



**HAL**  
open science

## Exploring the link between spectral variance and upper canopy taxonomic diversity in a tropical forest: influence of spectral processing and feature selection

Colette Badourdine, Jean-baptiste Féret, Raphaël Pélissier, Grégoire Vincent

### ► To cite this version:

Colette Badourdine, Jean-baptiste Féret, Raphaël Pélissier, Grégoire Vincent. Exploring the link between spectral variance and upper canopy taxonomic diversity in a tropical forest: influence of spectral processing and feature selection. *Remote Sensing in Ecology and Conservation*, 2023, 9 (2), pp.235-250. 10.1002/rse2.306 . hal-03800415

**HAL Id: hal-03800415**

**<https://hal.science/hal-03800415>**

Submitted on 6 Oct 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

## RESEARCH ARTICLE

# Exploring the link between spectral variance and upper canopy taxonomic diversity in a tropical forest: influence of spectral processing and feature selection

Colette Badourdine<sup>1</sup> , Jean-Baptiste Féret<sup>2</sup>, Raphaël Pélissier<sup>1</sup> & Grégoire Vincent<sup>1</sup> <sup>1</sup>AMAP, Univ. Montpellier, IRD, CNRS, CIRAD, INRAE, Montpellier, France<sup>2</sup>TETIS, INRAE, AgroParisTech, Université Montpellier, Montpellier, France

## Keywords

Biodiversity mapping, features selection, forest canopy, imaging spectroscopy, spectral diversity, tropical forest biodiversity

## Correspondence

Colette Badourdine, AMAP, Univ. Montpellier, IRD, CNRS, CIRAD, INRAE, Montpellier, France.  
E-mail: colette.badourdine@ird.fr

**Funding Information.** J.-B. Féret acknowledges financial support from Agence Nationale de la Recherche (France) (BioCop project—ANR-17-CE32-0001) and TOSCA program grant from the French Space Agency (CNES) (HyperTropik project). Colette Badourdine benefited from a PhD grant co-funded by CNES and IRD, France.

Editor: Temuulen Sankey  
Associate Editor: Feng Ling

Received: 5 April 2022; Revised: 22 August 2022; Accepted: 5 September 2022

doi: 10.1002/rse2.306

## Introduction

Tropical forests are the largest terrestrial reservoir of biodiversity (ter Steege et al., 2013). They are subjected to a rapid loss of biodiversity due to climate change, land-use change (O'Neill et al., 2018), and anthropogenic pressure (Alamgir et al., 2017). Developing monitoring systems able to assess tropical forest degradation in terms of carbon stocks and biodiversity is a pressing challenge. The scale and speed of change in tropical forests far outpace our ability to monitor them through ground inventories alone. Developing new remote sensing approaches capable of providing rapid estimates of biodiversity over large

## Abstract

The rapid loss of biodiversity in tropical rainforests calls for new remote sensing approaches capable of providing rapid estimates of biodiversity over large areas. Imaging spectroscopy has shown potential for the estimation of taxonomic diversity, but the link with spectral diversity has not been investigated extensively with experimental data so far. We explored the relationship between taxonomic diversity and visible to near infrared spectral variance derived from various spectral processing techniques by means of a labeled dataset comprising 2000 individual tree crowns from 200 species from an experimental tropical forest station in French Guiana. We generated a set of artificially assembled communities covering a broad range of taxonomic diversity from this experimental dataset. We analyzed the impact of various processing steps: spectral normalization, spectral transformation through principal component analysis, and feature selection. Correlation between taxonomic diversity and inter-specific spectral variance was strong. Correlation was lower with total spectral variance, with or without normalization and transformation. Dimensionality reduction through feature selection resulted in dramatic improvement of the correlation between Shannon index and spectral variance. While airborne diversity mapping of tropical forest may not be at hand yet, our results confirm that spectral diversity metrics, when computed on properly preprocessed and selected spectral information can predict taxonomic diversity in tropical ecosystems.

areas is necessary to better understand the functioning of tropical rainforests.

Spectral diversity is deemed to integrate information on taxonomic, functional, and phylogenetic diversity and similarly to predict ecosystem function (Cavender-Bares et al., 2020). The variation of spectral patterns can thus be related to functional and structural properties that vary between species or functional groups (Gamon et al., 1997; Ustin & Gamon, 2010; Wang & Gamon, 2019). Operational multispectral satellites have been able to predict plant functional traits when combined with climatic and soil information (Aguirre-Gutiérrez et al., 2021; Ma et al., 2019; Rapinel et al., 2019). Airborne and satellite

imaging spectroscopy thus hold promise for detecting changes in the floristic and functional composition of temperate and tropical forests and estimating their biodiversity (Clark & Roberts, 2012; Féret & Asner, 2014; Somers & Asner, 2013).

Most studies that use imaging spectroscopy to estimate biodiversity rely on the Spectral Variation Hypothesis (SVH; Palmer et al., 2000), which posits that the variability in reflectance or “spectral variability” in space is an expression of spatial heterogeneity of ecosystems composition and is therefore related to plant diversity. However, the simple expectation that a higher spectral variation will indicate a higher plant diversity is not always corroborated by empirical studies. Schmidlein and Fassnacht (2017) showed, for instance, that the SVH does not hold over broad regions or across time using MODIS 500 × 500 m pixels. The coarser the spatial resolution of remote sensing data is, the more likely a pixel is to contain several species and the smaller the overall spectral variation across pixels of a given area (Fassnacht et al., 2022). Therefore, one may expect that, for a given region, the relationship between spectral variation and diversity will not be stable across spatial resolution scales. But, higher spatial resolution is not always better. For instance, if the pixel size is below the individual tree crown size, the within-individual variability increases, confounding the relationship between spectral variability and taxonomic diversity (Nagendra & Rocchini, 2008). The spatial resolution of remote sensing data should therefore be adapted to the size of individual organisms of a given ecosystem. Whether a higher spatial resolution allows a better assessment of tropical forest diversity thus remains an open question.

Spectral diversity of plant canopies can be measured in different ways. It can be estimated using the coefficient of variation of spectral indices (Oindo & Skidmore, 2002) or spectral bands among pixels (Gholizadeh et al., 2018, 2019; Hall et al., 2010; Wang et al., 2018), the mean distance of a group of pixels from their spectral centroid (Rocchini, 2007; Rocchini et al., 2010), the number of spectrally distinct clusters obtained by unsupervised classification or “spectral species” (Féret & Asner, 2014), and spectral variance (Laliberté et al., 2019).

One strong limitation to the design of a methodology adapted to tropical forests canopies so far is the relative scarcity of adequate working datasets, i.e., sky-view spectral images over a study site where individual tree crowns have been delineated and identified at the species level to serve as a control for the assessment of species diversity (but see Baldeck et al., 2015). In the present paper, we take advantage of a dataset including very-high spatial resolution spectral information extracted from visible to near infrared imaging spectroscopy acquired over an

experimental tropical forest station in French Guiana, encompassing about 2000 individual tree crowns (ITCs) from 200 species. Albeit important in size, the data at hand are incomplete because they contain only the most easily discernible crowns whose segmentation and label could be confirmed beyond reasonable doubt during field checking. This rich dataset however offers the possibility to explore the link between spectral diversity and taxonomic diversity by allowing the reconstruction of pseudo images through re-assembling labeled tree crowns drawn from the database. We generated a set of artificially assembled communities covering a broad range of taxonomic diversity from our global tree crown dataset to explore the relationship between spectral variance and taxonomic diversity. We hypothesised, following the SVH, that the spectral variance measured on these communities would be correlated to their taxonomic diversity. We applied a hierarchical variance partitioning to the artificial tree communities to evaluate the contribution of species (inter-species spectral variance) and individuals (inter-crown spectral variance within species) to the total spectral variance. We evaluated the influence of spectral processing and feature selection on the correlation between spectral variance and taxonomic diversity, and on the hierarchical structure of spectral variance. These analyses allowed us to highlight the importance of spectral processing and feature selection for the estimation of taxonomic diversity using spectral variance. Moreover, we were able to test if the correlation between spectral variance and taxonomic diversity was related to the proportion of interspecies spectral variance. Practical implications for biodiversity monitoring from airborne imaging spectroscopy are finally discussed.

## Material and Methods

### Study area

The study site is located at Paracou, on the coastal part of French Guiana (51°8' N, 52°53' W), where a set of 16 permanent forest plots (118.75 ha in total) are regularly monitored. All stems above 10 cm diameter at breast height (DBH) have been inventoried regularly for more than 25 years. More than 750 tree species have been listed on the site. A detailed description of the site and experimental design can be found in Gourlet-Fleury et al. (2004). In this study, we used data from the forest inventory conducted in 2015.

### Imaging spectroscopy

Imaging spectroscopy was acquired with a Hypspec VNIR-1600 (Hypspec NEO, Skedsmokorset, Norway) sensor

coupled with a Riegl LMSQ780 laser scanner. The 160 spectral bands span the visible to near infrared (VNIR) range, from 414 to 994 nm, with a spectral sampling distance of 3.64 nm. The flight of the King Air B200 airplane took place on cloudless conditions on September 19th, 2016, from 15:00 to 17:00, solar time, at an average altitude of 920 m and covered all the plots of the experimental site of Paracou. Images were orthorectified, georeferenced, and atmospheric corrections were applied using ATCOR-4. The final product obtained had a 1-m spatial resolution. The full spectral range of the VNIR sensor was considered in this study.

We extracted spectral information corresponding to the bottom of atmosphere VNIR reflectance from pixels corresponding to the ITCs from the image mosaic. Several pre-processing steps were performed, including filtering of irrelevant pixels and spectral domains, spectral normalization and transformation. This included: (1) spatial masking of pixels corresponding to defoliated tree crowns and shaded pixels, (2) spectral masking to remove atmospheric water absorption bands, (3) reflectance normalization based on continuum removal (CR), and (4) spectral transformation of normalized reflectance with principal component analysis (PCA).

### Removal of irrelevant pixels and noisy spectral domains

We masked pixels corresponding to lower values of Normalized Difference Vegetation Index (NDVI) and fixed a threshold for minimum NDVI of 0.5 to remove non-vegetated or non-foliated pixels (Asner & Martin, 2009; Féret & Asner, 2014). Shaded parts of the ITCs are characterized by low overall reflectance compared with sunlit pixels, particularly in the near infrared (NIR) domain; therefore, pixels with NIR reflectance inferior to a threshold of 20% reflectance were also masked (Féret & de Boissieu, 2020). We discarded all spectral bands between 887 nm and 994 nm, due to low signal to noise ratio. This reflectance dataset is hereafter referred to as Level-1 spectral processing (Table 1).

### Normalization of reflectance data using continuum removal

Continuum removal (CR) or convex-hull transform is a normalization procedure allowing comparison of individual absorption features from a common baseline (Clark & Roush, 1984). The CR aims at performing albedo normalization, allowing comparison of individual absorption features among reflectance spectra (Cavender-Bares et al., 2020). It also reduces structure-induced reflectance effects of canopies such as within-canopy scattering of

**Table 1.** Number of bands, pixels, ITC, and species at each step of reflectance data processing.

Filtering/processing	Number of bands	Number of pixels	Number of species	Number of ITC
Initial dataset	160	46 571	246	2246
Discard shaded and defoliated pixels	160	41 564	246	2239
Discard noisy spectral domains	124	41 564	246	2239
Removal of ITCS with less than 10 pixels (preprocessing level-1)	124	37 533	199	1595
Continuum removal (preprocessing level-2)	122	37 533	199	1595

diffuse radiation (Serbin & Townsend, 2020), and the effect of changing illumination from sunlit to shaded parts of the tree crown.

The continuum is a convex hull fit over the top of a reflectance spectrum using straight-line segments that connect local spectra maxima. This continuum is then used as a reference, and the continuum-removed spectrum is computed as the ratio between the original spectrum and the continuum, as follows (Eq. 1):

$$R_{CR} = (R/C) \quad (1)$$

where  $R_{CR}$  corresponds to the continuum-removed reflectance,  $R$  is the original reflectance, and  $C$  is the continuum.  $R_{CR}$  ranges between 0 and 1 if  $R$  is positive, which, unless improper preprocessing, is the case. This continuum-removed reflectance dataset is hereafter referred to as Level-2 spectral processing (see Table 1).

### Principal component analysis

We applied a standardized PCA (SPCA) to the Level-2 data. PCA is a commonly used method in imaging spectroscopy analysis for spectral transformation and dimensionality reduction (Ruiz Hidalgo et al., 2021; Theodoridis et al., 2014). PCA is particularly relevant for dimensionality reduction of hyperspectral information, as it produces uncorrelated (orthogonal) latent variables corresponding to the principal components (PCs) from original spectral information characterized by strong correlation and redundancy between neighboring spectral bands. PCs are ranked by decreasing contribution to total explained variance. Components with low eigenvalues explaining a marginal proportion of the total variance in an image may still contain useful information for specific tasks such as species discrimination and biodiversity analysis: Mather (1999) suggested that the choice of PCA components should not be based on eigenvalues alone

and Rodarmel and Shan (2002) recommended performing visual selection of the components in order to select those including relevant biological spatial patterns. In the case of imaging spectroscopy, PCs explaining an important part of the spectral variance may also include patterns related to sensor artefacts or conditions of acquisition. PC selection based on visual analysis has been suggested by multiple authors using PCA for dimensionality reduction of imaging spectroscopy in the context of biodiversity analysis (Asner et al., 2012; Féret & Asner, 2014; Laliberté et al., 2019). Here, we performed PC selection based on a statistical criterion corresponding to the correlation between taxonomic diversity and spectral variance, and we used Sequential Forward Selection (SFS) to identify feature relevance (see 3.5).

This SPCA-transformed continuum-removed reflectance dataset is hereafter referred to as Level-3 spectral processing.

### Standardization of spectral information

For each of the three spectral datasets corresponding to the different levels of spectral processing, a standardization across features was performed in order to give the same importance to each spectral feature, and the same contribution to spectral variance.

### Delineated tree crowns

Field survey was conducted to build a large ground truth dataset. Easily discernible crowns were delineated manually using the canopy height model derived from LiDAR (Light Detection And Ranging) acquisitions and the high spatial resolution RGB (Red, Green, Blue) images acquired simultaneously with imaging spectroscopy. The correct delineation of these ITCs was then validated in the field and the corresponding species ascertained from the ground inventory data (Aubry-Kientz et al., 2019). A total of 2246 ITCs from 246 species were delineated over the plot network, and the corresponding pixels were extracted from the imaging spectroscopy based on the crown outlines. This dataset was subsequently reduced when discarding shaded pixels and by eliminating ITCs with less than 10 pixels were removed from the dataset (Table 1).

### Generation of artificial populations

We aimed at quantifying the link between spectral variance and taxonomic diversity for different communities of various richness.

Species richness ( $S$ ) is the number of species present in the considered community. Shannon diversity index ( $H'$ )

is a measure of entropy. Simpson diversity index ( $D$ ) is the probability that two individuals randomly selected belong to different species. Both Shannon and Simpson indices combine species richness and evenness, but Simpson index gives more weight to abundant species (Marcon, 2015).

We took advantage of the experimental dataset combining georeferenced delineated tree crowns and corresponding reflectance acquired from the imaging spectroscopy in order to explore a broad range of taxonomic diversity and composition. As this initial dataset was strongly imbalanced in terms of number of pixels and crowns per species (pixels per crowns range = 138 and interquartile range = 15), we developed a procedure for the generation of populations with controlled species richness.

Each artificial population comprised 100 ITCs selected from the crown database, and each ITC included 10 pixels, resulting in individual populations of 1000 pixels. For ITCs of more than 10 pixels, we performed a random selection of 10 pixels among all pixels for each ITC draw. To generate population samplings with a large range of species diversity, our simulation strategy was based on two successive steps. First, we produced 20 artificial populations corresponding to extreme levels of species richness ( $S$ ): 10 populations of maximum species richness ( $S = 100$ ), and 10 populations of low species richness ( $S = 2$ ). Then, we produced intermediate levels of diversity, based on successive resampling of the populations corresponding to extreme richness. We gradually decreased or increased the diversity of populations with high and low richness respectively, by sequentially deleting ITCs and drawing others from the whole ITCs dataset randomly and without replacement to ensure that no population will have pixels from the same ITC sampled twice. The procedure resulted in 220 populations covering the full gradient of taxonomic diversity with different equitabilities (Table 2).

We repeated the simulation procedure in order to generate 501 sets of 220 populations. For each population (number of observations = 100), we computed three taxonomic diversity metrics to characterize tree species assemblages (Table 3). Figure 1 shows the relationship between  $S$ ,  $H'$  and  $D$  for the set of artificial populations

**Table 2.** Statistical summary of diversity metrics for the generated artificial populations.

Indice	Median	Max	Min
Richness ( $S$ )	50.00	100	2
Shannon index ( $H'$ )	3.34	4.50	0.33
Simpson index ( $D$ )	0.94	0.99	0.18

with median correlation for Level-2 data. Both Shannon and Simpson diversity indices showed negative skew, but this skewness was particularly strong for the Simpson index. Our protocol for generating artificial populations allowed us to explore a wide gradient of taxonomic diversity, even if extreme cases may appear unrealistic. To estimate the actual values of  $S$ ,  $H'$  and  $D$  that can be found in the study area, we selected from the forest inventory database all the ITCs with a DBH  $\geq 20$  cm, which are more likely to be seen from above. We drew 100 samples of 100 neighboring ITCs and compared their diversity metrics with those of the artificial populations. The median Shannon index was 3.7, the median Simpson index was 0.97, and the median species richness was 56 species, quite close to the medians obtained for the artificial populations (see Table 2).

### Spectral variance partitioning

We performed spectral variance analysis to study the influence of spectral processing on the relationship between taxonomic diversity and different components of spectral variance. Total spectral variance (or spectral inertia) is the sum over all variables (i.e. individual spectral or principal components) of the unidimensional variances, i.e.:

$$\text{Var}_{\text{TOT}} = \sum_{i=1}^n \frac{\sum_{j=1}^p (y_{ij} - \bar{y}_i)^2}{p} \quad (2)$$

where  $n$  is the number of spectral bands and  $p$  is the number of pixels,  $y_{ij}$  corresponds to the reflectance of spectral band  $i$  and  $j$ ,  $\bar{y}_i$  corresponds to the reflectance averaged over all pixels for spectral band  $i$ . We hierarchically partitioned the total spectral variance of each artificial population as:

$$\text{Var}_{\text{TOT}} = \text{Var}_{\text{SP}} + \text{Var}_{\text{ITC|SP}} + \text{Var}_{\text{RES}} \quad (3)$$

where  $\text{Var}_{\text{TOT}}$  is the total variance of the spectral data (corresponding to Level-1, Level-2, or Level-3 reflectance data).  $\text{Var}_{\text{SP}}$  is the part of this variance explained by species (inter-species variance).  $\text{Var}_{\text{ITC|SP}}$  is the part of the intra-species variance explained by ITCs (inter-crown

within species variance), and  $\text{Var}_{\text{RES}}$  is the residual (intra-crown) variance (Legendre & Legendre, 1998).

In order to highlight the influence of spectral processing steps, we compared the correlation between  $\text{Var}_{\text{TOT}}$  and taxonomic diversity for each level of processing of spectral information, and we analyzed the distribution of  $\text{Var}_{\text{SP}}$ ,  $\text{Var}_{\text{ITC|SP}}$  and  $\text{Var}_{\text{RES}}$ , for each of the 501 sets of populations.

### Feature selection

We used Sequential Feature Selection (SFS) to identify the combination of spectral features (spectral bands or components) maximizing the correlation between  $\text{Var}_{\text{TOT}}$  and taxonomic diversity. SFS is a greedy search algorithm seeking to optimize a criterion (corresponding here to the Pearson correlation coefficient ( $r$ ) between  $\text{Var}_{\text{TOT}}$  and a taxonomic diversity index) by adding features sequentially in forward mode and systematically evaluating the model for all available features to be added to an initial subset (Theodoridis et al., 2014). Here, SFS was performed on one set of 220 populations, hereafter named calibration set, and the generalization ability of the selected features was tested on the remaining 500 sets of 220 populations. The calibration set was identified based on the correlation between  $\text{Var}_{\text{TOT}}$  and the Shannon index when using all features: for each level of processing, we identified the set of populations of median correlation. We made the assumption that mixing of crowns and pixels when generating artificial populations was sufficient for feature selection to be robust enough from one population to another. We also assumed that the selection of features obtained using this set characterized by the median value of the correlation between  $\text{Var}_{\text{TOT}}$  and the Shannon index would be an acceptable trade-off for the applicability of selected features to the largest number of populations. We selected the Shannon index as it was better correlated to  $\text{Var}_{\text{TOT}}$  ( $r = 0.28$ ) than species richness ( $r = 0.24$ ) and better correlated to  $\text{Var}_{\text{SP}}$  than Simpson index ( $r = 0.88$  for Shannon index,  $r = 0.80$  for Simpson index).

Figure 2 summarizes the full analysis procedure.

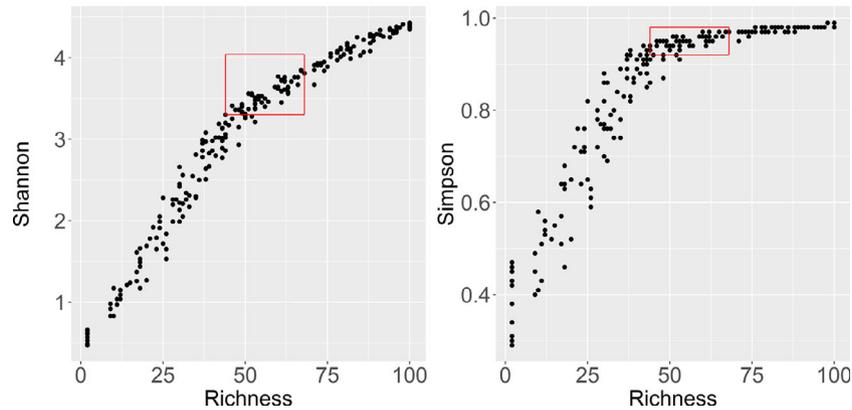
## Results

### Impacts of reflectance preprocessing on the relationship between spectral variance and taxonomic diversity

We computed the correlation between  $\text{Var}_{\text{TOT}}$  and the Shannon index for the 501 sets of 220 populations for each of the three levels of spectral processing. The variability of this correlation was large for each processing level (Fig. 3) and were approximately normally

**Table 3.** Summary of taxonomic diversity metrics used in this study ( $p_i$  corresponds to the relative abundance of species  $i$ ).

Diversity metrics	Description/equation
Species richness ( $S$ )	Number of species
Shannon index ( $H'$ ) (Shannon, 1948)	$H' = -\sum_i^S p_i \times \ln(p_i)$
Simpson index ( $D$ ) (Simpson, 1949)	$D = 1 - \sum_i^S (p_i)^2$



**Figure 1.** Distribution of calibration populations for Level-2 dataset along the Richness/Shannon or Richness/Simpson Index gradient. The range of diversity measured over Paracou is represented by the red box.

distributed. Our results highlight the strong influence of spectral normalization, as the median correlation increased from 0.15 for Level-1 (prior to normalization) to 0.30 and 0.35 for Level-2 and Level-3, respectively. The influence of PCA in addition to spectral normalization resulted in marginal overall increase of the correlation. However, this correlation remained low overall, with 15% of the populations showing correlation higher than 0.50.

### Effects of feature selection on the relationship between spectral variance and taxonomic diversity

We computed the evolution of the correlation between  $\text{Var}_{\text{TOT}}$  and diversity indices with an increasing number of features for each set to test the robustness of feature ranking obtained from the calibration set. The evolution of the correlation between  $\text{Var}_{\text{TOT}}$  and taxonomic diversity when increasing the number of features was remarkably similar across taxonomic diversity indices (Fig. 4).

Feature selection strongly improved the correlation overall for the three levels of processing, and the improvement was stronger on the calibration set. The maximum correlation was obtained for one spectral band when using Level-1 data, 12 bands for Level-2 data and 12 PCs for Level-3 data. The evolution of the correlation with dimensionality reduction gave very different results between the calibration set and the validation sets for Level-1 while it remained similar for Level-2 and Level-3 with differences of around 0.02 to 0.05 between the Pearson correlation coefficient obtained for calibration set and the median Pearson correlation coefficient obtained for validation sets.

For the validation sets, the correlation decreased slowly but steadily as features were added for Level-1. For Level-2, the median correlation remained consistently strong

( $r > 0.70$ ) from one selected band to 25 bands, then steadily decreased to reach a median correlation around 0.29 when using all bands. For Level-3, the maximum median correlation was reached for only 7 selected PCs and was lower than the maximum obtained on the calibration set.

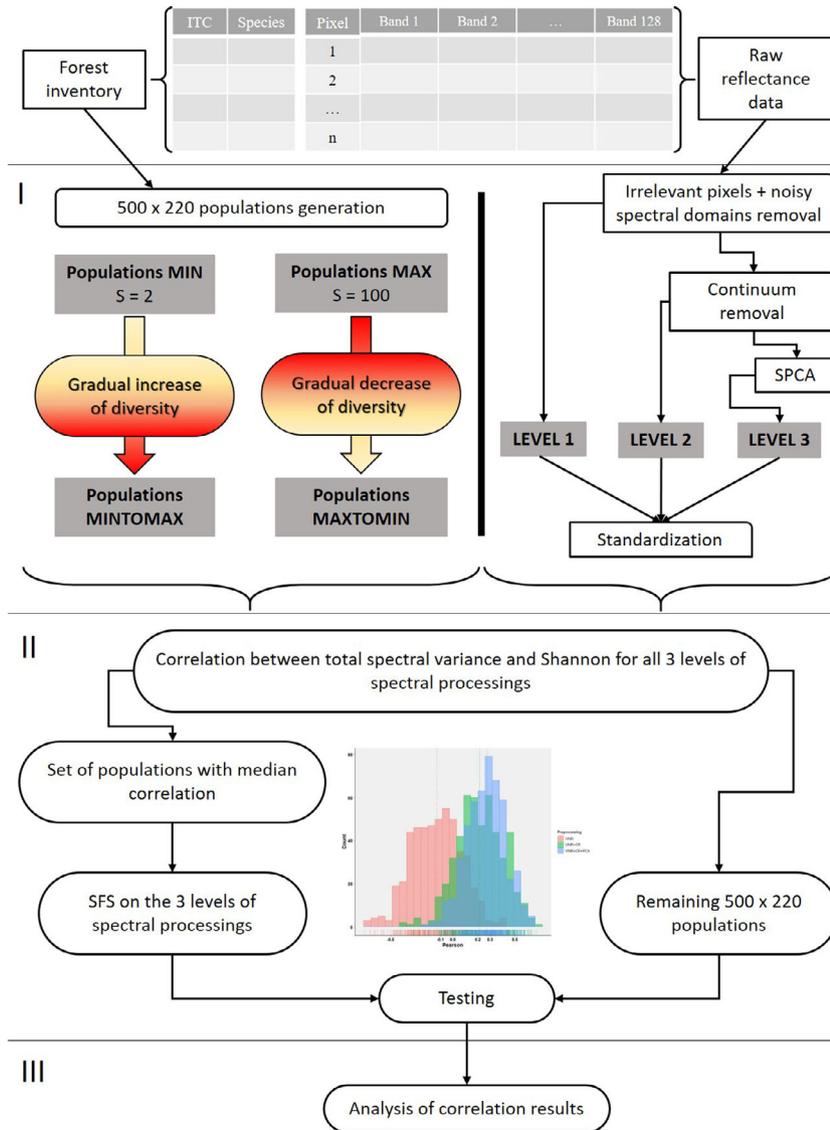
The comparison of the distribution of correlations when using all spectral features (Fig. 4) and when using the optimal number of features (Fig. 5) highlights the dramatic improvement of the strength of the linear relationship between taxonomic diversity and  $\text{Var}_{\text{TOT}}$  resulting from feature selection.

### Effect of SFS on the partitioning of spectral variance

We analyzed the influence of feature selection on variance partitioning into inter-species, inter-crown, and residual components. We computed the evolution of the correlation between  $\text{Var}_{\text{SP}}$  and taxonomic diversity with an increasing number of spectral features, for all 501 sets (Fig. 6) and compared the distribution of  $\text{Var}_{\text{SP}}$ ,  $\text{Var}_{\text{ITCI}_{\text{SP}}}$ , and  $\text{Var}_{\text{RES}}$  when using the full spectral information or the optimal set of features identified from SFS (Fig. 7).

$\text{Var}_{\text{SP}}$  showed systematically stronger correlation with diversity than  $\text{Var}_{\text{TOT}}$ . The correlation between  $\text{Var}_{\text{SP}}$  and any of the three diversity indices was stable over all the sets, with limited influence of the feature selection. In the case of Level-2 processing, dimensionality reduction resulted in a strong decrease of its variability across sets (Fig. 6B), which was not the case for other levels of processing.

The improvement in correlation between diversity and  $\text{Var}_{\text{TOT}}$  following dimensionality reduction did not reflect a change in the structure of spectral variance and notably no systematic increase in the share of interspecies variance

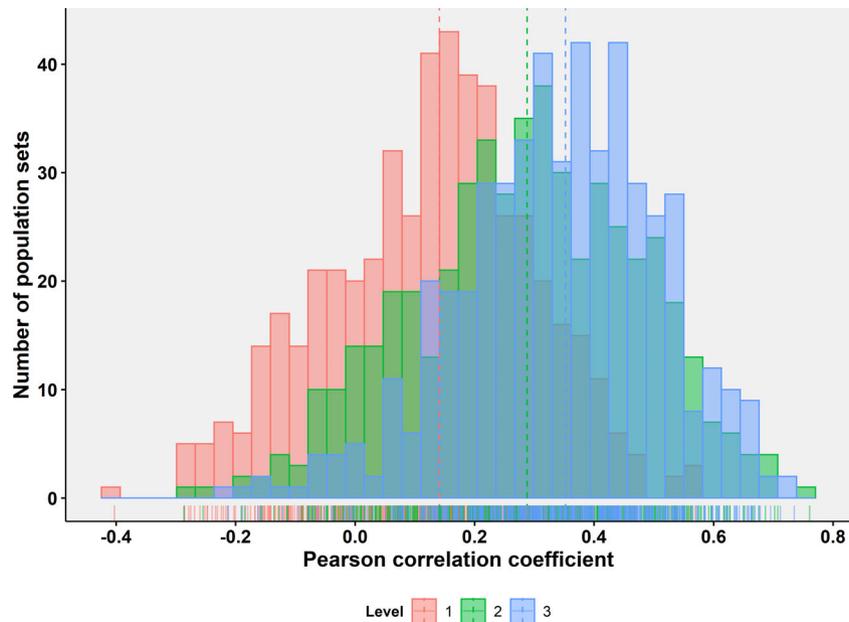


**Figure 2.** Summary of the different treatments applied to the data. The effect of these three levels of spectral processings is then compared with the results obtained with the SFS.

(Fig. 7). Feature selection did not increase the proportion of  $Var_{SP}$  for Level-1 and Level-2, confirming that the proportion of  $Var_{SP}$  is a poor predictor of the correlation between  $Var_{TOT}$  and the Shannon index. Moreover, for Level-3, feature selection resulted in a decrease in the share of  $Var_{RES}$  in favour of  $Var_{SP}$ , but this had no direct effect on the correlation between  $Var_{SP}$  and Shannon index since it remained constant (Fig. 6C). The selected features maximized the correlation between  $Var_{TOT}$  and diversity indices without changing either the proportion of  $Var_{SP}$  or its correlation with these indices, which means that SFS does not prioritize bands related to the species contribution to the total spectral variance.

**Characterization of the bands selected by the SFS and identification of the spectral domains of interest**

The contribution of the spectral information corresponding to bands selected from SFS applied on Level-2 data is illustrated in Figure 8. The maximum correlation ( $R_{MAX}$ ) between  $Var_{TOT}$  and the Shannon index was reached for the first 12 selected features. Most of the twelve features were located in the red edge part of the near infrared domain. The correlation remained greater than or equal to 95% of  $R_{MAX}$  for 3 to 32 selected features, including spectral bands in the red edge, near infrared, green



**Figure 3.** Distribution of the Pearson correlation coefficient between the Shannon diversity index and the VarTOT computed on all spectral features, for the 501 sets of 220 populations and the three levels of spectral processing.

domain of the visible region (530 to 560 nm), and one spectral band in the blue domain corresponding to the first spectral band. The spectral bands corresponding to the blue and the red domains showed negative contribution to the link between taxonomic diversity and spectral variance, as their addition resulted in suboptimal correlation. These spectral domains correspond to absorption domains of chlorophylls, and the absorption corresponding to these domains tends to saturate for dense and highly photosynthetic canopies. The results obtained for Level-1 data are presented in Appendix S1. The bands belonging to the red edge were not selected by the SFS, which shows that the existing information in the raw reflectance data needs to be processed using appropriate transformation or normalization to get the most of spectral information.

## Discussion

The aim of this study was to investigate and quantify the link between spectral variance and taxonomic diversity on artificial populations covering a wide range of taxonomic diversity. Our results highlighted the importance of spectral processing and feature selection in maximizing this relationship. We showed that the correlation between total variance and taxonomic diversity was relatively weak, while, as expected, interspecies variance and taxonomic diversity were strongly correlated. However, the interspecies variance represented less than 50% of the total spectral variance and feature selection did not significantly increase this proportion.

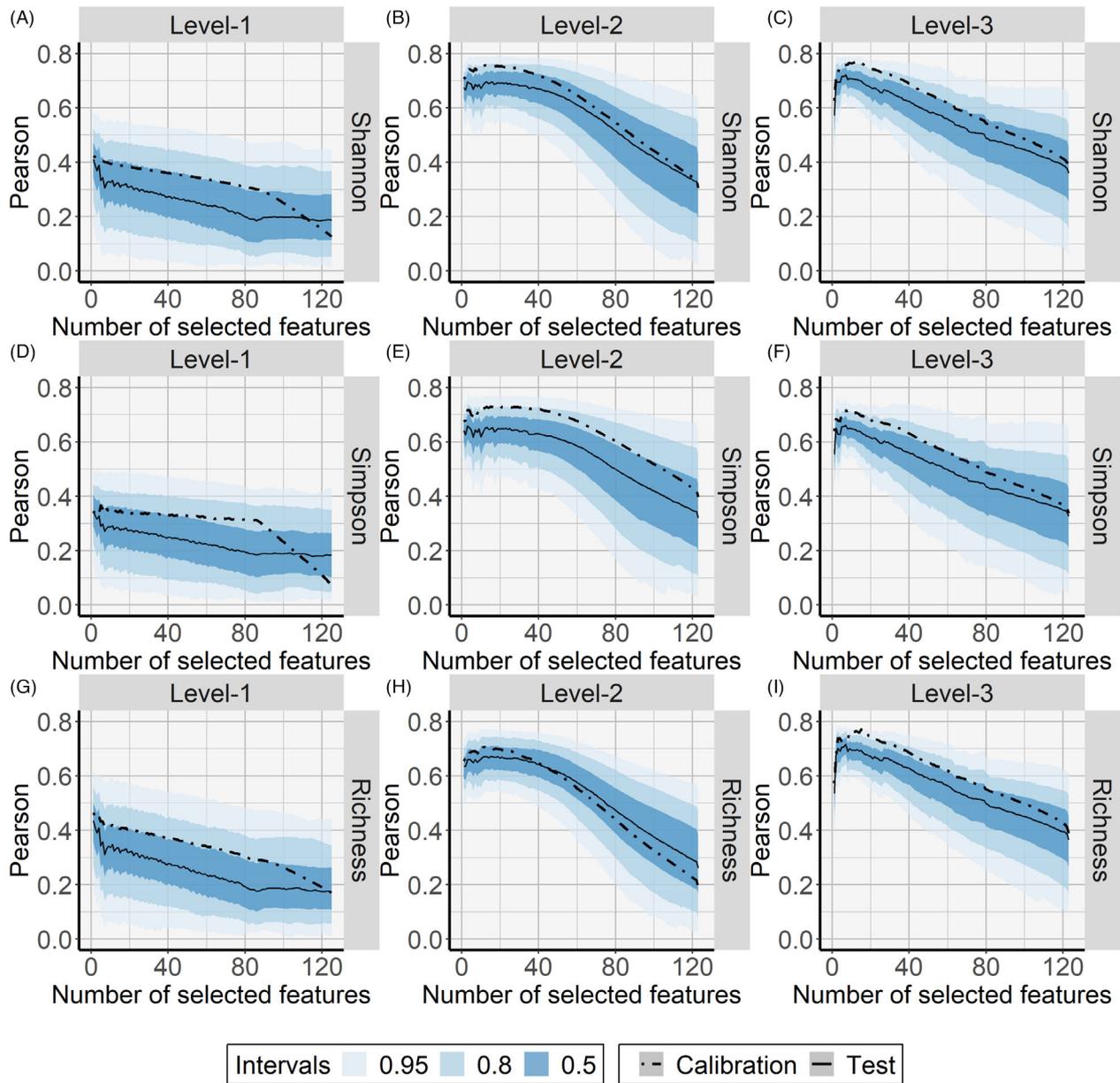
## Impact of spectral processing on the correlation between spectral variance and taxonomic diversity

We observed that Var<sub>TOT</sub> derived from Level-1 reflectance was poorly correlated with taxonomic diversity, but spectral processing including normalization and feature selection resulted in a strong increase of this correlation. The different population sets showed highly variable correlations between Var<sub>TOT</sub> and diversity indices when all features were used. After optimal feature selection from a calibration set, the variability of the correlations between Var<sub>TOT</sub> and Shannon diversity index obtained from independent test sets was strongly reduced. These results are encouraging with regard to the possibility of mapping upper canopy taxonomic diversity from imaging spectroscopy. The applicability of the findings obtained with Level-3 processing on this subset of the image is less straightforward than those corresponding to Level-2 processing. The computation of PCA over this full imaging spectroscopy dataset will inevitably result in very different components, therefore it will be necessary to perform the feature selection again on these new components.

## Methodological considerations

### Artificial population generation

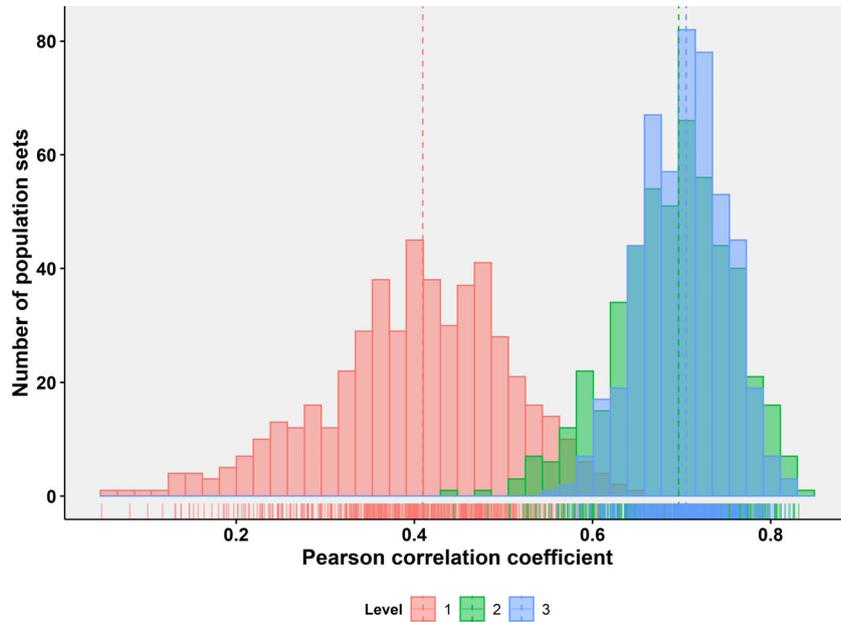
Our population generation procedure implies that the size of all crowns of the artificial populations is the same. We made this choice in order to guarantee that the



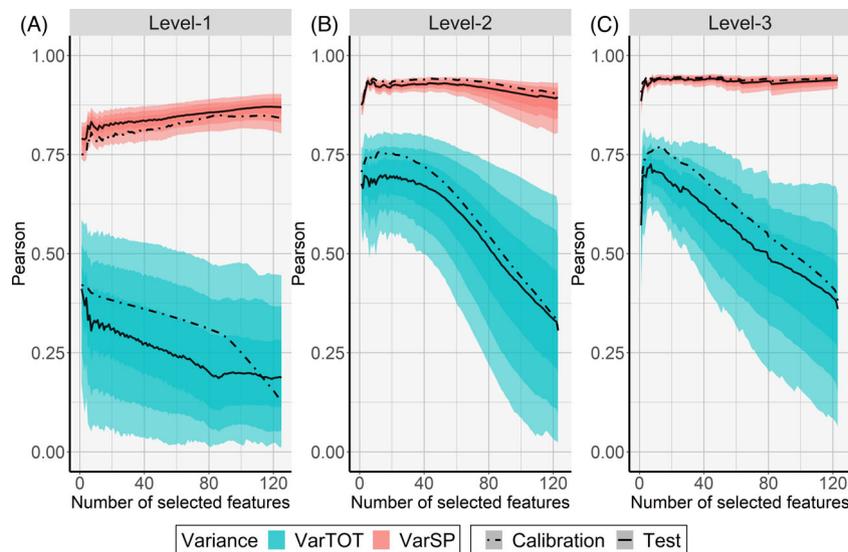
**Figure 4.** Evolution of the Pearson correlation coefficient between total spectral variance and taxonomic diversity, for an increasing number of spectral features selected from SFS applied to Level-1 spectral processing (VNIR, A, E, G), SFS applied to Level-2 spectral processing (VNIR+CR, B, F, H) and SFS applied to Level-3 spectral processing (VNIR+CR + PCA, C, G, I). The taxonomic diversity is expressed as Shannon index (A–C), Simpson index (D–F), and species richness (G–I).

estimation of the taxonomic diversity of the communities is done on communities of the same size (number of individuals and pixels, hence same surface). This choice was made to evaluate the link between taxonomic diversity and spectral variance under controlled conditions in order to limit the possible additional source of variation on the spectral variance. This crown size effect, and thus the relative proportion of spectral variance represented by

some individuals, will have to be taken into account when applying this method to real-world data. Moreover, the reflectance of some pixels may be the result of a mixture of several individuals that overlap or are intermingled. The crowns were carefully selected and identified in the study, which allowed to limit occurrence of mixed pixels without resorting to a spatial buffer applied to each delineated crown.



**Figure 5.** Distribution of the Pearson correlation coefficient calculated on the 501 sets of 220 populations between VarTOT calculated on the best features selection and Shannon index for the three types of preprocessing.

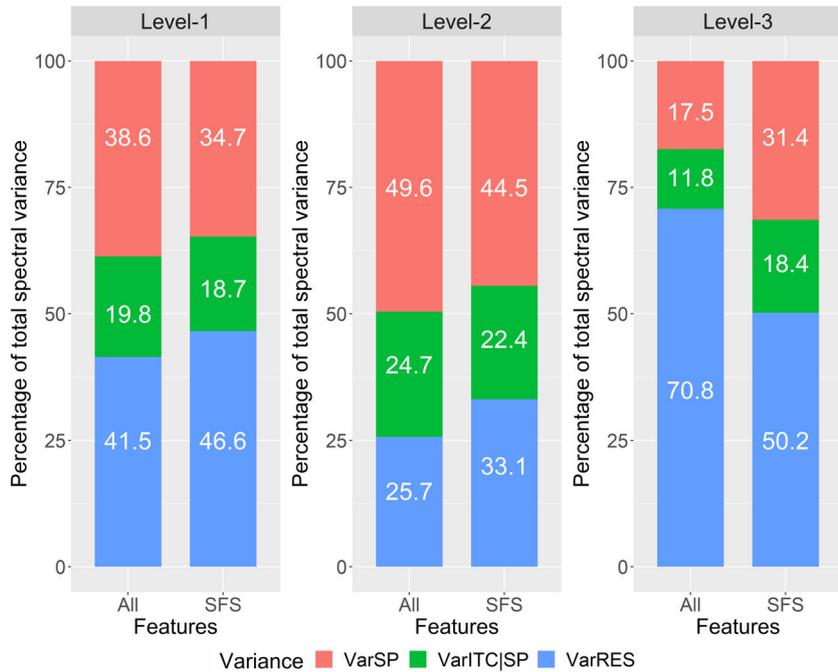


**Figure 6.** Evolution of the Pearson correlation coefficient between the Shannon index and VarTOT or VarSP, for each spectral processing.

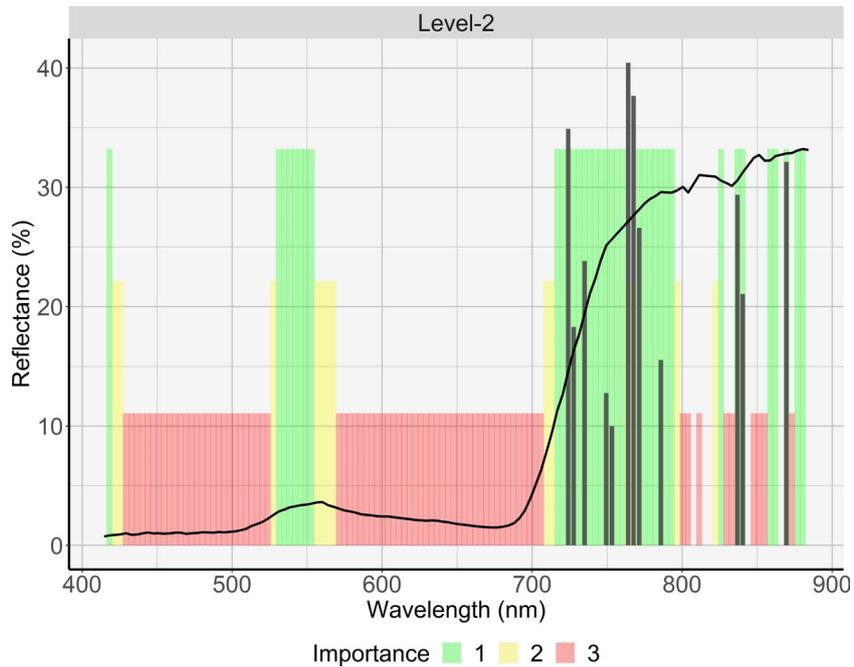
**Interpretation of the spectral features selected by SFS**

The first spectral band, belonging to the blue domain, has been identified by the SFS as important. The application of CR on a VNIR reflectance spectrum corresponding to green vegetation is most likely to produce a first segment between the first spectral band and the NIR shoulder corresponding to higher wavelengths of the red edge. In this

situation, the CR reflectance of the first spectral band from the blue domain would be directly related to the slope of the segment, and to the value of NIR reflectance. The green and red edge domains are particularly sensitive to photosynthetic pigment content of foliage in the case of dense canopies with high photosynthetic pigment content, suggesting the link between taxonomic information, photosynthetic activity and spectral information (Cavender-Bares et al., 2020).



**Figure 7.** Comparison of the distribution of inter-species spectral variance (VarSP), inter-ITC within species spectral variance (VarITC|SP), and residual spectral variance (VarRES) using all features or the selected features for the three pretreatments.



**Figure 8.** Spectral features selection obtained with SFS applied on calibration dataset for Level-2 processing. Feature importance: 1 = selection of features required to reach maximum correlation (RMax) between Shannon’s H and VarTOT, and remain  $\geq 95\%$  of RMax; 2 = additional features resulting in correlation between 90 and 95% of RMax ( $0.69 \leq R < 0.72$ ); 3 = addition of features resulting in correlation  $< 90\%$  of RMax. Gray bars correspond to the 12 features identified to reach maximum correlation, with importance represented by their height.

These results are in line with other studies, in particular, with the fact that the absence of SWIR has an influence on the selection of bands in the visible, red-edge and to some extent the NIR (Hennessy et al., 2020; Rivard et al., 2008). We did not examine the influence of spectral resolution in this study; however, the importance of the blue domain and the red-edge were reported for both hyperspectral (Hennessy et al., 2020) and multispectral data (Immitzer et al., 2016).

### Partial estimation of biodiversity of tropical forest

This study focused on canopy tree diversity, which is merely a subset of forest plant diversity. Tropical forests contain many forms of plant growth. Trees account for only 25% of all tropical forest species (Gentry & Dodson, 1987; Wright, 2002), and lianas can account for up to 23% of the remaining species (Gentry & Dodson, 1987). Our approach does not take into account the species in the lower layers of the canopy. A review of the recent progress made in the study of liana ecology using terrestrial, aerial, and space-based remote sensing is provided by Heijden et al. (2022).

### Effectiveness of feature selection for prediction of taxonomic diversity

Our population generation procedure was designed to test the relationship between spectral variance and taxonomic diversity for different combinations of richness and equitability. For the populations, which are the most representative of the actual diversity observed in Paracou (Fig. 1) the correlation between  $\text{Var}_{\text{TOT}}$  and Shannon index became positive when using the selected features: it reached a median value of 0.21 for the Level-1 and 0.23 for Level-2 processing. For Level-3 processing, we found a positive correlation both when using all the components and when using the selection and the correlation reached the median value of 0.29 when using the selected features. The maximum correlation obtained for this range of taxonomic diversity remained weak, suggesting that the direct relationship between spectral variance and taxonomic diversity is challenging the applicability of an approach based on spectral variance for operational biodiversity mapping over tropical rainforests using imaging spectroscopy. Here, we used the full spectral range of the VNIR sensor, but the airplane also carried a sensor for SWIR (Short Wave Infrared) data acquisition. The addition of SWIR improves discrimination of species (Laybros et al., 2019) and thus could have provided useful information. However, the signal-to-noise ratio of SWIR is much lower compared with VNIR (Laybros et al., 2020). It is possible that expanding the spectrum by

incorporating SWIR data into the analysis would have led to improved results, but because different sensors were used for the acquisition, geometric coherence was more difficult to ensure over the entire 400 to 2500 nm spectrum, which resulted in artefacts when applying continuum removal.

$\text{Var}_{\text{SP}}$  was consistently more strongly correlated with taxonomic diversity than  $\text{Var}_{\text{TOT}}$ . However, in an operational framework of biodiversity mapping,  $\text{Var}_{\text{SP}}$  is not accessible, hence the importance of developing methods to optimize the link between spectral variance as measured on an image and taxonomic diversity. As shown in Figures 6 and 7, the increase in correlation between spectral variance and diversity indices obtained with SFS was not related to a change in the level of inter-species variance nor to an increase in correlation between inter-species variance and diversity indices. In order to understand the drivers of increased correlation induced by SFS, we have examined different hypotheses. Since the increase in correlation is not related to an increase in the proportion of variance explained by the species, we surmised that it could be due to a homogenisation of intra-species variance within populations. However, calculations of intra-specific variances did not unambiguously support this hypothesis. We also analyzed the distribution and variability of the paired distances between mean spectra of species but found that the SFS did not clearly reduce their variability. The mechanisms by which feature selection strongly contributed to improve the prediction of taxonomic diversity from spectral variance are not well understood yet and deserve further scrutiny.

The general application of the method described in this study needs careful evaluation. The procedure developed to select spectral features from SFS applied on a calibration set led to suboptimal performances when applied to test datasets. Although this feature selection significantly improved overall performances, the selection of features obtained using the calibration set was not the optimal selection for all populations sets. This procedure may be adjusted by applying SFS to multiple calibration datasets and selecting features based on the ranking and frequency of occurrences among multiple calibration datasets. This may improve the stability of the performance of the SFS, but the relevance of this operation on pseudo-images is questionable: it would not guarantee a better performance once applied on the whole image.

### Comparison with other approaches

We chose total variance as spectral diversity metric, but other metrics could have been used, such as the mean distance from the spectral centroid (Rocchini, 2007;

Rocchini et al., 2010) or the coefficient of variation (CV) of spectral reflectance (Wang et al., 2018). We calculated the CV of each population and the linear correlation results with the three diversity indices were similar to those shown in Figure 5.

An advantage of methods using spectral variation to estimate taxonomic diversity over methods based on spectral species clustering (Féret & Asner, 2014) is that they require less user choice and are therefore more objective. Féret and Asner (2014), Laliberté et al. (2019), Wang et al. (2018) applied PCA to reflectance data and selected relevant components based on visual analysis. Our feature selection method using SFS algorithm circumvents the need for subjective choices to be made. However, SFS depends on ground information availability.

Oldeland et al. (2010) and Wang et al., 2018 found better correlations between spectral variability metrics and Shannon or Simpson indices, which consider both richness and abundance, than with species richness alone. In our study, the correlation coefficient between spectral variance and taxonomic diversity showed minor differences when using species richness or Shannon index as diversity metric (Fig. 5). In this same study, Wang et al. showed that when the richness of simulated communities increased but equitability was low (high abundance of dominant species) the spectral variability, measured *via* the CV, remained low. They also reported a case where the spectral variability of a plot with high simulated diversity was exceeded by one low diversity plot dominated by a species with high intraspecific variability. The leaf traits of some species and their structure had a strong influence on optical diversity, as did the structure of the canopy, and could modify the optical diversity-vegetation diversity relationship. Rossi et al. (2021) concluded in their study that the spectral diversity-biodiversity relationship studied at the leaf level (Frye et al., 2021; Schweiger et al., 2018) cannot be easily transferred to the plant or community level and that the impact of canopy structure on the relationship between spectral diversity and taxonomic diversity needs to be better understood.

An earlier study by Vaglio Laurin et al. (2014) conducted in Ghana examined how diversity of 64 0.125 ha tropical forest plots could be predicted from imaging spectroscopy. The authors reported a pseudo- $R^2$  value of 0.85 when comparing the Shannon diversity measured from plot inventories with the predicted value derived from spectral information acquired at 1-m spatial. The approach developed in that study predicted the Shannon diversity in 64 plots from 744 predictors (186 bands  $\times$  4 metrics) using random forest regression. In our study, we used a single predictor (spectral variance) and a simple

linear model to relate this predictor to taxonomic diversity.

## Conclusion

Our study demonstrates the link between spectral variance and taxonomic diversity in a simulated diversity gradient built from spectral information extracted from imaging spectroscopy acquired over hyperdiverse dense tropical forest canopy. We found that the inter-species spectral variance represented between 17% and 50% of the total spectral variance depending on image processing (Fig. 8). We evidenced the importance of reflectance normalization for improving the correlation between spectral variance and taxonomic diversity indices. We showed that feature selection significantly enhanced predictability of taxonomic diversity from spectral variance. This study allowed us to identify which spectral domains maximized the link between spectral variance and taxonomic diversity. We found that the red-edge and NIR domains contained the most useful information for estimating taxonomic diversity in a tropical rainforest. This study has contributed to the foundation for a spectral-spatial analysis to be conducted at the landscape scale at our site. Forthcoming analyses will test its ability to map spatial patterns of taxonomic diversity using spectral variance.

## ACKNOWLEDGEMENTS

The authors would like to thank the CNES for the acquisition of hyperspectral data over Paracou experimental site and the Laboratoire d'Excellence CEBA (ANR-10-LABX-25). J.-B. Féret acknowledges financial support from Agence Nationale de la Recherche (France) (BioCop project—ANR-17-CE32-0001) and TOSCA program grant from the French Space Agency (CNES) (HyperTropik project). Colette Badourdine benefited from a PhD grant co-funded by CNES and IRD, France.

## REFERENCES

- Aguirre-Gutiérrez, J., Rifai, S., Shenkin, A., Oliveras, I., Bentley, L.P., Svátek, M. et al. (2021) Pantropical modelling of canopy functional traits using Sentinel-2 remote sensing data. *Remote Sensing of Environment*, **252**, 112122. <https://doi.org/10.1016/j.rse.2020.112122>
- Alamgir, M., Campbell, M.J., Sloan, S., Goosem, M., Clements, G.R., Mahmoud, M.I. et al. (2017) Economic, socio-political and environmental risks of road development in the tropics. *Current Biology*, **27**, R1130–R1140. <https://doi.org/10.1016/j.cub.2017.08.067>
- Asner, G.P., Knapp, D.E., Boardman, J., Green, R.O., Kennedy-Bowdoin, T., Eastwood, M. et al. (2012) Carnegie

- airborne Observatory-2: increasing science data dimensionality via high-fidelity multi-sensor fusion. *Remote Sensing of Environment*, **124**, 454–465. <https://doi.org/10.1016/j.rse.2012.06.012>
- Asner, G.P. & Martin, R.E. (2009) Airborne spectranomics: mapping canopy chemical and taxonomic diversity in tropical forests. *Frontiers in Ecology and the Environment*, **7**, 269–276. <https://doi.org/10.1890/070152>
- Aubry-Kientz, M., Dutrieux, R., Ferraz, A., Saatchi, S., Hamraz, H., Williams, J. et al. (2019) A comparative assessment of the performance of individual tree crowns delineation algorithms from ALS data in tropical forests. *Remote Sensing*, **11**, 1086. <https://doi.org/10.3390/rs11091086>
- Baldeck, C.A., Asner, G.P., Martin, R.E., Anderson, C.B., Knapp, D.E., Kellner, J.R. et al. (2015) Operational tree species mapping in a diverse tropical Forest with airborne imaging spectroscopy. *PLoS One*, **10**, e0118403. <https://doi.org/10.1371/journal.pone.0118403>
- Cavender-Bares, J., Gamon, J.A. & Townsend, P.A. (2020) *Remote sensing of plant biodiversity*. Cham: Springer International Publishing. <https://doi.org/10.1007/978-3-030-33157-3>
- Clark, M.L. & Roberts, D.A. (2012) Species-level differences in hyperspectral metrics among tropical rainforest trees as determined by a tree-based classifier. *Remote Sensing*, **4**, 1820–1855. <https://doi.org/10.3390/rs4061820>
- Clark, R.N. & Roush, T.L. (1984) Reflectance spectroscopy: quantitative analysis techniques for remote sensing applications. *Journal of Geophysical Research*, **89**, 6329–6340. <https://doi.org/10.1029/JB089iB07p06329>
- Fassnacht, F.E., Müllerová, J., Conti, L., Malavasi, M. & Schmidlein, S. (2022) About the link between biodiversity and spectral variation. *Applied Vegetation Science*, **25**, e12643. <https://doi.org/10.1111/avsc.12643>
- Féret, J.-B. & Asner, G.P. (2014) Mapping tropical forest canopy diversity using high-fidelity imaging spectroscopy. *Ecological Applications*, **24**, 1289–1296. <https://doi.org/10.1890/13-1824.1>
- Féret, J.-B. & de Boissieu, F. (2020) biodivMapR: an R package for  $\alpha$ - and  $\beta$ -diversity mapping using remotely sensed images. *Methods in Ecology and Evolution*, **11**, 64–70. <https://doi.org/10.1111/2041-210X.13310>
- Frye, H.A., Aiello-Lammens, M.E., Euston-Brown, D., Jones, C.S., Kilroy Mollmann, H., Merow, C. et al. (2021) Plant spectral diversity as a surrogate for species, functional and phylogenetic diversity across a hyper-diverse biogeographic region. *Global Ecology and Biogeography*, **30**, 1403–1417. <https://doi.org/10.1111/geb.13306>
- Gamon, J.A., Serrano, L. & Surfus, J.S. (1997) The photochemical reflectance index: an optical indicator of photosynthetic radiation use efficiency across species, functional types, and nutrient levels. *Oecologia*, **112**, 492–501. <https://doi.org/10.1007/s004420050337>
- Gentry, A.H. & Dodson, C. (1987) Contribution of nontrees to species richness of a tropical rain Forest. *Biotropica*, **19**, 149. <https://doi.org/10.2307/2388737>
- Gholizadeh, H., Gamon, J.A., Townsend, P.A., Zyguelbaum, A.I., Helzer, C.J., Hmimina, G.Y. et al. (2019) Detecting prairie biodiversity with airborne remote sensing. *Remote Sensing of Environment*, **221**, 38–49. <https://doi.org/10.1016/j.rse.2018.10.037>
- Gholizadeh, H., Gamon, J.A., Zyguelbaum, A.I., Wang, R., Schweiger, A.K. & Cavender-Bares, J. (2018) Remote sensing of biodiversity: soil correction and data dimension reduction methods improve assessment of  $\alpha$ -diversity (species richness) in prairie ecosystems. *Remote Sensing of Environment*, **206**, 240–253. <https://doi.org/10.1016/j.rse.2017.12.014>
- Gourlet-Fleury, S., Guehl, J.-M. & Laroussinie, O. (2004) *Ecology and management of a neotropical rainforest: lessons drawn from Paracou, a long-term experimental research site in French Guiana*. Paris: Elsevier.
- Hall, K., Johansson, L.J., Sykes, M.T., Reitalu, T., Larsson, K. & Prentice, H.C. (2010) Inventorying management status and plant species richness in semi-natural grasslands using high spatial resolution imagery. *Applied Vegetation Science*, **13**, 221–233. <https://doi.org/10.1111/j.1654-109X.2009.01063.x>
- Heijden, G.M.F., Proctor, A.D.C., Calders, K., Chandler, C.J., Field, R., Foody, G.M. et al. (2022) Making (remote) sense of lianas. *Journal of Ecology*, **110**, 498–513. <https://doi.org/10.1111/1365-2745.13844>
- Hennessy, A., Clarke, K. & Lewis, M. (2020) Hyperspectral classification of plants: a review of waveband selection Generalisability. *Remote Sensing*, **12**, 113. <https://doi.org/10.3390/rs12010113>
- Immitzer, M., Vuolo, F. & Atzberger, C. (2016) First experience with Sentinel-2 data for crop and tree species classifications in Central Europe. *Remote Sensing*, **8**, 166. <https://doi.org/10.3390/rs8030166>
- Laliberté, E., Schweiger, A.K. & Legendre, P. (2019) Partitioning plant spectral diversity into alpha and beta components (preprint). *Ecology*, **23**, 370–380. <https://doi.org/10.1101/742080>
- Laybros, A., Aubry-Kientz, M., Féret, J.-B., Bedeau, C., Brunaux, O., Derroire, G. et al. (2020) Quantitative airborne inventories in dense tropical forest using imaging spectroscopy. *Remote Sensing*, **12**(10), 1577. <https://doi.org/10.3390/rs12101577>
- Laybros, A., Schläpfer, D., Féret, J.-B., Descroix, L., Bedeau, C., Lefevre, M.-J. et al. (2019) Across date species detection using airborne imaging spectroscopy. *Remote Sensing*, **11**, 789. <https://doi.org/10.3390/rs11070789>
- Legendre, P. & Legendre, L. (1998) *Numerical ecology, 2nd English ed. ed, developments in environmental modelling*. Amsterdam; New York: Elsevier.
- Ma, X., Mahecha, M.D., Migliavacca, M., van der Plas, F., Benavides, R., Ratcliffe, S. et al. (2019) Inferring plant

- functional diversity from space: the potential of Sentinel-2. *Remote Sensing of Environment*, **233**, 111368. <https://doi.org/10.1016/j.rse.2019.111368>
- Marcon, E. (2015) Mesures de la Biodiversité. Master. Kourou, France. (cel-01205813v2).
- Mather, P.M. (1999) *Computer processing of remotely-sensed images: an introduction*, 2nd edition. New York: Wiley, Chichester.
- Nagendra, H. & Rocchini, D. (2008) High resolution satellite imagery for tropical biodiversity studies: the devil is in the detail. *Biodiversity and Conservation*, **17**, 3431–3442. <https://doi.org/10.1007/s10531-008-9479-0>
- Oindo, B.O. & Skidmore, A.K. (2002) Interannual variability of NDVI and species richness in Kenya. *International Journal of Remote Sensing*, **23**, 285–298. <https://doi.org/10.1080/01431160010014819>
- Oldeland, J., Wesuls, D., Rocchini, D., Schmidt, M. & Jürgens, N. (2010) Does using species abundance data improve estimates of species diversity from remotely sensed spectral heterogeneity? *Ecological Indicators*, **10**, 390–396. <https://doi.org/10.1016/j.ecolind.2009.07.012>
- O'Neill, D.W., Fanning, A.L., Lamb, W.F. & Steinberger, J.K. (2018) A good life for all within planetary boundaries. *Nature Sustainability*, **1**, 88–95. <https://doi.org/10.1038/s41893-018-0021-4>
- Palmer, M.W., Wohlgemuth, T., Earls, P., Arévalo, J.R. & Thompson, S. (2000) Opportunities for long-term ecological research at the tallgrass prairie preserve, Oklahoma. Cooperation in long term ecological research in central and Eastern Europe: proceedings 123–128.
- Rapinel, S., Mony, C., Lecoq, L., Clément, B., Thomas, A. & Hubert-Moy, L. (2019) Evaluation of Sentinel-2 time-series for mapping floodplain grassland plant communities. *Remote Sensing of Environment*, **223**, 115–129. <https://doi.org/10.1016/j.rse.2019.01.018>
- Rivard, B., Sanchez-Azofeifa, A., Foley, S. & Calvo-Alvarado, J. (2008) Species classification of tropical tree leaf reflectance and dependence on selection of spectral bands. In: Kalacska, M. & Sanchez-Azofeifa, A. (Eds.) *Hyperspectral remote sensing of tropical and sub-tropical forests*. CRC Press, pp. 141–159. <https://doi.org/10.1201/9781420053432.ch6>
- Rocchini, D. (2007) Effects of spatial and spectral resolution in estimating ecosystem  $\alpha$ -diversity by satellite imagery. *Remote Sensing of Environment*, **111**, 423–434. <https://doi.org/10.1016/j.rse.2007.03.018>
- Rocchini, D., Balkenhol, N., Carter, G.A., Foody, G.M., Gillespie, T.W., He, K.S. et al. (2010) Remotely sensed spectral heterogeneity as a proxy of species diversity: recent advances and open challenges. *Ecological Informatics*, **5**, 318–329. <https://doi.org/10.1016/j.ecoinf.2010.06.001>
- Rodarmel, C. & Shan, J. (2002) Principal component analysis for hyperspectral image classification. *Surveying and Land Information Science*, **62**, 115–122.
- Rossi, C., Kneubühler, M., Schütz, M., Schaepman, M.E., Haller, R.M. & Risch, A.C. (2021) Spatial resolution, spectral metrics and biomass are key aspects in estimating plant species richness from spectral diversity in species-rich grasslands. *Remote Sensing in Ecology and Conservation*, **8**, 297–314. <https://doi.org/10.1002/rse2.244>
- Ruiz Hidalgo, D., Bacca Cortés, B. & Caicedo Bravo, E. (2021) Dimensionality reduction of hyperspectral images of vegetation and crops based on self-organized maps. *Information Processing in Agriculture*, **8**, 310–327. <https://doi.org/10.1016/j.inpa.2020.07.002>
- Schmidtlein, S. & Fassnacht, F.E. (2017) The spectral variability hypothesis does not hold across landscapes. *Remote Sensing of Environment*, **192**, 114–125. <https://doi.org/10.1016/j.rse.2017.01.036>
- Schweiger, A.K., Cavender-Bares, J., Townsend, P.A., Hobbie, S.E., Madritch, M.D., Wang, R. et al. (2018) Plant spectral diversity integrates functional and phylogenetic components of biodiversity and predicts ecosystem function. *Nature Ecology & Evolution*, **2**, 976–982. <https://doi.org/10.1038/s41559-018-0551-1>
- Serbin, S.P. & Townsend, P.A. (2020) Scaling functional traits from leaves to canopies. In: Cavender-Bares, J., Gamon, J.A. & Townsend, P.A. (Eds.) *Remote sensing of plant biodiversity*. Cham: Springer International Publishing, pp. 43–82. [https://doi.org/10.1007/978-3-030-33157-3\\_3](https://doi.org/10.1007/978-3-030-33157-3_3)
- Shannon, C.E. (1948) A mathematical theory of communication. *Bell System Technical Journal*, **27**, 379–423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>
- Simpson, E.H. (1949) Measurement of diversity. *Nature*, **163**, 688. <https://doi.org/10.1038/163688a0>
- Somers, B. & Asner, G.P. (2013) Multi-temporal hyperspectral mixture analysis and feature selection for invasive species mapping in rainforests. *Remote Sensing of Environment*, **136**, 14–27. <https://doi.org/10.1016/j.rse.2013.04.006>
- ter Steege, H., Pitman, N.C.A., Sabatier, D., Baraloto, C., Salomao, R.P., Guevara, J.E. et al. (2013) Hyperdominance in the Amazonian tree Flora. *Science*, **342**, 1243092. <https://doi.org/10.1126/science.1243092>
- Theodoridis, S., Theodoridis, S. & Koutroumbas, K. (2014) *Pattern recognition*. Burlington: Elsevier Science.
- Ustin, S.L. & Gamon, J.A. (2010) Remote sensing of plant functional types. *New Phytologist*, **186**, 795–816. <https://doi.org/10.1111/j.1469-8137.2010.03284.x>
- Vaglio Laurin, G., Chan, J.C.-W., Chen, Q., Lindsell, J.A., Coomes, D.A., Guerriero, L. et al. (2014) Biodiversity mapping in a tropical west African Forest with airborne hyperspectral data. *PLoS One*, **9**, e97910. <https://doi.org/10.1371/journal.pone.0097910>
- Wang, R. & Gamon, J.A. (2019) Remote sensing of terrestrial plant biodiversity. *Remote Sensing of Environment*, **231**, 111218. <https://doi.org/10.1016/j.rse.2019.111218>
- Wang, R., Gamon, J.A., Cavender-Bares, J., Townsend, P.A. & Zyguelbaum, A.I. (2018) The spatial sensitivity of the

spectral diversity–biodiversity relationship: an experimental test in a prairie grassland. *Ecological Applications*, **28**, 541–556. <https://doi.org/10.1002/eap.1669>

Wright, J.S. (2002) Plant diversity in tropical forests: a review of mechanisms of species coexistence. *Oecologia*, **130**, 1–14. <https://doi.org/10.1007/s004420100809>

## Supporting Information

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

**Table S1** Distribution for each level of spectral treatments of the features providing the maximum Pearson correlation coefficient, at least 95% of the maximum and at least 70% of the maximum.

**Figure S2.** Spectral features selection obtained with SFS applied on calibration dataset for Level-1 processing. Feature importance.