation indices. *Remote Sensing of Environment* **1996**, *55*, 95–107.
97. Gamon, J.; Peñuelas, J.; Field, C.B. A narrow-waveband spectral index that tracks diurnal changes in photosynthetic efficiency. *Remote Sensing of environment* **1992**, *41*, 35–44.
98. Roujean, J.L.; Breon, F.M. Estimating PAR absorbed by vegetation from bidirectional reflectance measurements. *Remote Sensing of Environment* **1995**, *51*, 375–384.
99. Horler, D.; Dockray, M.; Barber, J. The red edge of plant leaf reflectance. *International Journal of Remote Sensing* **1983**, *4*, 273–288.
100. Horler, D.; Dockray, M.; Barber, J.; Barringer, A. Red edge measurements for remotely sensing plant chlorophyll content. *Advances in Space Research* **1983**, *3*, 273–277.
101. Gitelson, A.A.; Vina, A.; Ciganda, V.; Rundquist, D.C.; Arkebauer, T.J. Remote estimation of canopy chlorophyll content in crops. *Geophysical Research Letters* **2005**, *32*, L08403.
102. Cho, M.A.; Skidmore, A.K. A new technique for extracting the red edge position from hyperspectral data: The linear extrapolation method. *Remote Sensing of Environment* **2006**, *101*, 181–193.
103. Guyot, G.; Baret, F. Utilisation de la haute résolution spectrale pour suivre l'état des couverts végétaux. Signatures spectrales d'objets en télédétection. 4 ème Colloque international.; Agence Spatiale Européenne: Aussois, France, 1988; Vol. 287, pp. 279–286.
104. Zhu, Y.; Yao, X.; Tian, Y.; Liu, X.; Cao, W. Analysis of common canopy vegetation indices for indicating leaf nitrogen accumulations in wheat and rice. *International Journal of Applied Earth Observation and Geoinformation* **2008**, *10*, 1–10.
105. Xue, L.; Cao, W.; Luo, W.; Dai, T.; Zhu, Y. Monitoring leaf nitrogen status in rice with canopy spectral reflectance. *Agronomy Journal* **2004**, *96*, 135–142.
106. Peñuelas, J.; Filella, I.; Lloret, P.; Muñoz, F.; Vilajeliu, M. Reflectance assessment of mite effects on apple trees. *International Journal of Remote Sensing* **1995**, *16*, 2727–2733.
107. Vincini, M.; Frazzi, E.; D'Alessio, P. Angular dependence of maize and sugar beet VIs from directional CHRIS/Proba data. Proceedings of the 4th ESA CHRIS/Proba Workshop; , 2006.
108. Jordan, C.F. Derivation of leaf-area index from quality of light on the forest floor. *Ecology* **1969**, pp. 663–666.
109. McMurtrey, J.; Chappelle, E.; Kim, M.; Meisinger, J.; Corp, L. Distinguishing nitrogen fertilization levels in field corn (*Zea mays* L.) with actively induced fluorescence and passive reflectance measurements. *Remote Sensing of Environment* **1994**, *47*, 36–44.
110. Chappelle, E.W.; Kim, M.S.; McMurtrey, J.E. Ratio analysis of reflectance spectra (RARS): an algorithm for the remote estimation of the concentrations of chlorophyll a, chlorophyll b, and carotenoids in soybean leaves. *Remote Sensing of Environment* **1992**, *39*, 239–247.
111. Zarco-Tejada, P.J.; Miller, J.R. Land cover mapping at BOREAS using red edge spectral parameters from CASI imagery. *Journal of Geophysical Research: Atmospheres (1984–2012)* **1999**, *104*, 27921–27933.
112. Lichtenthaler, H.; Lang, M.; Sowinska, M.; Heisel, F.; Miehe, J. Detection of vegetation stress via a new high resolution fluorescence imaging system. *Journal of Plant Physiology* **1996**, *148*, 599–612.
113. Elvidge, C.D.; Chen, Z. Comparison of broad-band and narrow-band red and near-infrared vegetation indices. *Remote Sensing of Environment* **1995**, *54*, 38–48.
114. Broge, N.H.; Leblanc, E. Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density. *Remote Sensing of Environment* **2001**, *76*, 156–172.
115. Vogelmann, J.; Rock, B.; Moss, D. Red edge spectral measurements from sugar maple leaves. *International Journal of Remote Sensing* **1993**, *14*, 1563–1575.

116. Pu, R.; Foschi, L.; Gong, P. Spectral feature analysis for assessment of water status and health level in coast live oak (*Quercus agrifolia*) leaves. *International Journal of Remote Sensing* **2004**, *25*, 4267–4286.
117. Peñuelas, J.; Pinol, J.; Ogaya, R.; Filella, I. Estimation of plant water concentration by the reflectance water index WI (R900/R970). *International Journal of Remote Sensing* **1997**, *18*, 2869–2875.
118. Fisher, R.A. *Statistical methods for research workers*; Genesis Publishing Pvt Ltd, 1925.
119. Mann, H.B.; Whitney, D.R. On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics* **1947**, pp. 50–60.
120. Kruskal, W.H.; Wallis, W.A. Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association* **1952**, *47*, 583–621.
121. Jensen, J.R. *Introductory digital image processing: a remote sensing perspective.*, second ed.; Prentice-Hall, 1996.
122. Salisbury, F.B.; Ross, C.W. *Plant Physiology*; Wadworth, Belmont, CA, 1992.
123. Gausman, H.W. Visible light reflectance, transmittance, and absorptance of differently pigmented cotton leaves. *Remote Sensing of Environment* **1983**, *13*, 233–238.
124. Clevers, J. The use of imaging spectrometry for agricultural applications. *ISPRS Journal of Photogrammetry and Remote Sensing* **1999**, *54*, 299–304.
125. Mutanga, O.; Skidmore, A.K. Red edge shift and biochemical content in grass canopies. *ISPRS Journal of Photogrammetry and Remote Sensing* **2007**, *62*, 34–42.
126. Woolley, J.T. Reflectance and transmittance of light by leaves. *Plant Physiology* **1971**, *47*, 656–662.
127. Boyer, M.; Miller, J.; Belanger, M.; Hare, E.; Wu, J. Senescence and spectral reflectance in leaves of northern pin oak (*Quercus palustris* Muenchh.). *Remote Sensing of Environment* **1988**, *25*, 71–87.
128. Fourty, T.; Baret, F.; Jacquemoud, S.; Schmuck, G.; Verdebout, J. Leaf optical properties with explicit description of its biochemical composition: direct and inverse problems. *Remote sensing of Environment* **1996**, *56*, 104–117.
129. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **2011**, *12*, 2825–2830.
130. Breiman, L. Random forests. *Machine learning* **2001**, *45*, 5–32.
131. Belgiu, M.; Drăguţ, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing* **2016**, *114*, 24–31.
132. Vapnik, V.N. *Statistical learning theory*; Vol. 1, Wiley New York, 1998.
133. Cawley, G.C.; Talbot, N.L. Gene selection in cancer classification using sparse logistic regression with Bayesian regularization. *Bioinformatics* **2006**, *22*, 2348–2355.
134. Dumont, J.; Hirvonen, T.; Heikkinen, V.; Mistretta, M.; Granlund, L.; Himanen, K.; Fauch, L.; Porali, I.; Hiltunen, J.; Keski-Saari, S.; others. Thermal and hyperspectral imaging for Norway spruce (*Picea abies*) seeds screening. *Computers and Electronics in Agriculture* **2015**, *116*, 118–124.
135. Wold, S.; Sjöström, M.; Eriksson, L. PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systemsn* **2001**, *58*, 109–130.
136. Barker, M.; Rayens, W. Partial least squares for discrimination. *Journal of Chemometrics* **2003**, *17*, 166–173.
137. Castillo, R.; Otto, M.; Freer, J.; Valenzuela, S. Multivariate strategies for classification of Eucalyptus globulus genotypes using carbohydrates content and NIR spectra for evaluation of their cold resistance. *Journal of Chemometrics* **2008**, *22*, 268–280.
138. Lê Cao, K.A.; Boitard, S.; Besse, P. Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems. *BMC bioinformatics* **2011**, *12*, 253.
139. Peerbhay, K.Y.; Mutanga, O.; Ismail, R. Commercial tree species discrimination using airborne AISA Eagle hyperspectral imagery and partial least squares discriminant analysis (PLS-DA) in KwaZulu–Natal, South Africa. *ISPRS Journal of Photogrammetry and Remote Sensing* **2013**, *79*, 19–28.
140. Apolloni, B.; Pedrycz, W.; Bassis, S.; Malchiodi, D. *The Puzzle of Granular Computing*; Studies in Computational Intelligence, Springer Berlin Heidelberg, 2008.
141. Szmidt, E. *Distances and Similarities in Intuitionistic Fuzzy Sets*; Studies in Fuzziness and Soft Computing, Springer International Publishing, 2013.
142. Aggarwal, C.C.; Hinneburg, A.; Keim, D.A. On the surprising behavior of distance metrics in high dimensional space. International Conference on Database Theory. Springer, 2001, pp. 420–434.