Title page

Species recognition in herbaria and standing trees using an accessible portable spectrometer

Flavia Machado Durgante[a,b,c,*], Caroline da Cruz Vasconcelos[a,b], Hilana Louise Hadlich[a], Caroline Lorenci Mallmann[d], Ricardo de Oliveira Perdiz[e], Kaio Cesar Marinho da Cunha[a,b], Samyra Ramos Chaves[a,b], Celio Pasquini[f], Kelly Torralvo[g], Michael John Gilbert Hopkins[b], Jochen Schöngart[a], Maria Teresa Fernandez Piedade[a], Priscila Amaral de Sá[h], Maíra Luciana Guimarães Conde[a], Florian Wittmann[a,c]

[a]Ecology, Monitoring and Sustainable Use of Wetlands (MAUA), National Institute of Amazonian Research (INPA), Manaus, Amazonas, Brazil
[b]Herbarium, National Institute of Amazonian Research (INPA), Manaus, Amazonas, Brazil
[c]Department of Wetland Ecology, Karlsruhe Institute of Technology (KIT), Rastatt, Germany
[d]Department of Geoscience, Federal University of Santa Maria (UFSM), Santa Maria, Rio Grande do Sul, Brazil
[e]Postgraduate Program in Natural Resources (PRONAT), Federal University of Roraima (UFRR), Boa Vista, Roraima, Brazil
[f]Department of Analytical Chemistry, State University of Campinas (UNICAMP), Campinas, Brazil
[g]Integrative Ecology Lab, Mamirauá Sustainable Development Institute (IDSM), Tefé, Amazonas, Brazil
[h]Postgraduate Program in Ecology and Conservation (PPGEC), State University of Mato Grosso (UNEMAT), Nova Xavantina, Mato Grosso, Brazil

* Corresponding author at: Ecology, Monitoring and Sustainable Use of Wetlands (MAUA), National Institute of Amazonian Research (INPA), Manaus, Amazonas, Brazil
E-mail address: flaviamdflorestal@gmail.com (F. M. Durgante)

ORCID

Flavia Machado Durgante https://orcid.org/0000-0002-5517-8821
Caroline da Cruz Vasconcelos https://orcid.org/0000-0002-7850-6672

Hilana Louise Hadlich	https://orcid.org/0000-0002-3054-8574
Caroline Lorenci Mallmann	https://orcid.org/0000-0002-2967-1313
Ricardo de Oliveira Perdiz	https://orcid.org/0000-0002-2333-6549
Kaio Cesar Marinho da Cunha	https://orcid.org/0000-0003-2197-9687
Samyra Ramos Chaves	https://orcid.org/0000-0002-8937-1419
Celio Pasquini	https://orcid.org/0000-0002-2445-1354
Kelly Torralvo	https://orcid.org/0000-0001-8089-7652
Michael John Gilbert Hopkins	https://orcid.org/0000-0001-5473-2942
Jochen Schöngart	https://orcid.org/0000-0002-7696-9657
Maria Teresa Fernandez Piedade	https://orcid.org/0000-0002-7320-0498
Priscila Amaral de Sá	https://orcid.org/0009-0000-4555-9658
Maíra Luciana Guimarães Conde	https://orcid.org/0000-0003-1074-3687
Florian Wittmann	https://orcid.org/0000-0001-9180-356X

Email

Flavia Machado Durgante	flaviamdflorestal@gmail.com
Caroline da Cruz Vasconcelos	cc_vasconcelos@hotmail.com
Hilana Louise Hadlich	hilanalouise@gmail.com
Caroline Lorenci Mallmann	carolinemallmann15@gmail.com
Ricardo de Oliveira Perdiz	ricoperdiz@gmail.com
Kaio Cesar Marinho da Cunha	kaiocesar2013.kc19@gmail.com
Samyra Ramos Chaves	samyraramos.1@gmail.com
Celio Pasquini	pasquini@unicamp.br
Kelly Torralvo	kellytorralvo1@gmail.com
Michael John Gilbert Hopkins	mikehopkins44@hotmail.com
Jochen Schöngart	jschongart@gmail.com
Maria Teresa Fernandez Piedade	maua.manaus@gmail.com
Priscila Amaral de Sá	priscilaasa@yahoo.com.br

Maíra Luciana Guimarães Conde conde.lucianagc@gmail.com

Florian Wittmann florian.wittmann@kit.edu

Funding

This study was funded by the German-Brazilian project ATTO (Amazon Tall Tower Observatory) that was supported by German Federal Ministry of Education and Research (BMBF contract 01LK1602F and 01LK2101D); the Brazilian Ministry of Science, Technology, Innovation, and Communication (FINEP/MCTIC, contract 640 01.11.01248.00); PELD/MAUA project (process nº 441811/2020-5 and 446044/2024-5 (CNPq), 01.02.016301.02630/2022-76 (FAPEAM); Call CNPq/MCTI/CONFAP-FAPs/PELD Nº 21/2020); ATTO Hydrotraits project funded by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq; Call nº 01/2022, process nº 440166/2022-5); SpectraPop project (process nº 01.02.016301.04984/2024-17) funded by the Fundação de Amparo à Pesquisa do Estado do Amazonas (FAPEAM; Programa Mulher Faz Ciência Call nº 006/2024); HerbSpectra-Amazônia project funded by CNPq (Pró-Amazônia Call nº 03/2025, process nº 442571/2025-9).

Declaration of generative AI and AI-assisted technologies

Generative AI tools were used to support language editing and to assist in improving R scripts. The authors reviewed, validated, and took full responsibility for all analyses and interpretations presented in this manuscript.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank the collaborators Celso Costa, Teófilo de Souza Valente Junior, Elisson Merquides Fernandes (Gigante), and Evelyn Lopes Ferreira for overall support in the field, as well as the parabotanist José Edmilson da Costa Souza for helping to identify species in the field. We thank the MAUA/INPA research group and ATTO staff for the logistical support. CCV acknowledges the research grant awarded by the International Association for Plant Taxonomy (IAPT, round 2025).

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at <link here>:

Data availability

The dataset used and/or analysed during the current study are available from the corresponding author on reasonable request and will be deposited in the ATTO data portal repository (https://www.attodata.org/home/Start).

Abstract

Species identification in hyperdiverse regions such as the Amazon forest, is time-consuming and often prone to errors. In light of the current rapid biodiversity loss, developing fast and low-cost methods to support species identification in both field and herbarium samples is fundamental to improve the quality of tropical forest inventories. We tested two spectrometers - the high-performance portable ASD FieldSpec® 4 (350–2500 nm; 6–10 kg) and the handheld ISC NIR-S-G1 (900–1700 nm; 150 g), to assess their ability to discriminate tree species from bark and leaf samples. Both instruments achieved accuracy higher than 90% using herbarium leaf samples and 81 to 90% with bark spectra. Although the ISC NIR-S-G1 performance was slightly worse than the ASD, increasing the number of scans per sample improved its accuracy. Demonstrating the effectiveness of this compact, low-cost device for accurate recognition of plant species highlights new opportunities to expand spectral applications for biodiversity monitoring, species identification to forest management and forest conservation specially to countries of the Global South.

Keywords

Amazonian tree flora; low-cost handheld spectrometer; dry leaves; species discrimination; miniaturized NIR spectroscopy; reflectance spectra; tree bark

1. Introduction

Plant spectroscopy has emerged as a powerful non-destructive approach for advancing taxonomy, trait ecology, and biodiversity monitoring. By capturing the chemical and structural signatures of plant tissues, full-range spectroscopy (350–2500 nm) can reveal both phylogenetic (Durgante et al., 2013; Meireles et al., 2020) and functional patterns (Costa et al., 2018, 2023; Robin et al., 2025; Dawson et al., 2025), offering a complementary perspective to traditional morphology and molecular tools. The development of spectral herbaria databases linking spectral and specimen information is transforming biological collections into quantitative, digital resources that can support species identification and trait prediction through machine learning (Cavander-Bares et al., 2025; Neto-Bradley et al., 2025).

Despite its potential, the expansion of spectroscopy in ecology and taxonomy remains constrained by instrument cost and accessibility. High-end spectrometers such as the ASD FieldSpec® series or laboratory-based FT-NIR systems provide high-resolution spectra across a broad wavelength range but are expensive (often up to US$80,000), sensitive to environmental conditions (FT-NIR), and cumbersome for field or herbarium use. These limitations particularly affect institutions in tropical regions, where biodiversity is greatest but financial and logistical resources are limited. To democratize spectral technologies, it is crucial to evaluate the performance of lower-cost, miniaturized devices and determine how far the data they deliver can be harmonized with reference-grade instruments.

Accurate species identification is a persistent challenge in tropical forests, where over 15,000 tree species are estimated to occur in the Amazon alone (ter Steege et al., 2020). Taxonomic identification is hindered by high species richness, morphological similarity among taxa, and declining numbers of expert parabotanists (Hopkins et al., 2007; Carvalho et al., 2023). Cryptic and morphologically variable species frequently lead to misidentifications (Bickford et al., 2007; Gomes et al., 2013), while the reliance on fertile material limits identification during most of the year. These challenges have direct consequences for forest management, where misidentification can lead to inappropriate harvesting decisions, affecting both sustainability and timber evaluation (Lacerda & Nimmo, 2010). Spectroscopy offers a rapid, non-destructive alternative that can be applied to both field-collected and herbarium specimens (Costin et al., 2011; Durgante et al., 2013; Hadlich et al., 2018; Prata et al., 2018; Gaem et al., 2022; Vasconcelos et al., 2025; Hadlich et al., 2025).

Near-infrared (NIR) spectroscopy measures light absorption and reflection to probe molecular-bound vibrations associated with the chemical and physical properties of samples

(Pasquini, 2018). Foundation studies have shown that NIR spectroscopy (1000–2500 nm) can accurately classify species using dry leaves (herbarium samples) (Durgante et al., 2013; Vasconcelos et al., 2021; 2025), recognize species from ontogenetic stages (Lang et al., 2015; Carvalho et al., 2024), and estimate life-history traits from herbarium material (Costa et al., 2018, 2023; Kothari et al., 2023; White et al., 2025; Robin et al., 2025). In the field, hyperspectral devices have been used to discriminate tropical tree species based on leaf or bark spectra (Hadlich et al., 2018, 2025; Mallmann et al., 2023). However, the adoption of these techniques is limited by the size, fragility, and cost of current instruments.

Recent advances in micro-electromechanical systems (MEMS) and miniaturized detectors have produced portable spectrometers that are low-cost, lightweight, and increasingly reliable. Instruments such as the ISC NIR-S-G1 (ISC Corp., Taiwan; hereafter ISC), operate in the 900–1700 nm range, weight < 150 g, and cost less than US$2,000. These devices are already used in the agricultural and pharmaceutical industries (Gullifa et al., 2025; Hu et al., 2025), but their ecological potential remains poorly tested. It is still unclear whether their reduced spectral range and lower signal-to-noise ratio significantly limit their capacity for species discrimination or whether data collected from such devices can be effectively integrated with spectra from high-end instruments.

Here, we evaluated the performance of the portable ISC against the reference-grade ASD FieldSpec® 4 spectroradiometer (350–2500 nm; hereafter ASD) for species identification in tropical trees. We collected spectra from dry leaves of 98 trees (10 species) and from inner and outer bark of 74 standing trees (nine species) in a central Amazon upland (*terra firme*) forest. By comparing performance across tissues, instruments, and preprocessing strategies, we address four key methodological questions: 1. Does the ISC achieve species-level classification accuracy comparable to the ASD FieldSpec® 4? 2. Does the restricted spectral range of the ISC (950–1650 nm) limit prediction performance? 3. How do the number of spectral readings and preprocessing approaches affect model accuracy? 4. Can spectral data collected from different instruments be integrated into a single classification model?

By directly comparing high-end and miniaturized spectrometers under controlled conditions, this study provides an empirical assessment of trade-offs between cost, performance, and interoperability. Demonstrating comparable accuracy between devices would represent a significant step toward scalable, low-cost spectroscopy for herbaria and fieldwork in biodiversity-rich, yet resource-limited, regions. Ultimately, our findings aim to

inform the design of standardized, accessible protocols for the next generation of spectral biodiversity infrastructure.

2. Material and Methods

2.1. Taxon sampling

We selected abundant species representing major Amazonian wood lineages, all taxonomically verified and from distinct families and orders, with ten species for leaves and nine for bark (Table 1). Selection was based on previous floristic inventories in permanent plots at the *terra-firme* forest of the Amazon Tall Tower Observatory (ATTO) site, Uatumã Sustainable Development Reserve (USDR), central Amazonia (Andreae et al., 2015; Fig. S1). Individual trees were marked in the field for on-site spectral data collection, forming the first stage of the study.

Table 1

List of tree species used to compare the two portable spectrometers: ASD FieldSpec® 4 and ISC NIR-S-G1.

Tissue	Family	Scientific name	Label	Number of trees
Dry leaf	Lecythidaceae	*Corythophora alta* R.Knuth	CORALTA	9
	Sapotaceae	*Ecclinusa guianensis* Eyma	ECCGUIA	10
	Sapotaceae	*Manilkara elata* (Allemão ex Miq.) Monach.	MANELAT	9
	Euphorbiaceae	*Micrandropsis scleroxylon* (W.A.Rodrigues) W.A.Rodrigues	MICSCLE	10
	Sapotaceae	*Nemaluma anomala* (Pires) Swenson	NEMANOM	10
	Burseraceae	*Protium apiculatum* Swart	PROAPIC	10
	Violaceae	*Rinorea guianensis* Aubl.	RINGUIA	10
	Humiriaceae	*Sacoglottis guianensis* Benth.	SACGUIA	10
	Malvaceae	*Scleronema grandiflorum* Huber	SCLGRAN	10
	Fabaceae	*Swartzia reticulata* Ducke	SWARETI	10
			Total number of trees	98
Bark	Lecythidaceae	*Corythophora alta* R.Knuth	CORALTA	9
	Sapotaceae	*Ecclinusa guianensis* Eyma	ECCGUIA	8

Sapotaceae	*Manilkara elata* (Allemão ex Miq.) Monach.	MANELAT	9
Euphorbiaceae	*Micrandropsis scleroxylon* (W.A.Rodrigues) W.A.Rodrigues	MICSCLE	9
Sapotaceae	*Nemaluma anomala* (Pires) Swenson	NEMANOM	8
Burseraceae	*Protium apiculatum* Swart	PROAPIC	10
Violaceae	*Rinorea guianensis* Aubl.	RINGUIA	8
Humiriaceae	*Sacoglottis guianensis* Benth.	SACGUIA	6
Fabaceae	*Swartzia reticulata* Ducke	SWARETI	7
		Total number of trees	74

Additionally, botanical samples, including sterile and/or fertile branches, were collected from the same sampling group of individuals and processed using standard herbarium procedures (Mori et al., 2011): they were collected, pressed, dried, and stored on newsprint. These specimens were later used for spectral data acquisition from dry leaves, comprising the second stage of the study.

2.2. Spectra acquisition using two handheld portable spectrometers

An overview of the instruments is provided in Table 2. Each ASD FieldSpec® 4 spectrum (ASD Inc., USA; Fig. 1a, c, d) contains 2,151 reflectance values from 350–2500 nm. The optical resolution of 3 nm in the VIS (350–700 nm) and NIR (700–1000 nm) regions, and 8 nm in the SWIR (1000–2500 nm). The device is operated via the RS^3 software and connects to a laptop using either an Ethernet cable or Bluetooth. Measurements are obtained in contact mode using a contact probe attached via an optical fiber cable, with specific accessories depending on the plant material.

Table 2

Comparison of key technical specifications between the ASD FieldSpec® 4 and the ISC NIR-S-G1 spectrometers.

Specifications	ASD FieldSpec® 4	ISC NIR-S-G1
Manufacturer	Malvern Panalytical	InnoSpectra Corp.
Detector technology	512 element NIR-enhanced silicon array (VIS–NIR), Graded Index InGaAs Photodiode, 2 Stage TE	1 mm InGaAs (uncooled)

	cooled (SWIR)	
Wavelength range	350–2500 nm	900–1700 nm
Wavelength accuracy	± 1 nm	± 1 nm
Spectral regions	VIS–NIR–SWIR1–SWIR2	NIR–SWIR1
Spectral resolution	3 nm (VIS–NIR), 8 nm (SWIR)	~10 nm
Number of variables	2,151	228
Weight	5.44 kg (instrument only) + 0.7 kg (contact probe)	< 0.15 kg
Scanning time	0.1 s	1 s
Average number of scans	10	20

Each spectrum collected with the ISC NIR-S-G1 spectrometer (InnoSpectra Corp., Taiwan) (Fig. 1b, e, f) comprises 228 reflectance values spanning the 900–1700 nm range, with an optical resolution of 10 nm, covering part of the NIR regions. This device is controlled via the ISC-NIRScan application and connects to a laptop or smartphone via USB or Bluetooth.

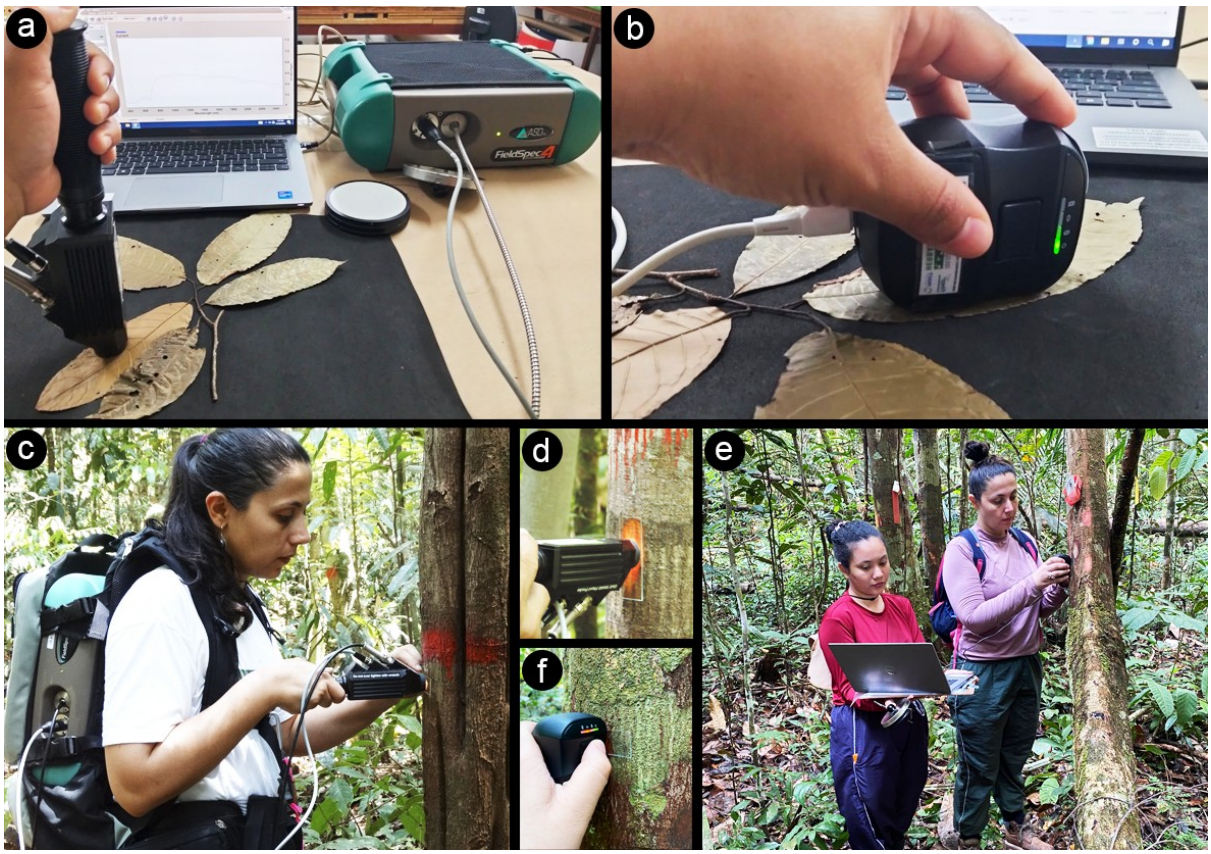


Fig. 1. Spectral data collection in the laboratory for dry leaves (a, b) and in the field for outer and inner bark (c-f) using the two portable spectrometers compared in this study: ASD FieldSpec® 4 (a) and ISC NIR-S-G1 (b). Details of the EVA black background (a, b) used for dry leaves in contact measurements and the glass sheet for bark of standing trees (d, f).

2.2.1. Dry leaves

For each specimen, we collected eight spectra, four from the adaxial surface and four from the abaxial surface, across at least four leaves using both instruments. A black ethylene vinyl acetate (EVA) background was used to reduce light scattering, reflecting <7% in ASD measurements and <3% with the ISC measurements (Fig. S2). Before each specimen, a white reference was taken with a Spectralon® standard to calibrate the instruments.

2.2.2. Bark of standing trees

The number of spectra per tree differed between instruments: three spectra of both outer and inner bark were collected with the ASD FieldSpec® 4 (Hadlich et al., 2018), and five to fourteen spectra per bark tissue with the ISC NIR-S-G1. Outer bark spectra were sampled directly from the trunk, avoiding mosses, lichens, and fungi. Inner bark spectra were obtained by gently removing outer layers to expose internal tissue, without cutting the cambium or exposing wood. Both instruments were calibrated using a Spectralon® white reference, and a glass sheet was placed over the inner bark during measurements to prevent contamination from tree exudates (Fig. 1d, f).

2.3. Spectral data handling and analysis

Spectra with reading errors or atypical profiles were removed based on expert assessment and statistical filtering (high standard deviation), indicating noise or contamination (Fig. S3).

ASD FieldSpec® 4 were trimmed to the range of 400–2400 nm for the and ISC spectra to 950–1650 nm to exclude noisy edges. All spectra were preprocessed using Standard Normal Variate (SNV) transformation using the 'prospectr' package, followed by visual inspection (Fig. S3). This transformation enhances the important signals while minimizing the impact of unwanted artifacts, like the multiplicative light scatter (Barnes et al., 1989).

Species identification was evaluated using Linear Discriminant Analysis (LDA) and Partial Least Squares Discriminant Analysis (PLS-DA) implemented in the 'caret' package, with PLS-DA tuned to 30 latent variables. We tested three input strategies to assess scan averaging: (1) a single random spectrum per individual; (2) the mean of three scans; and (3) the mean of all available scans per individual (eight scans for dry leaves, and four to fourteen scans for bark). Comparative analyses accounted for the distinct spectral ranges of ASD (400–2400 nm) and ISC (950–1650 nm) by evaluating: (i) full spectral ranges; (ii) the overlapping region (950–1650 nm); (iii) raw and preprocessed data; and (iv) most informative wavelengths for each device. Dimensionality reduction used stepwise selection (forward/backward), retaining at most one-third sample size (Williams & Titus, 1988; Durgante et al., 2013).

To compare instruments, we calculated Pearson correlations between ASD and ISC reflectance at matching wavelengths in the overlapping spectral region (950–1650 nm) and conducted Principal Component Analysis (PCA) using mean spectra per species to visualize multiple variate displacement.

Instrument-transferability was tested in three scenarios: (i) ASD-trained models tested on ISC; (ii) ISC-trained models tested on ASD; and (iii) mixed models with 50% ASD and 50% ISC spectra in both training and testing sets (Fig. 2). All transferability tests used only the overlapping region (950–1650 nm), applying ASD wavelengths as the reference set.

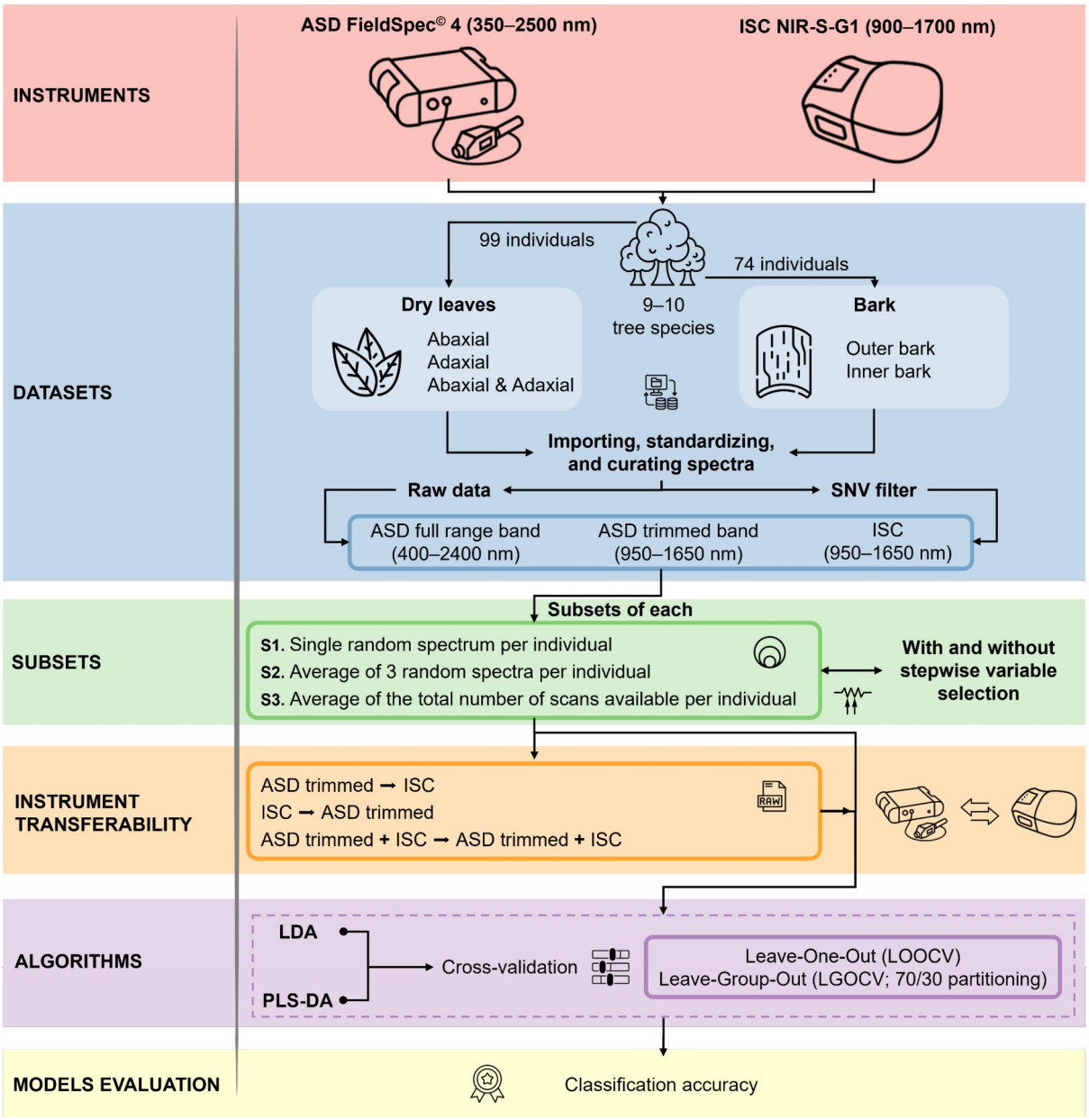


Fig. 2. Workflow for evaluating the performance of two spectrometers (ASD FieldSpec® 4 and ISC) in classifying Amazonian tree species based on spectral data from dry leaves and bark of standing trees.

Model performance was evaluated using two cross-validation techniques: Leave-One-Out Cross-Validation (LOOCV) and Leave-Group-Out-Cross-Validation (LGOCV), with a 70/30 train/test partition repeated 100 times using random splits to ensure robustness. Overall classification accuracy with 95% confidence intervals was used as the primary performance metric (Fig. 2).

All data processing and statistical analyses were performed in R (v4.4.3; R Core Team, 2024), using custom scripts and the packages mentioned above developed for spectral

analysis and classification models. A complete list of R packages used in our analyses is provided in Table S1.

3. Results

3.1. Dry leaf spectra

3.1.1. Results of leaf classification models for each instrument

Leaf spectra from the ASD FieldSpec® 4 achieved consistently high species identification accuracy (> 0.90) across all tested scenarios, including single spectra per individual, averages of three spectra, and averages of eight spectra with raw and processed data (Standard Normal Variate (SNV) transformation) (Table 3; Fig. 3 and Fig. S4). Reducing the spectral range from 350–2500 nm to the shared range between devices (955–1645 nm) caused only a minor decrease in accuracy, particularly for single-spectrum models (0.88).

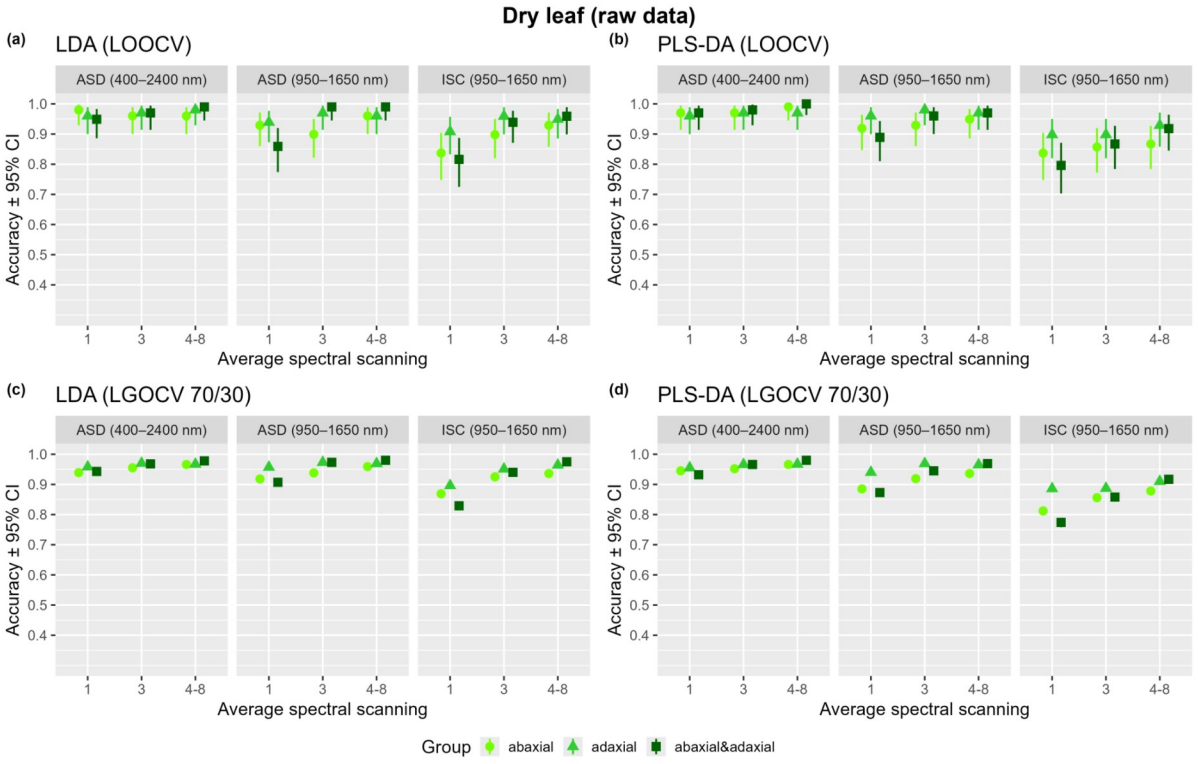


Fig. 3. Classification accuracy based on raw data from the dry leaf spectra. Panels show the results of (a) Linear Discriminant Analysis (LDA) with Leave-One-Out Cross-Validation

(LOOCV), (b) Partial Least Squares Discriminant Analysis (PLS-DA) with LOOCV, (c) LDA with 70/30 Leave-Group-Out Cross-Validation (LGOCV), and (d) PLS-DA with LGOCV. Each panel compares model performance across instruments (ASD full range 400–2400 nm, and trimmed band 950–1650 nm; ISC: 950–1650 nm) and different levels of spectral scan averaging (1, 3, and 4–8 scans). Marker shapes indicate the leaf surface scanned: abaxial (squares), adaxial (triangles), and both surfaces combined (circles).

For the ISC, single-scans models reached 0.80 to 0.81 accuracy, which increased to 0.88 to 0.90 when averaging three scans per individual. Averaging eight spectra per specimen (four adaxial and four abaxial) yielded performance comparable to the ASD, with accuracies of 0.94 to 0.96 (ISC) versus 0.95 to 0.99 (ASD full range) and 0.97 to 0.98 (ASD trimmed) across models and validation methods. The pre-processed data did not increase the accuracy significantly. However, raw or processed by SNV achieved prediction accuracies above 94% for both devices (Table 3). Stepwise variable selection identified the most informative wavelengths in the method located at the region 2000–2500 nm for the full ASD range, and between 1400–1650 nm in the common range for both instruments (Fig. S5).

Table 3

Summary of species classification accuracy for dry leaf spectra (abaxial & adaxial sets only) obtained with ASD FieldSpec® 4 and ISC NIR-S-G1 across all analytical scenarios (see Fig. 2 for workflow). Overall mean accuracy for models based on preprocessing and feature-selection settings. Each subset represents a different level of spectral scan averaging per individual (1, 3, and 4–8 scans). Results summarize the performance of LDA and PLS-DA classification models under both Leave-One-Out (LOOCV) and Leave-Group-Out (LGOCV; 70/30 partitioning) Cross-Validation. Processed data refers to the application of the Standard Normal Variate (SNV) transformation.

Device/ Wavelength	Analysis/ Validation	4–8 scans				3 scans		1 scan	
		Raw data		Preprocessed data		Raw data	Preprocessed data	Raw data	Preprocessed data
		Full range	Selected variables	Full range	Selected variables	Full range	Full range	Full range	Full range
ASD 400–2400 nm	LDA LOO	0.99	0.98	0.99	0.99	0.97	0.99	0.95	0.95
	LDA LGO	0.98	0.97	0.98	0.98	0.97	0.96	0.94	0.93
	PLS-DA LOO	1.00	0.92	1.00	0.95	0.98	0.98	0.97	0.93

	PLS-DA LGO	0.98	0.93	0.98	0.94	0.97	0.97	0.93	0.90
	Mean accuracy	0.99	0.95	0.99	0.97	0.97	0.98	0.95	0.93
ASD 950–1650 nm	LDA LOO	0.99	0.99	0.97	0.99	0.99	0.97	0.86	0.85
	LDA LGO	0.98	0.98	0.98	0.99	0.97	0.97	0.91	0.91
	PLS-DA LOO	0.97	0.97	0.98	0.96	0.96	0.99	0.89	0.93
	PLS-DA LGO	0.97	0.97	0.97	0.95	0.95	0.96	0.87	0.91
	Mean accuracy	0.98	0.98	0.98	0.97	0.97	0.97	0.88	0.90
ISC 950–1650 nm	LDA LOO	0.96	0.98	0.96	0.99	0.94	0.85	0.82	0.80
	LDA LGO	0.98	0.95	0.96	0.98	0.94	0.92	0.83	0.82
	PLS-DA LOO	0.92	0.92	0.94	0.94	0.87	0.87	0.80	0.83
	PLS-DA LGO	0.92	0.92	0.93	0.93	0.86	0.87	0.77	0.80
	Mean accuracy	0.94	0.94	0.95	0.96	0.90	0.88	0.80	0.81

3.1.2. Leaf spectral pattern between instruments

The spectral pattern of dry leaves from 10 species was similar between ASD FieldSpec® 4 and ISC NIR-S-G1 within their overlapping spectral range (950–1650 nm), with both instruments showing a characteristic low reflectance peak around 1400 nm (Fig. S6a). PCA of mean raw leaf spectra revealed tight species level clustering for both devices. Although ISC measurements indicated a slight shift along PC1 relative to ASD, the spectra remained close in ordination space (Fig. S6b). For some species, such as *Sacoglottis guianensis* and *Protium apiculatum*, the displacement along PC1 was minimal, and for most species the separation along PC2 was smaller than along PC1. After spectral pre-processing, the PCA further reduced discrepancies between the two instruments, bringing their spectral signatures even closer. The greatest separation along PC1 in the pre-process data corresponded to the distinction of *Rinorea guianensis* and *Nemaluma anomala* from the other species. But the difference between instruments decreased at both axes. Pearson correlations between instruments using raw data vary among species. Only two species showed consistently high and significant correlations between 1450 and 1550 nm, while the remaining species did not exhibit significant correlations. After pre-processing, correlations increased for most species to approximately 75%, with significance at the 0.05 level (Fig. S7a, d).

339

340 *3.1.3. Leaf classification models with mixed data between instruments*

341

342 Using data from both sources within the same classification models for dried leaves,
343 the datasets were highly comparable. When models were trained with ASD data and species
344 were predicted using ISC, model accuracies ranged from 0.66 to 0.79. Conversely, when
345 models were built with ISC data to predict species using ASD spectra, accuracies varied
346 between 0.86 to 0.92. Additionally, when data from both instruments were combined for
347 training and testing datasets, the accuracies increased, ranging from 0.83 to 0.94 (Table 4).
348 No significant increase in accuracy can be observed with data pre-processing.

349

350 **Table 4**

351 Classification performance of cross-instrument models based on raw and preprocessed dry
352 leaf spectra. Species classification models were trained on spectra from one instrument and
353 tested on the other (ASD → ISC and ISC → ASD), or trained and tested on the combined
354 dataset (ASD + ISC → ASD + ISC). Two classification methods were evaluated: Linear
355 Discriminant Analysis (LDA) and Partial Least Squares Discriminant Analysis (PLS-DA)
356 using Leave-Group-Out Cross-Validation (LGOCV) for internal validation. Values represent
357 overall accuracy with 95% confidence intervals (CI) based on external validation.

Dry leaf models	Analysis/ Validation	Raw data		Preprocessed data	
		Overall accuracy	95% CI	Overall accuracy	95% CI
ASD → ISC	LDA LGOCV	0.66	0.56–0.76	0.30	0.21–0.40
	PLS-DA LGOCV	0.79	0.69–0.86	0.78	0.68–0.85
ISC → ASD	LDA LGOCV	0.92	0.85–0.96	0.91	0.83–0.96
	PLS-DA LGOCV	0.86	0.77–0.92	0.82	0.73–0.89
ASD + ISC → ASD + ISC	LDA LGOCV	0.94	0.87–0.98	0.88	0.80–0.94

	PLS-DA LGOCV	0.83	0.74–0.90	0.78	0.68–0.86

3.2. Bark spectra

3.2.1. Bark classification models

Species classification models based on ASD FieldSpec® bark spectra showed high accuracy, ranging from 0.66 (single scan) to 0.93 (averaging 3 scans) for outer bark and 0.84 (single) to 0.97 (3 scans) for inner bark. Limiting the ASD range to 950–1650 nm produced similar results — outer from 0.69 (single) to 0.95 (3 scans) and inner from 0.89 (single) to 0.99 (3 scans) (Figs. 4 and S8; Tables 5 and 6).

ISC-based models were generally less accurate compared to ASD. However, ISC models performance improved with multiple scans: for outer bark, accuracy increased from 0.48 to 0.58 in a single scan, to 0.65 to 0.77 in 3 scans, and to 0.72 to 0.87 in 4–14 scans, and for inner bark from 0.58 to 0.74 in a single scan to 0.73 to 0.78 in 3 scans and 0.74 to 0.97 in 4–14 scans, respectively. The models developed using pre-processed data showed accuracies similar to those obtained with models built from the raw data (Tables 5 and 6).

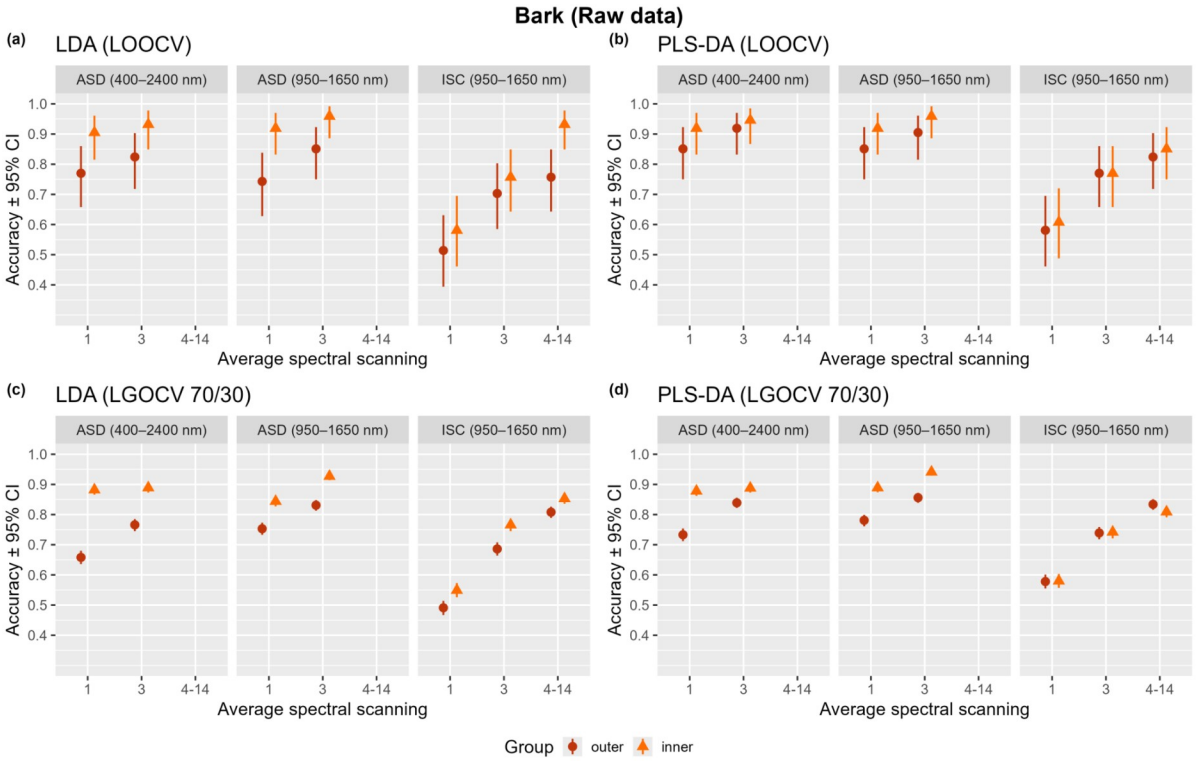


Fig. 4. Classification accuracy based on raw data from the bark of standing trees. Panels show the results of (a) Linear Discriminant Analysis (LDA) with Leave-One-Out Cross-Validation (LOOCV), (b) Partial Least Squares Discriminant Analysis (PLS-DA) with LOOCV, (c) LDA with 70/30 Leave-Group-Out Cross-Validation (LGOCV), and (d) PLS-DA with LGOCV. Each panel compares the performance across instruments (ASD: 400–2450 nm and 950–1650 nm; ISC: 950–1650 nm) and different levels of spectral scan averaging (1, 3, and 4–14 scans). Marker shapes indicate the scanned bark surface: outer (circles) and inner (triangles) bark.

Table 5

Summary of species classification accuracy for outer bark spectra obtained with ASD FieldSpec® 4 and ISC NIR-S-G1 across all analytical scenarios (see Fig. 2 for workflow). Overall mean accuracy for models based on preprocessing and feature-selection settings. Each subset corresponds to a different number of averaged spectral scans per individual (1, 3, and 4–14 scans). Results summarize the performance of LDAand PLS-DA classification models under both Leave-One-Out (LOOCV) and Leave-Group-Out (LGOCV; 70/30 partitioning) Cross-Validation. Processed data refers to the application of the Standard Normal Variate (SNV) transformation.

Device/ Wavelength	Analysis/ Validation	3 scans				1 scan	
		Raw data		Preprocessed data		Raw data	Preprocessed data
		Full range	Selected variables	Full range	Selected variables	Full range	Full range
ASD 400–2400 nm	LDA LOO	0.82	0.93	0.82	0.97	0.77	0.80
	LDA LGO	0.76	0.90	0.77	0.89	0.66	0.77
	PLS-DA LOO	0.91	0.84	0.85	0.91	0.85	0.89
	PLS-DA LGO	0.84	0.85	0.81	0.89	0.73	0.82
	Mean accuracy	0.83	0.88	0.81	0.91	0.75	0.82
ASD 950–1650 nm	LDA LOO	0.82	0.95	0.82	0.91	0.74	0.80
	LDA LGO	0.82	0.88	0.79	0.84	0.75	0.69
	PLS-DA LOO	0.89	0.87	0.89	0.88	0.85	0.80
	PLS-DA LGO	0.85	0.84	0.84	0.84	0.78	0.72
	Mean accuracy	0.85	0.88	0.84	0.86	0.78	0.75
Device/ Wavelength	Analysis/ Validation	4–14 scans				3 scans	1 scan
		Raw data		Preprocessed data		Raw data Preprocessed	Raw data Preprocessed

							data		data
		Full range	Selected variables	Full range	Selected variables	Full range	Full range	Full range	Full range
ISC 950–1650 nm	LDA LOO	0.76	0.82	0.77	0.87	0.70	0.72	0.51	0.51
	LDA LGO	0.81	0.77	0.81	0.79	0.69	0.65	0.49	0.51
	PLS-DA LOO	0.82	0.81	0.82	0.74	0.77	0.74	0.58	0.49
	PLS-DA LGO	0.83	0.78	0.79	0.72	0.74	0.70	0.58	0.48
	Mean accuracy	0.81	0.80	0.80	0.78	0.72	0.70	0.54	0.50

394

Table 6

Summary of species classification accuracy for inner bark spectra obtained with ASD FieldSpec® 4 and ISC NIR-S-G1 across all analytical scenarios (see Fig. 2 for workflow). Overall mean accuracy for models based on preprocessing and feature-selection settings. Each subset corresponds to a different number of averaged spectral scans per individual (1, 3, and 4–14 scans). Results summarize the performance of LDA and PLS-DA classification models under both Leave-One-Out (LOOCV) and Leave-Group-Out (LGOCV; 70/30 partitioning) Cross-Validation. Processed data refers to the application of the Standard Normal Variate (SNV) transformation.

Device/ Wavelength	Analysis/ Validation	3 scans				1 scan	
		Raw data		Preprocessed data		Raw data	Preprocessed data
		Full range	Selected variables	Full range	Selected variables	Full range	Full range
ASD 400–2400 nm	LDA LOO	0.93	0.97	0.95	0.96	0.91	0.95
	LDA LGO	0.89	0.95	0.93	0.93	0.88	0.89
	PLS-DA LOO	0.95	0.91	0.95	0.91	0.92	0.88
	PLS-DA LGO	0.89	0.87	0.90	0.87	0.88	0.84
	Mean accuracy	0.91	0.92	0.93	0.92	0.90	0.89
ASD 950–1650 nm	LDA LOO	0.96	0.99	0.96	0.96	0.92	0.89
	LDA LGO	0.93	0.95	0.94	0.89	0.84	0.89
	PLS-DA LOO	0.96	0.95	0.95	0.96	0.92	0.93
	PLS-DA LGO	0.94	0.93	0.93	0.91	0.89	0.90
	Mean accuracy	0.95	0.95	0.94	0.93	0.89	0.90

Device/ Wavelength	Analysis/ Validation	4–14 scans				3 scans		1 scan	
		Raw data		Preprocessed data		Raw data	Preprocessed data	Raw data	Preprocessed data
		Full range	Selected variables	Full range	Selected variables	Full range	Full range	Full range	Full range
ISC 950–1650 nm	LDA LOO	0.93	0.92	0.88	0.97	0.76	0.73	0.58	0.61
	LDA LGO	0.85	0.83	0.82	0.90	0.77	0.74	0.55	0.61
	PLS-DA LOO	0.85	0.80	0.78	0.88	0.77	0.78	0.61	0.74
	PLS-DA LGO	0.81	0.74	0.77	0.83	0.74	0.74	0.58	0.67
	Mean accuracy	0.86	0.82	0.81	0.90	0.76	0.75	0.58	0.66

405 *3.2.2. Bark spectral pattern between instruments*

406

407 Outer bark (Fig. S9a) and the inner bark spectra (Fig. S10a) showed similar overall
408 patterns to raw data between ASD and ISC, including the characteristic water absorption peak
409 at 1450 nm. Inner bark additionally displayed low reflectance peaks at 900 and 1250 nm.
410 With the pre-processed data, it is possible to recognize greater discrepancies in the spectral
411 patterns when comparing the two instruments. The PCA results differed between the raw and
412 pre-processed datasets (Figs. S9b, S10b). In the raw data, for the outer bark, spectra from the
413 same species were more dispersed along the first principal component (PC1) but closer along
414 PC2. For the inner bark, however, spectra from the same species showed dispersion along
415 both axes. In contrast, the pre-processed data revealed shorter distances among objects of the
416 same species in the PCA, indicating spectral coherence after data treatment. Despite these
417 similarities, Pearson correlation within the shared spectral range was low. Using raw data,
418 correlations with the outer bark were significant in only two species, and no significant
419 correlations were found for the inner bark in any species (Fig. S7b-c). After data pre-
420 processing, correlations increased and became significant for five species in both bark
421 tissues; however, this improvement was restricted to specific wavelength regions (Fig. S7e-f).
422 A higher variance (shade area) was observed in bark spectra outer (Fig. S11) and
423 inner (Fig. S12) when acquired with the ISC compared to the ASD FieldSpec®. Furthermore,
424 the most informative spectral regions selected for each model —full-range spectra (ASD),
425 950–1650 nm (ASD), and 950–1650 nm (ISC)—were not consistent across discriminant
426 models for each tissue type.

427

428 *3.2.3. Bark classification models with mixed data between instruments*

429

Using data from both sources within the same classification models for bark tissues significantly reduced model accuracy for raw and pre-processed data. For raw data, the outer bark, where ASD-collected data were used to predict species from ISC data, the accuracy varied from 0.35 to 0.47. When ISC collected data to predict species from ASD data, the accuracy was close to 0.50. When the two datasets were combined in a single model, the accuracy was between 0.46 to 0.76. For inner bark, the results were even poorer: models trained with ASD-collected data to predict species from ISC data or vice versa, achieved the accuracy of only 0.14 to 0.27. However, when data from both sensors were merged into a single model, the accuracy improved to 0.65 to 0.74 (Table 7).

Table 7

Classification performance of cross-instrument models based on raw and preprocessed bark spectra. Species classification models were trained on spectra from one instrument and tested on the other (ASD → ISC and ISC → ASD), or trained and tested on the combined dataset (ASD + ISC → ASD + ISC). Two classification methods were evaluated: Linear Discriminant Analysis (LDA) and Partial Least Squares Discriminant Analysis (PLS-DA) using Leave-Group-Out Cross-Validation (LGOCV) for internal validation. Values represent overall accuracy with 95% confidence intervals (CI) based on external validation. Processed data refers to the application of the Standard Normal Variate (SNV) transformation.

Bark tissue	Model	Analysis/ Validation	Raw data		Preprocessed data (SNV)	
			Overall Accuracy	95% CI	Overall Accuracy	95% CI
Outer	ASD → ISC	LDA LGOCV	0.47	0.36–0.59	0.43	0.32–0.55
		PLS-DA LGOCV	0.35	0.24–0.47	0.35	0.24–0.47
	ISC → ASD	LDA LGOCV	0.54	0.42–0.66	0.58	0.46–0.69
		PLS-DA LGOCV	0.51	0.39–0.63	0.48	0.37–0.61
	ASD + ISC → ASD + ISC	LDA LGOCV	0.76	0.64–0.85	0.70	0.59–0.80

		PLS-DA LGOCV	0.46	0.34–0.58	0.42	0.31–0.54
Inner	ASD → ISC	LDA LGOCV	0.18	0.10–0.28	0.12	0.06–0.22
		PLS-DA LGOCV	0.27	0.17–0.39	0.20	0.12–0.31
	ISC → ASD	LDA LGOCV	0.19	0.11–0.30	0.11	0.05–0.20
		PLS-DA LGOCV	0.14	0.07–0.23	0.27	0.17–0.39
	ASD + ISC → ASD + ISC	LDA LGOCV	0.74	0.63–0.84	0.80	0.69–0.88
		PLS-DA LGOCV	0.65	0.53–0.76	0.59	0.47–0.71

4. Discussion

Portable, low-cost spectrometers such as the ISC can provide reliable species-level identification for herbarium leaves and bark samples, approaching the performance of high-end devices when multiple scans per specimen are averaged. Differences in predictive accuracy are driven mainly by sensor sensitivity and signal stability rather than spectral range, indicating that compact instruments can capture the chemical features most relevant for classification. Despite the challenges of integrating data across different spectrometers for bark spectra, dried herbarium leaves showed promising cross-instrument transferability. Overall, our results reinforce the potential to expand the use of spectroscopy in herbaria and forest inventories. Portable NIR spectroscopy, therefore, represents a practical and scalable pathway to democratize species identification, modernize biological collections, and strengthen global biodiversity monitoring infrastructures.

4.1. Spectral model performance

4.1.1. Pressed dry leaves: an application niche for ISC NIR-S-G1

The ISC demonstrated strong performance in species identification using herbarium leaf samples. Models built with ASD spectra achieved >90% accuracy across validation schemes, while single-scan ISC models achieved around 80%. Averaging multiple scans per specimen markedly improved performance (ASD: 97%; ISC: 95%), confirming that replication stabilizes spectral quality (Durgante et al., 2013). Eight scans per specimen (four adaxial + four abaxial) proved optimal, and replication strategies ≥4 scans are therefore recommended for herbarium applications (see Dawson et al., 2025 and Cavander-Bares et al., 2025).

Despite its lower resolution and sensitivity, the ISC achieved comparable results to the ASD while costing 88% less, highlighting its potential for large-scale spectral digitization of herbarium collections. Handheld devices, however, are more sensitive to operator movement than benchtop spectrometers (Beć et al., 2020) and require proper training in both spectral acquisition and data curation. Detecting and re-measuring atypical or noisy from the expected pattern is essential to maintain data reliability and model consistency.

Spectral models integrating data from both instruments maintained >80% accuracy, demonstrating robustness and cross- instrument compatibility for leaf spectra. Although preprocessing increased spectral correlation and improved alignment in PCA space, it did not translate into higher classification accuracy. Calibration consistency was ensured by using Spectralon® 100% reference standard and identical black EVA backgrounds, minimizing reflectance contamination (<3% for ISC and <7% for ASD; Fig. S2). Because herbaria employ diverse background materials (Neto-Bradley et al., 2025), future studies should test how performance type influences spectral readings to enhance long-term data comparability.

4.1.2. Bark Spectra: Performance under field conditions

Bark spectral posed greater challenges for classification, consistent with earlier ASD-based studies that reported higher accuracy for inner than outer bark (Hadlich et al., 2018; 2025). Our results corroborate with these patterns: ASD outer bark accuracy increased from 0.75 to 0.78 (single scan) to 0.84 to 0.88 (averaged scans), while inner bark improved from 0.89 to 0.90 to 0.91 to 0.95.

ISC spectra exhibited higher variability, resulting in lower single scans accuracies (outer bark: 0.50 to 0.54; inner bark: 0.58 to 0.66). Nonetheless, averaging multiple scans substantially improved performance (outer: 0.78 to 0.81; inner 0.81 to 0.90), reducing the gap

with ASD to only 3–9%. These results emphasize that multi-scan averaging is essential for the reliable classification of bark tissues in field conditions.

Pre-processing with Standard Normal Variate (SNV) increased the correlation between instruments from <25% in raw data to ~50% in some bark spectral regions and reduced inter-intrument distances in PCA space. However, model accuracy remained unchanged, indicating that noise reduction improved comparability without altering predictive power, especially to transferability of models (outer: ~40%; inner: 13%). This highlights the need for methodological standardization when the goal is to integrate bark spectra from multiple instruments.

Despite the limitations in cross-instrument transferability, the ISC demonstrated strong predictive power for identifying species in the field at a fraction of the cost of high-end spectrometers. This strengthens its potential use in forest inventories, sustainable forest management and timber sector, where reliable species identification is essential for regulatory compliance and market transparency. Field spectroscopy can flag individuals with potential misidentifications for subsequent verification and enables the calculation of a botanical identification quality index, which could become a valuable metric for improving management plans and verifying the legitimacy of traded and certified timber.

4.2. Instrument sensitivity and spectral quality

ISC data quality depended strongly on preprocessing. Many spectra exhibited irregularities that required filtering and averaging (>3 scans for leaves, >5 for bark). These issues likely stem from lower spectral resolution and detector sensitivity. Nevertheless, consistent and accurate models were obtained under controlled conditions. Successful use therefore requires operator training to standardize sample procedures, minimize artefacts, and ensure reproducibility, key aspects for establishing compact spectroscopy as a reliable tool in biodiversity studies.

4.3. Comparative evaluation: ISC vs. ASD

When averaged spectra were used, both instruments performed similarly for leaves and bark, although the ASD consistently outperformed the ISC. Accuracy differences were not primarily driven by the narrower spectral range (950–1650 nm vs 400–2400 nm), as ASD full-range and trimmed models showed only minor variation (Tables 3, 5, and 6). Instead,

discrepancies reflect differences in detector performance, resolution, and overall signal quality.

The ISC's main advantages are its low cost and portability, it is approximately 88% cheaper and far more compact, making it particularly valuable for remote regions, mobile collection campaigns, and resource-limited institutions, especially in the Global South. However, cross-instrument equivalence cannot be assumed, as differences in detector types, calibration, resolution, and preprocessing may affect outcomes. Validation studies like this one are crucial to ensure interoperability.

Despite these differences, the two instruments identified similar informative spectral regions (notably 1400–1700 nm for leaves), indicating that the ISC captures the key chemical signatures relevant for these species discrimination. Thus, the remaining accuracy gap likely reflects higher noise and lower sensitivity rather than missing spectral information from different regions. With appropriate calibration and standardized acquisition protocols, compact spectrometers offer a promising and scalable solution for expanding operational biodiversity monitoring.

4.4. Toward an Interoperable Spectral Herbarium Network

The ISC demonstrates strong potential to modernize herbarium workflows, particularly in collections where dry leaves constitute the dominant tissue type. Its affordability and portability can accelerate the development of regional spectral libraries, enabling cross-referencing of species and improving identification and quality control.

To ensure data interoperability, a hierarchical system is needed in which high-end spectrometers provide calibration references for portable devices, ensuring their predictions remain accurate and comparable across instruments. Inter-laboratory collaboration and shared databases will be essential to prevent data silos. Building on this concept, spectral herbarium networks should operate at international and regional levels. Global initiatives such as the Next-Generation Specimen Digitization consortium (Cavander-Bares et al., 2025) are defining standards for spectral acquisition and metadata curation, while regional herbaria can provide calibration spectra and training hubs, generating reference spectra, developing transfer models, and training technicians. Such collaborations promote capacity building and data harmonization among institutions that share floras but have unequal infrastructure.

Demonstrating that a small, more accessible spectrometer can enable species recognition from the bark spectra of standing trees represents a significant step toward

transforming species identification in sustainable forest management. In regions such as the Amazon, where misidentification rates can exceed 40% of inventoried trees, there is an urgent need to operationalize this technology within forest management systems. Integrating spectroscopy into routine forest inventories could support logging companies in improving species identification, reducing the risk of overharvesting rare species, and ensuring accurate timber classification and valuation. The application of spectroscopy in field inventories also opens the door to the development of robust quality assurance and quality control protocols for species identification.

Together, these networks can establish a globally consistent yet locally grounded spectral infrastructure, transforming herbaria from passive specimen repositories into active nodes in digital biodiversity monitoring. Similarly, by reducing errors in forest inventories, improving regulatory enforcement systems, and increasing both reliability and financial returns, this approach could evolve into a powerful biodiversity monitoring system through the construction of large-scale spectral databases of bark traits. If adopted not only by private companies but also by regulatory agencies, such field-based tools could substantially enhance the accuracy, transparency, and scalability of biodiversity assessments. Despite current limitations, compact devices like the ISC represent an important step toward scalable and rapid biodiversity assessment in tropical ecosystems.

4.5. Limitations and future directions

Three main limitations emerge: (1) broader taxonomic (including plants with varying leaf anatomy) and geographic sampling is required to test generality; (2) cross-instrument transfer remains promising but requires systematic calibration-transfer experiments. Future work should also evaluate ISC for functional trait prediction and dry wood identification, expanding its applications in forest monitoring and traceability. These efforts will contribute to democratizing spectroscopy in biodiversity science through reproducible, low-cost protocols applicable in both herbarium and field contexts.

In addition to addressing these limitations, ISC-based methods offer several important gains. First, they enable the evaluation of preliminary identifications, allowing the recognition of errors, reassessment of specimens, and improved consistency in taxonomic assignments. Second, they can help detect potentially new species or specimens that differ from those represented in existing reference databases. Third, with the development of large spectral reference libraries, ISC can facilitate rapid pre-identification of specimens based on their spectral signatures.

Importantly, these advances also have direct implications for forest management. By improving species identification in field inventories, spectroscopy can reduce misidentification rates, support more accurate timber classification, and enhance traceability along the supply chain. This, in turn, can contribute to better decision-making in harvesting operations, reduce the risk of overexploiting rare or protected species, and increase the economic value and reliability of forest products.

Together, these benefits underscore the potential of spectroscopy to improve accuracy, efficiency, and discovery in biodiversity research while strengthening sustainable forest management practices and complementing traditional morphological approaches.

References

Andreae, M. O., Acevedo, O. C., Araújo, A., Artaxo, P., Barbosa, C. G. G., Barbosa, H. M. J., Brito, J., Carbone, S., Chi, X., Cintra, B. B. L., da Silva, N. F., Dias, N. L., Dias-Júnior, C. Q., Ditas, F., Ditz, R., Godoi, A. F. L., Godoi, R. H. M., Heimann, M., Hoffmann, T., Kesselmeier, J., Könemann, T., Krüger, M. L., Lavric, J. V., Manzi, A. O., Lopes, A. P., Martins, D. L., Mikhailov, E. F., Moran-Zuloaga, D., Nelson, B. W., Nölscher, A. C., Santos Nogueira, D., Piedade, M. T. F., Pöhlker, C., Pöschl, U., Quesada, C. A., Rizzo, L. V., Ro, C.-U., Ruckteschler, N., Sá, L. D. A., de Oliveira Sá, M., Sales, C. B., dos Santos, R. M. N., Saturno, J., Schöngart, J., Sörgel, M., de Souza, C. M., de Souza, R. A. F., Su, H., Targhetta, N., Tóta, J., Trebs, I., Trumbore, S., van Eijck, A., Walter, D., Wang, Z., Weber, B., Williams, J., Winderlich, J., Wittmann, F., Wolff, S., and Yáñez-Serrano, A. M. (2015). The Amazon Tall Tower Observatory (ATTO): overview of pilot measurements on ecosystem ecology, meteorology, trace gases, and aerosols. *Atmospheric Chemistry and Physics*, 15(18), 10723–10776. https://doi.org/10.5194/acp-15-10723-2015

Beć, K. W, Grabska, J., Siesler, H. W., Huck, C. W. (2025). Handheld near-infrared spectrometers: Where are we heading? *NIR news*, 31(3-4), 28–35. https://doi.org/10.1177/0960336020916815

Barnes R. J., Dhanoa M. S., & Lister S. J. (1898). Standard Normal Variate Transformation and De-trending of Near-Infrared Diffuse Reflectance Spectra. *Applied Spectroscopy*, 43(5), 772–777. https://doi.org/10.1366/0003702894202201

Bickford, D., Lohman, D. J., Sodhi, N. S., Ng, P. K. L., Meier, R., Winker, K., Ingram, K. K., & Das, I. (2007). Cryptic species as a window on diversity and conservation. *Trends in Ecology & Evolution*, 22(3), 148–155. https://doi.org/10.1016/j.tree.2006.11.004

Carvalho, R. L., Resende, A. F., Barlow, J., França, F. M., Moura, M. R., Maciel, R., Alves-Martins, F., Shutt, J., Nunes, C. A., Elias, F., Silveira, J. M., Stegmann, L., Baccaro, F. B., Juen, L., Schietti, J., Aragão, L., Berenguer, E., Castello, L., Costa, F. R. C., Guedes, M. L., Leal, C. G., Lees, A. C., Isaac, V., Nascimento, R. O., Phillips, O. L., Schmidt, F. A. Steege, H., Vaz-de-Mello, F., Venticinque, E. M., Vieira, I. C. G., Zuanon, J., The Synergize Consortium, Ferreira, J. (2023). Pervasive gaps in Amazonian ecological research. *Current Biology*, 33(16), 3495–3504. https://doi.org/10.1016/j.cub.2023.06.077

Carvalho, T. C. R., Vasconcelos, C. C., & Terra-Araujo, M. H. (2024). Testing the morphotype-based classification of two leaf shapes in *Ecclinusa guianensis* (Sapotaceae) from Central Amazonia. *Brittonia*, 76, 398–410. https://doi.org/10.1007/s12228-024-09809-8

Cavender-Bares, J., White, D. M., Ahlstrand, N. I., Austin, M. W., Bastianelli, D., Bazan, S., Boughalmi, K., Cardinal-McTeague, W., Chacón-Madrigal, E., Couvreur, T. L. P., Davis, C., Durgante, F. M., Grace, O. M., Guzmán Q., J. A., Hansen, K., Hernández-Leal, M. S., Hopkins, M. J. G., Jackson, R., Kothari, S., Lee, A. K., Léveillé-Bourret, É., Pinto-Ledezma, J., Quinteros-Casaverde, N. L., Meireles, J. E., Neto-Bradley, B., Nichodemus, C. O., Ree, R. H., Schmull, M., Soltis, D. E., Soltis, P.S., Tuomisto, H., Ustin, S. L., &Vasconcelos, C.C. (2025). Next-generation specimen digitization: capturing reflectance spectra from the world's herbaria for modeling plant biology across time, space, and taxa. *New Phytologist*. https://doi.org/10.1111/nph.70645

Costa, F. R. C., Lang, C., Almeida, D. R. A., Castilho, C., & Poorter, L. (2018). Near-infrared spectrometry allows fast and extensive predictions of functional traits from dry leaves and branches. *Ecological Applications*, 28(5), 1157–1167. https://doi.org/10.1002/eap.1728

Costa, F. R. C., Lang, C., Sousa, T. R., Emilio, T., Esquivel-Muelbert, A., & Schietti, J. (2023). Fine-grained water availability drives divergent trait selection in Amazonian trees. *Frontiers in Forests and Global Change*, 6, 1112560. https://doi.org/10.3389/ffgc.2023.1112560

Costion, C., Ford, A., Cross, H., Crayn, D., Harrington, M., & Lowe, A. (2011). Plant DNA barcodes can accurately estimate species richness in poorly known floras. *PLoS ONE*, *6*(11), e26841. https://doi.org/10.1371/journal.pone.0026841

Durgante, F. M., Higuchi, N., Almeida, A., & Vicentini, A. (2013). Species spectral signature: Discriminating closely related plant species in the Amazon with Near-Infrared Leaf-Spectroscopy. *Forest Ecology and Management*, *291*, 240–248. https://doi.org/10.1016/j.foreco.2012.10.045

Gaem, P. H., Andrade, A., Mazine, F. F., & Vicentini, A. (2022). Tree species delimitation in tropical forest inventories: Perspectives from a taxonomically challenging case study. *Forest Ecology and Management*, *505*, 119900. https://doi.org/10.1016/j.foreco.2021.119900

Gomes, A. C. S., Andrade, A. C. S., Barreto-Silva, J. S., Brenes-Arguedas, T., López, D. C., Freitas, C. C., Lang, C., Oliveira, A. A., Pérez, A. J., Perez, R., Silva, J. B., Silveira, A. M. F., Vaz, M. C., Vendrami, J., & Vicentini, A. (2013). Local plant species delimitation in a highly diverse Amazonian forest: Do we all see the same species? *Journal of Vegetation Science*, 24(1), 70–79. https://doi.org/10.1111/j.1654-1103.2012.01441.x

Gullifa, G., Albertini, C., Amoresano, A., Pinto, G., Illiano, A., Dirito, P., Materazzi, S., & Risoluti, R. (2025). A smart-based screening system by MicroNIR and chemometrics for on-site authentication of buffalo milk in the dairy industry. *Applied Food Research,* 5(2), 101159. https://doi.org/10.1016/j.afres.2025.101159

Hadlich, H. L., Durgante, F. M., Santos, J. dos, Higuchi, N., Chambers, J. Q., & Vicentini, A. (2018). Recognizing Amazonian tree species in the field using bark tissues spectra. *Forest Ecology and Management*, 427, 296–304. https://doi.org/10.1016/j.foreco.2018.06.002

Hadlich, H. L., Schöngart, J., Wittmann, F., Vasconcelos, C. C., Mallmann, C. L., Conde, M. L. G., Sá, P. A., Demarchi, L. O., Mori, G. B., Piedade, M. T. F., & Durgante, F. M. (2025). Exploring the potential of field spectroscopy for tree species identification in different Amazonian forest ecosystems. *Global Ecology and Conservation,* e03970. https://doi.org/10.1016/j.gecco.2025.e03970

Hopkins, M. J. G. (2007). Modelling the known and unknown plant biodiversity of the Amazon Basin. *Journal of Biogeography*, 8(34), 1400–1411. https://doi.org/10.1111/j.1365-2699.2007.01737.x

Hu, Q., Feng, X., Wang, L., Zhang, L., Chen, X., Wang, Y., Xie, G., & Peng, Q. (2025). Rapid discrimination of Chinese rice wine (Huangjiu) from various regions using benchtop-NIR and micro-NIR spectrometers in conjunction with chemometrics. *Infrared Physics & Technology,* 149, 105904. https://doi.org/10.1016/j.infrared.2025.105904

Kothari, S., Beauchamp-Rioux, R., Blanchard, F., Crofts, A. L., Girard, A., Guilbeault-Mayers, X., Hacker, P. W., Pardo, J., Schweiger, A. K., Demers-Thibeault, S., Bruneau, A., Coops, N. C., Kalacska, M., Vellend, M., & Laliberté, E. (2023). Predicting leaf traits across functional groups using reflectance spectroscopy. *New Phytologist*, 238(2), 549–566. https://doi.org/10.1111/nph.18713

Lacerda, A. E. B., & Nimmo, E. R. (2010). Can we really manage tropical forests without knowing the species within? Getting back to the basics of forest management through taxonomy. *Forest Ecology and Management, 259*(5), 995–1002. https://doi.org/10.1016/j.foreco.2009.12.005

Lang, C., Costa, F. R. C., Camargo, J. L. C., Durgante, F. M., & Vicentini, A. (2015). Near infrared spectroscopy facilitates rapid identification of both young and mature Amazonian tree species. *PLoS ONE*, 10(8), e0134521. https://doi.org/10.1371/journal.pone.0134521

Mallmann, C. L., Pereira-Filho, W., Dreyer, J. B. B., Tabaldi, L. A., & Durgante, F. M. (2023). Leaf-level field spectroscopy to discriminate invasive species (*Psidium guajava* L. and *Hovenia dulcis* Thunb.) from native tree species in the southern Brazilian Atlantic Forest. *Remote Sensing*, 15(3), 791. https://doi.org/10.3390/rs15030791

Meireles, J. E., Cavender-Bares, J., Townsend, P. A., Ustin, S., Gamon, J. A., Schweiger, A. K., Schaepman, M. E., Asner, G. P., Martin, R. E., Singh, A., Schrodt, F., Chlus, A., & O'Meara, B. C. (2020). Leaf reflectance spectra capture the evolutionary history of seed plants. *New Phytologist*, 228(2), 485–493. https://doi.org/https://doi.org/10.1111/nph.16771

Mori, S. A., Berkov, A., Gracie, C. A., & Hecklau, E. F. (2011). Tropical Plant Collecting: From the Field to the Internet. TECC.

731 Neto-Bradley, B.M., Bonnet, P., Goëau, H., Joly, A., Cavender-Bares, J. & Coomes, D.A.
732 (2025). Using reflectance spectra and Pl@ntNet to identify herbarium specimens: a
733 case study with *Lithocarpus*. *New Phytologist*. https://doi.org/10.1111/nph.70258
734 Pasquini, C. (2018). Near infrared spectroscopy: A mature analytical technique with new
735 perspectives – A review. *Analytica Chimica Acta*, 1226, 8–36.
736 https://doi.org/10.1016/j.aca.2018.04.004
737 Prata, E. M. B., Sass, C., Rodrigues, D. P., Domingos, F. M. C. B., Specht, C. D., Damasco,
738 G., Fine, P. V. A., Ribas, C. C., & Vicentini, A. (2018). Towards integrative
739 taxonomy in Neotropical botany: Disentangling the *Pagamea guianensis* species
740 complex (Rubiaceae). *Botanical Journal of the Linnean Society*, 188(2), 213–231.
741 https://doi.org/10.1093/botlinnean/boy051
742 R Core Team. (2024). *R: A Language and Environment for Statistical Computing* (4.4.2). R
743 Foundation for Statistical Computing. https://www.r-project.org/
744 Robin, M., Durgante, F. M., Mallmann, C. L., Hadlich, H. L., Römermann, C., Falcão, L. C.,
745 Lacerda, C. D., Duvoisin Jr., S., Wittmann, F., Piedade, M. T. F., Schöngart, J. &
746 Alves, E. (2025). Leaf spectroscopy as a tool for predicting the presence of isoprene
747 emissions and terpene storage in central Amazon forest trees. *Plant Methods*, 21(78),
748 213–231. https://doi.org/10.1186/s13007-025-01400-w
749 ter Steege, H., Prado, P. I., Lima, R. A. F., Pos, E., Souza Coelho, L., Andrade Lima Filho,
750 D., Salomão, R. P., Amaral, I. L., de Almeida Matos, F. D., Castilho, C. V, Phillips,
751 O. L., Guevara, J. E., de Jesus Veiga Carim, M., Cárdenas López, D., Magnusson, W.
752 E., Wittmann, F., Martins, M. P., Sabatier, D., Irume, M. V., … Pickavance, G.
753 (2020). Biased-corrected richness estimates for the Amazonian tree flora. *Scientific*
754 *Reports*, 10(1), 10130. https://doi.org/10.1038/s41598-020-66686-3
755 Vasconcelos, C. C., Ferraz, I. D. K., Durgante, F. M., Wittmann, F., Schöngart, J., Piedade,
756 M. T. F., Camargo, J. L. C., & Terra-Araujo, M. H. (2025). *Ecclinusa nervosa*
757 (Sapotaceae, Chrysophylloideae), a new species discovered in Central Amazonia.
758 *Systematic Botany*, 50(1), 55–66.
759 https://doi.org/10.1600/036364425X17466502618821
760 Vasconcelos, C., Ferraz, I., Adrianzén, M. U., Camargo, J., & Terra-Araujo, M. (2021).
761 *Chromolucuma brevipedicellata* (Sapotaceae, Chrysophylloideae), a new tree species
762 from central Amazonia, Brazil. *Brittonia*, *73*(2), 211–219.
763 https://doi.org/10.1007/s12228-021-09663-y

White, D. M., Cavender-Bares, J., Davis, C. C., Guzmán, J. A., Kothari, S., Robles, J. M., & Meireles, J. E. (2025). Seeing herbaria in a new light: Leaf reflectance spectroscopy unlocks trait and classification modeling in plant biodiversity collections. *New Phytologist*. https://doi.org/10.1111/nph.70357

Williams, B. K., & Titus, K. (1988). Assessment of sampling stability in ecological applications of discriminant analysis. *Ecology*, 69(4), 1275–1285. https://doi.org/10.2307/1941283

Appendix A. Supporting information

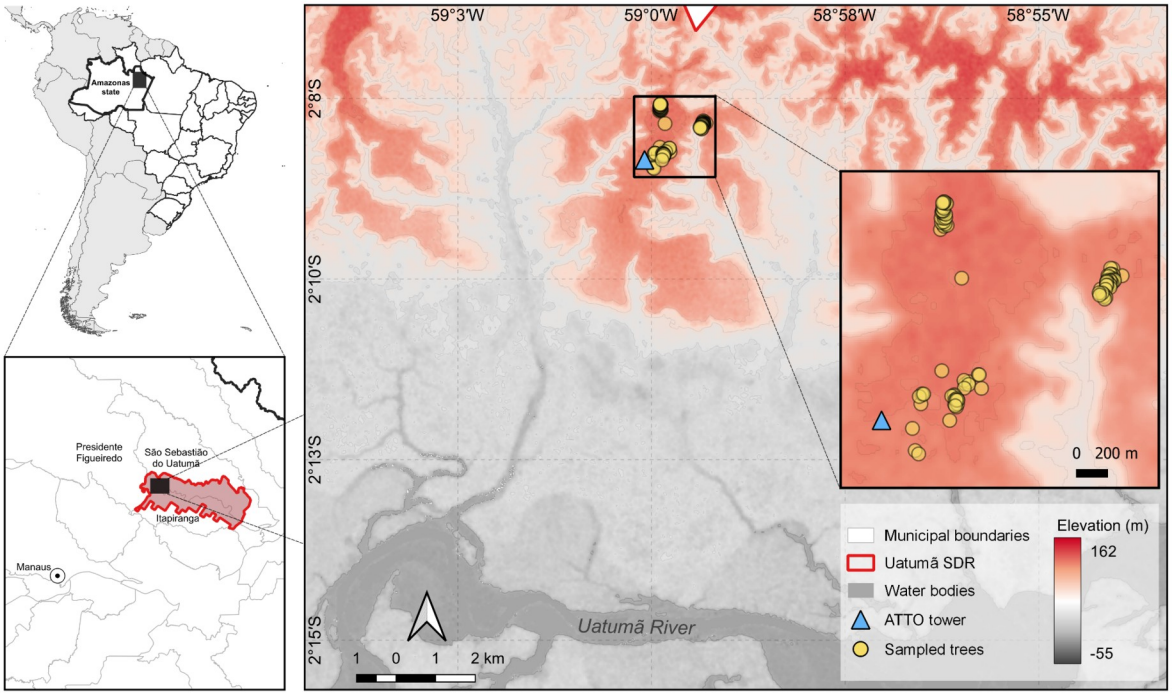


Fig. S1. Location of the Amazon Tall Tower Observatory (ATTO) in the Uatumã Sustainable Development Reserve (USDR), central Amazonia. The main panel depicts the spatial distribution of the sampled trees (yellow circles) around the ATTO tower (blue triangle). Insets indicate the regional context within South America and a zoomed view of the sampling area. Municipal boundaries, water bodies, and the USDR limits are also shown. The map was produced using QGIS (v3.28.1; QGIS Development Team 2022).

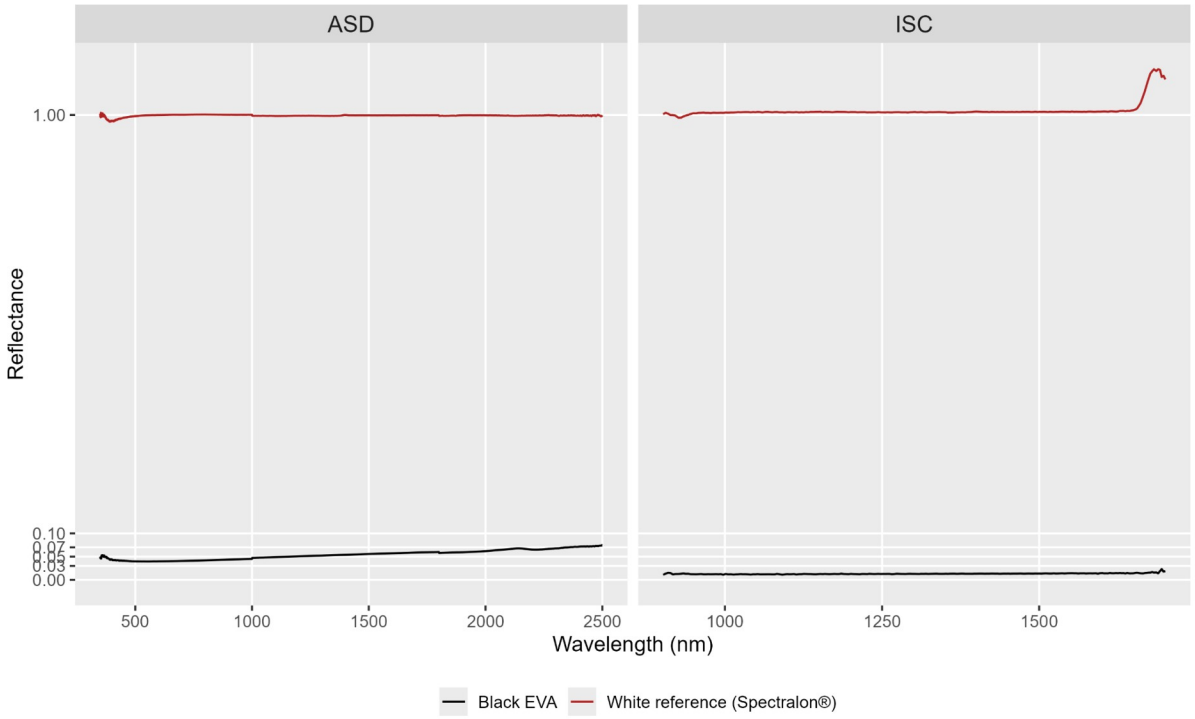


Fig. S2. Background reflectance measured on black EVA background using ASD FieldSpec® 4 and ISC NIR-S-G1. The spectra represent raw reflectance values and illustrate the low reflectance of the background material used during measurements.

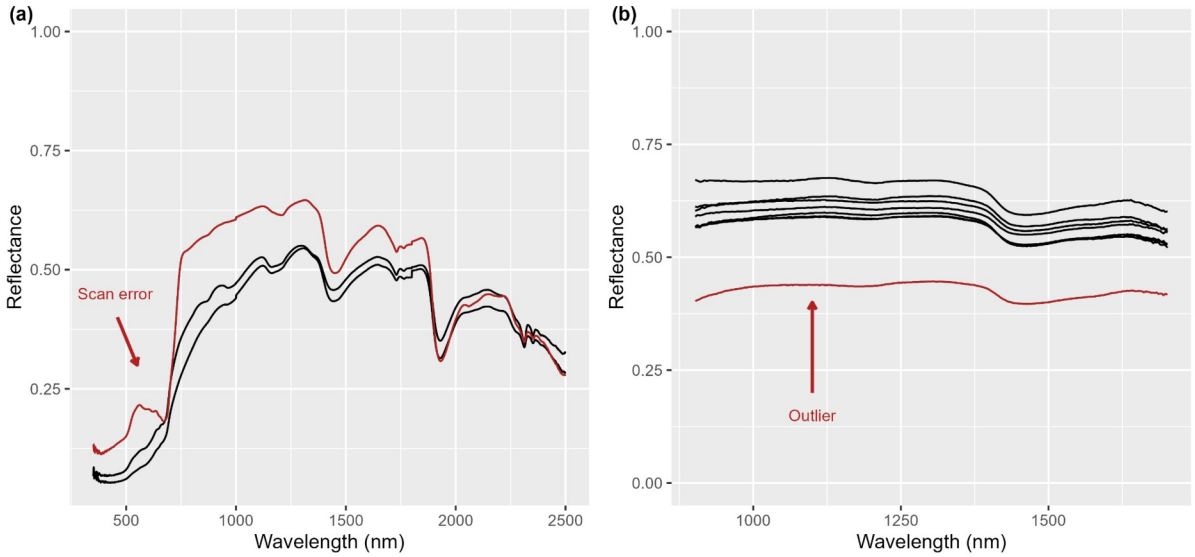
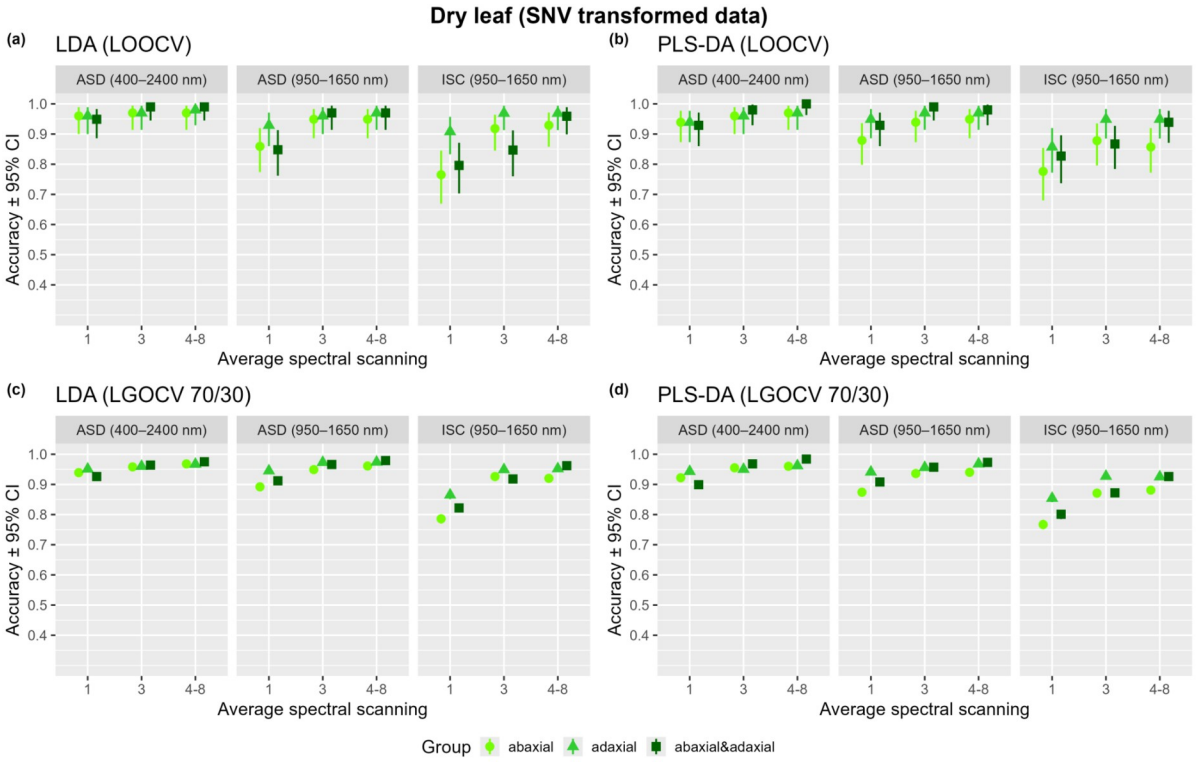


Fig. S3. Examples of deviations from expected spectral patterns excluded prior to analysis (quality control). (a) A reading error showing irregular reflectance values at shorter wavelengths, indicating acquisition failure. (b) An outlier with consistent deviation from the pattern of other spectra.



793

794 **Fig. S4.** Classification accuracy based on preprocessed data from the dry leaf spectra. Panels
795 show the results of (a) Linear Discriminant Analysis (LDA) with Leave-One-Out Cross-
796 Validation (LOOCV), (b) Partial Least Squares Discriminant Analysis (PLS-DA) with
797 LOOCV, (c) LDA with 70/30 Leave-Group-Out Cross-Validation (LGOCV), and (d) PLS-
798 DA with LGOCV. Each panel compares model performance across instruments (ASD full
799 range 400–2400 nm, and trimmed band 950–1650 nm; ISC: 950–1650 nm) and different
800 levels of spectral scan averaging (1, 3, and 4–8 scans). Marker shapes indicate the leaf
801 surface scanned: abaxial (squares), adaxial (triangles), and both surfaces combined (circles).

802

Fig. S5. Raw (a) and preprocessed (b) spectral reflectance of dry leaves measured with ASD FieldSpec® 4 and ISC NIR-S-G1**.** The dark green line represents the mean reflectance, and the green shaded area indicates the full range (minimum to maximum) of reflectance values. Vertical pink bands highlight wavelengths identified as most informative by LDA-based variable selection.

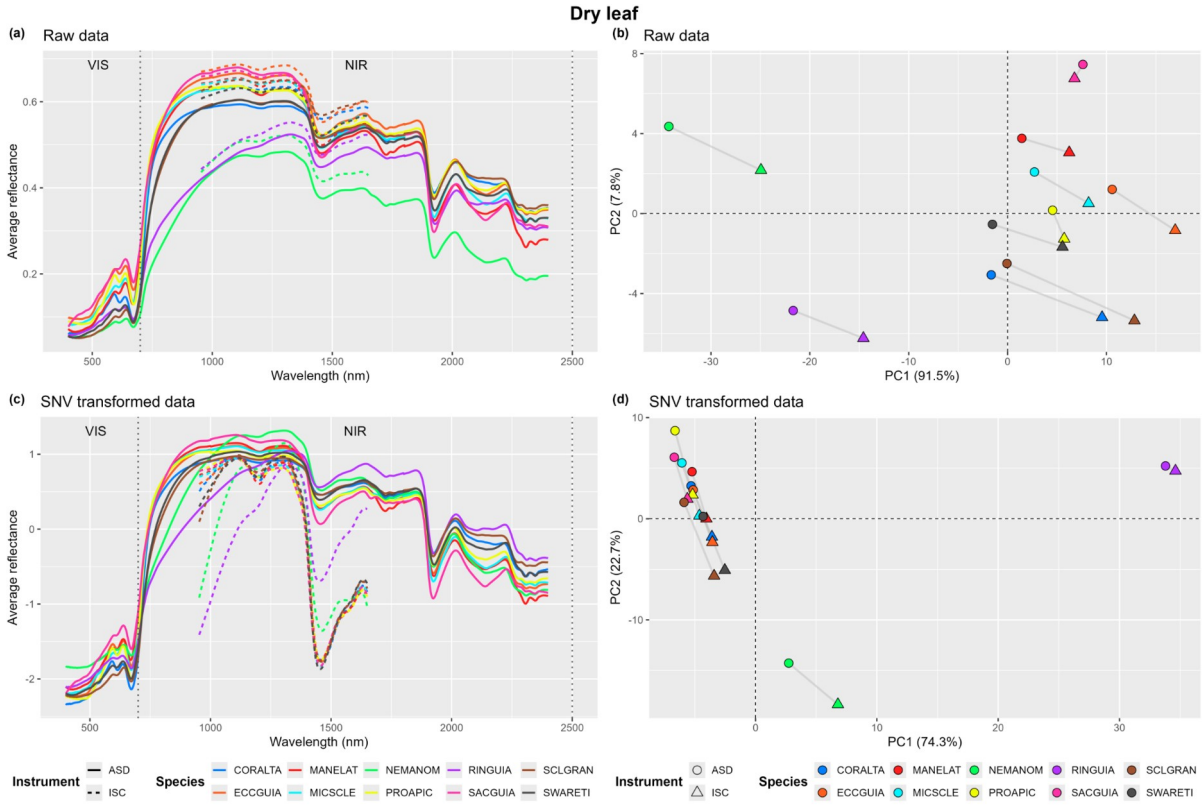


Fig. S6. Comparison between instruments based on raw (a, b) and preprocessed (c, d) dry leaf spectra. (a, c) Mean reflectance spectra measured by species with ASD (solid lines) and ISC (dashed lines). Different colors indicate tree species (Table 1), and spectral regions (VIS and NIR) are marked for reference. (b, d) Scores of principal component analysis (PCA) of mean spectra by species and instrument. Circles represent ASD and triangles represent ISC; colors correspond to tree species (Table 1). Lines connect paired points from the same species measured with both instruments.

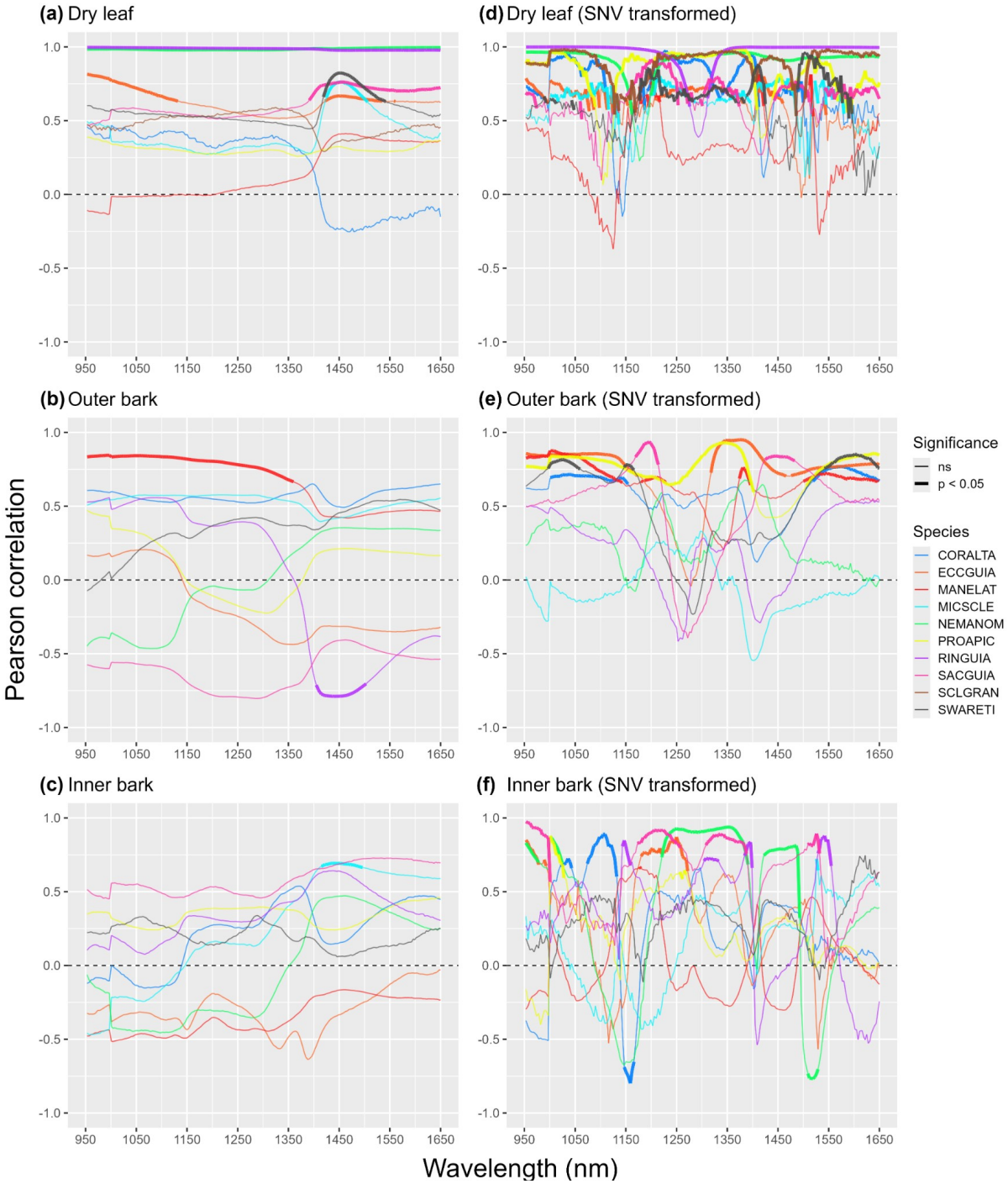


Fig. S7. Correlograms of Pearson correlations between ASD FieldSpec® 4 and ISC NIR-S-G1 reflectance values for each species, calculated across wavelengths for raw (a–c) and SNV-transformed (d–f) spectra of dry leaves (a, d), outer bark (b, e), and inner bark (c, f). Colored lines represent individual species. Line thickness denotes statistical significance of the correlation at each wavelength (thin = non-significant, $p \geq 0.05$; thick = significant, $p < 0.05$).

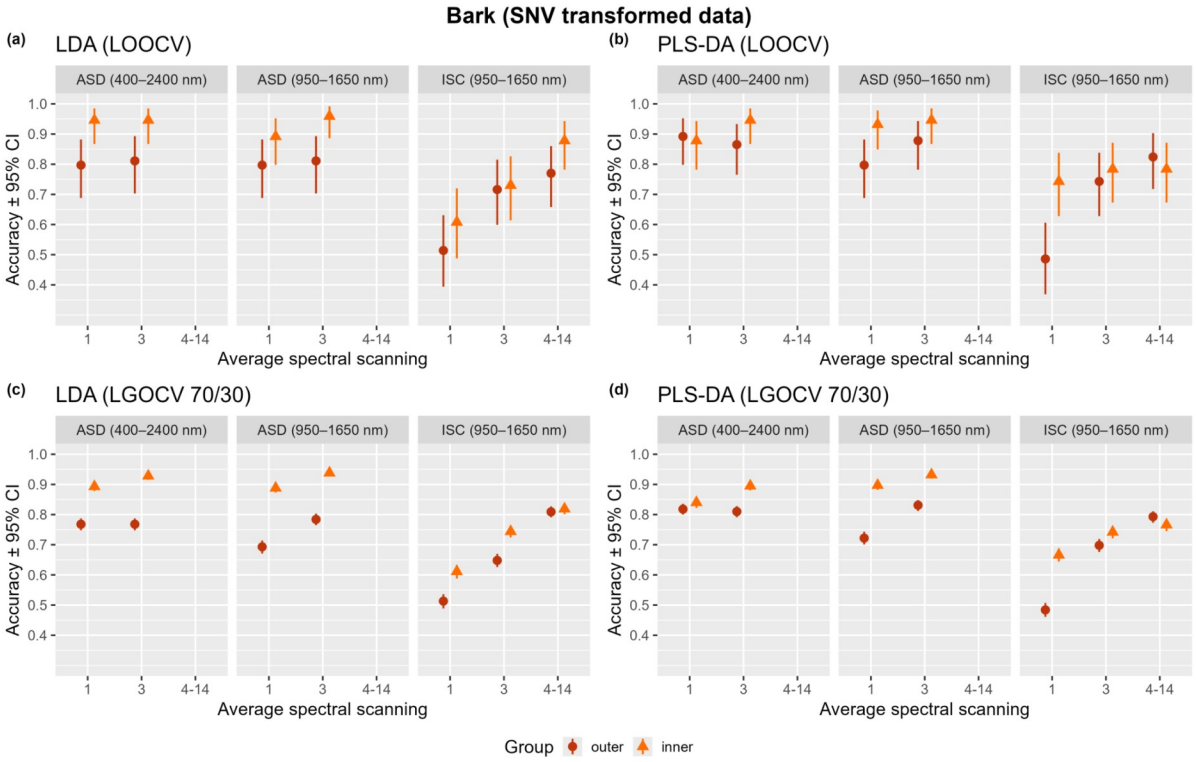


Fig. S8. Classification accuracy based on preprocessed data from the bark of standing trees. Panels show the results of (a) Linear Discriminant Analysis (LDA) with Leave-One-Out Cross-Validation (LOOCV), (b) Partial Least Squares Discriminant Analysis (PLS-DA) with LOOCV, (c) LDA with 70/30 Leave-Group-Out Cross-Validation (LGOCV), and (d) PLS-DA with LGOCV. Each panel compares the performance across instruments (ASD: 400–2450 nm and 950–1650 nm; ISC: 950–1650 nm) and different levels of spectral scan averaging (1, 3, and 4–14 scans). Marker shapes indicate the scanned bark surface: outer (circles) and inner (triangles) bark.

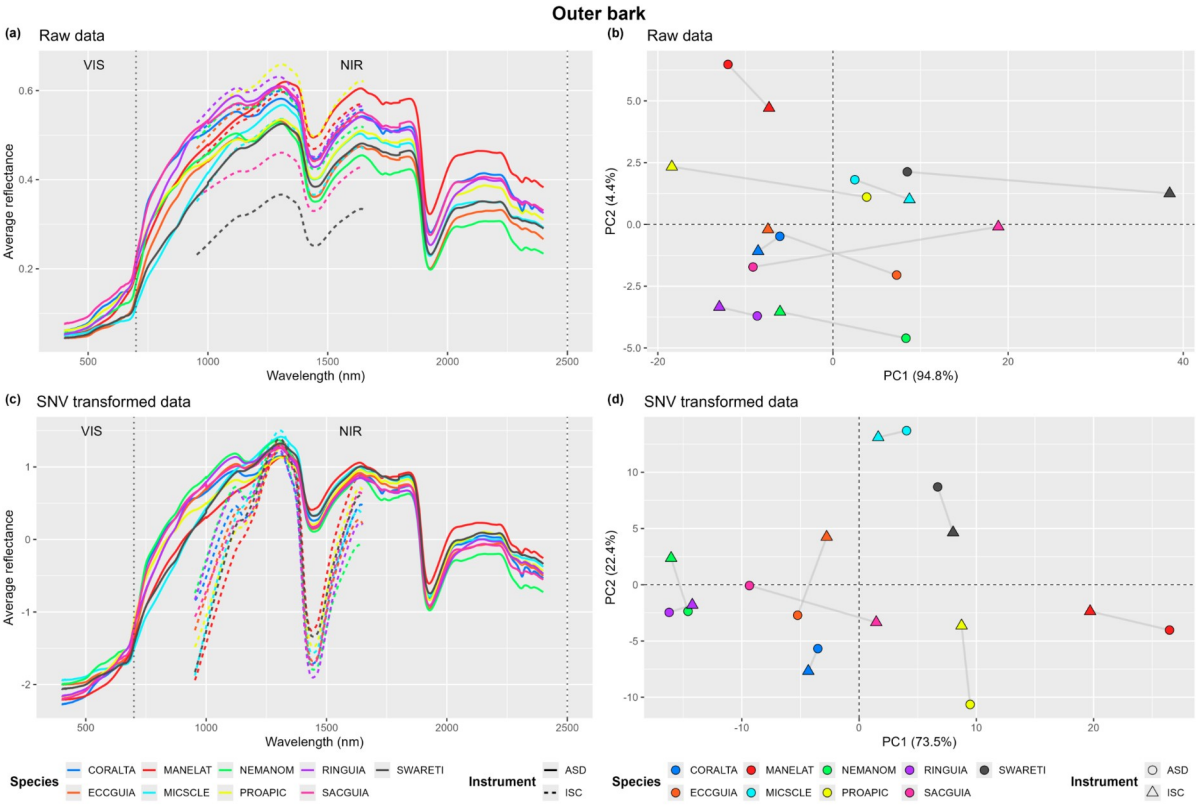


Fig. S9. Comparison between instruments based on raw (a, b) and preprocessed (c, d) outer bark spectra. (a, c) Mean reflectance spectra measured by species with ASD (solid lines) and ISC (dashed lines). Different colors indicate tree species (Table 1), and spectral regions (VIS and NIR) are marked for reference. (b, d) Principal component analysis (PCA) of mean spectra by species and instrument. Circles represent ASD and triangles represent ISC; colors correspond to tree species (Table 1). Lines connect paired points from the same species measured with both instruments.

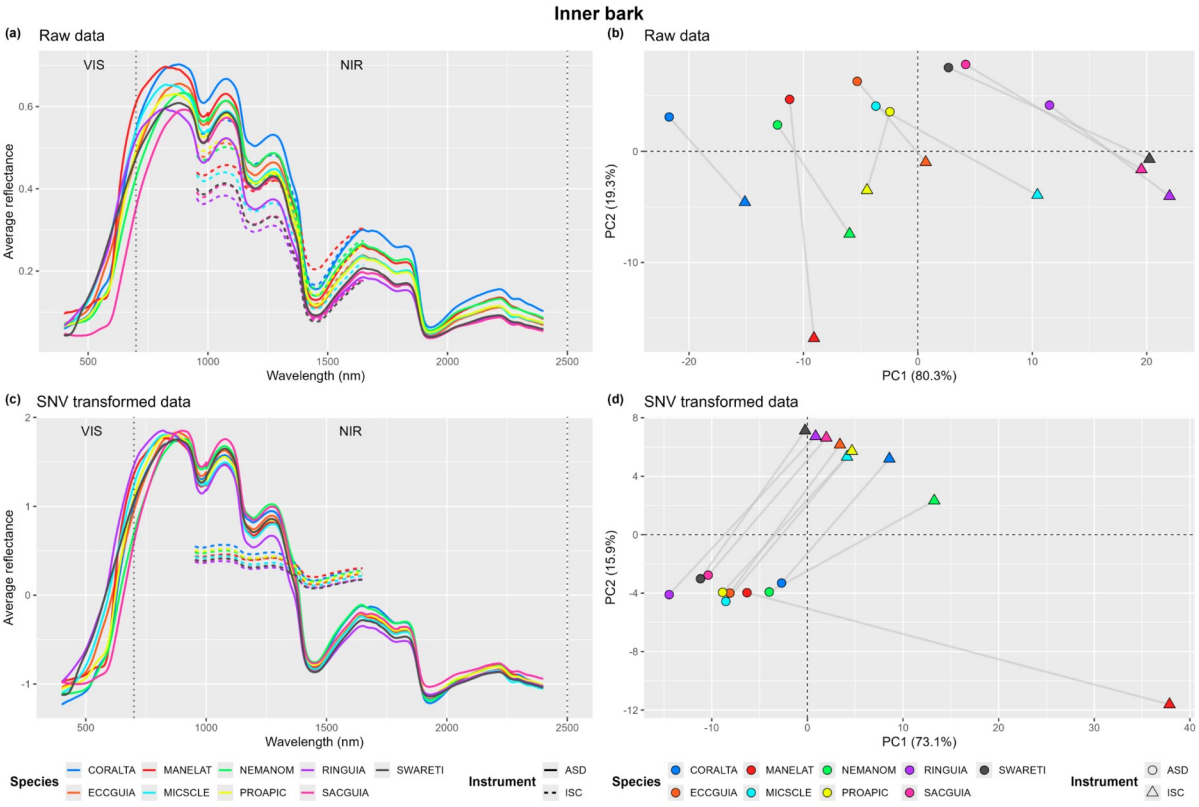


Fig. S10. Comparison between instruments based on raw (a, b) and preprocessed (c, d) inner bark spectra. (a, c) Mean reflectance spectra measured by species with ASD (solid lines) and ISC (dashed lines). Different colors indicate tree species (Table 1), and spectral regions (VIS and NIR) are marked for reference. (b, d) Principal component analysis (PCA) of mean spectra by species and instrument. Circles represent ASD and triangles represent ISC; colors correspond to tree species (Table 1). Lines connect paired points from the same species measured with both instruments.

Fig. S11. Raw (a) and preprocessed (b) spectral reflectance of outer bark of standing trees measured with ASD FieldSpec® 4 and ISC NIR-S-G1. Dark brown (a) and orange (b) lines represent the mean reflectance, while shaded areas indicate the full range (minimum to maximum) of reflectance values. Vertical pink bands highlight wavelengths identified as most informative by LDA-based variable selection.

Fig. S12. Raw (a) and preprocessed (b) spectral reflectance of inner bark of standing trees measured with ASD FieldSpec® 4 and ISC NIR-S-G1. Dark brown (a) and orange (b) lines represent the mean reflectance, while shaded areas indicate the full range (minimum to maximum) of reflectance values. Vertical pink bands highlight wavelengths identified as most informative by LDA-based variable selection.

Table S1

List of R packages used in our study.

Description	Name	Authorship	Identifier
Data importing	asdreader	Roudier (2017)	10.32614/CRAN.package.asdreader
	readr	Wickham et al. (2024)	10.32614/CRAN.package.readr
Data manipulation	data.table	Barrett et al. (2025)	10.32614/CRAN.package.data.table
	dplyr	Wickham et al. (2023)	10.32614/CRAN.package.dplyr
	purr	Wickham et al. (2025)	10.32614/CRAN.package.purrr
	reshape2	Wickham (2020)	10.32614/CRAN.package.reshape2
	stringr	Wickham (2023)	10.32614/CRAN.package.stringr
	tibble	Müller et al. (2025)	10.32614/CRAN.package.tibble

	tidyr	Wickham et al. (2024)	10.32614/CRAN.package.tidyr
Data analysis	caret	Kuhn et al. (2024)	10.32614/CRAN.package.caret
	MASS	Ripley et al. (2025)	10.32614/CRAN.package.MASS
	pls	Liland et al. (2024)	10.32614/CRAN.package.pls
	prospectr	Stevens & Ramirez-Lopez (2025)	10.32614/CRAN.package.prospectr
Data visualization	htmlwidgets	Vaidyanathan et al. (2023)	10.32614/CRAN.package.htmlwidgets
	ggplot2	Wickham et al. (2025)	10.32614/CRAN.package.ggplot2
	ggpubr	Kassambara (2025)	10.32614/CRAN.package.ggpubr
	plotly	Sievert et al. (2025)	10.32614/CRAN.package.plotly