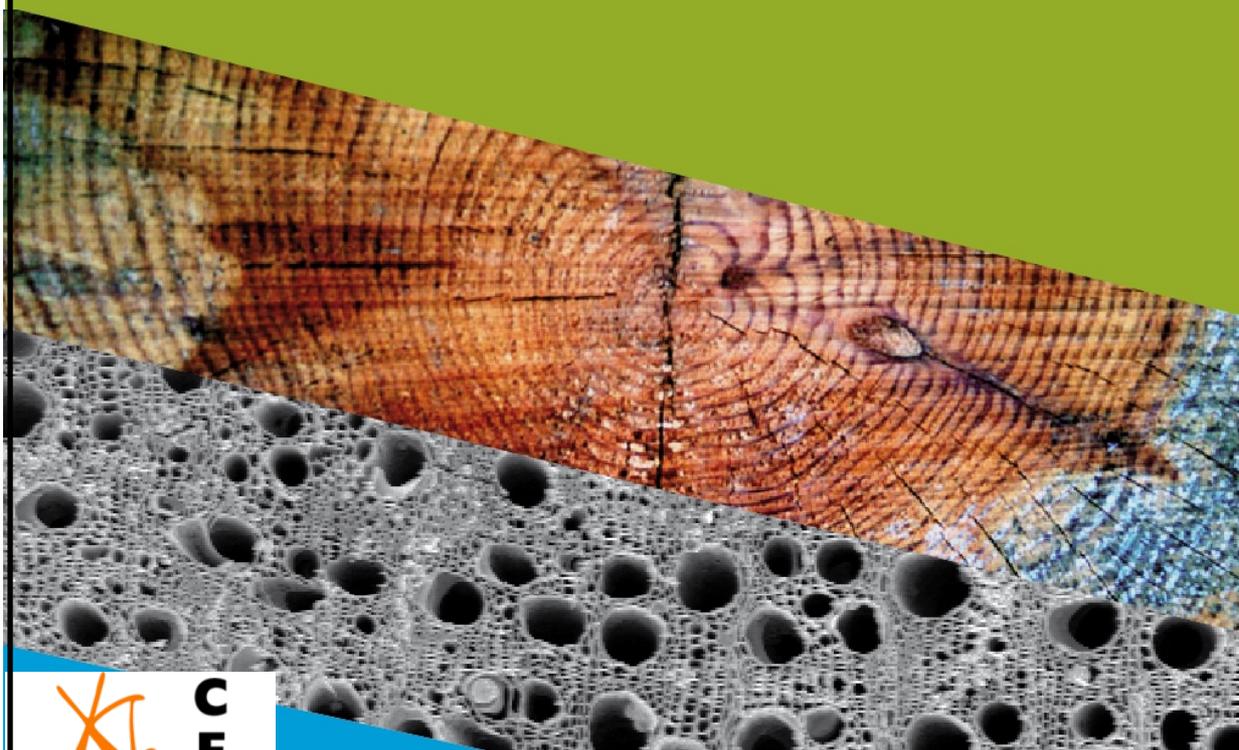


11èmes Journées Scientifiques du GDR3544 Sciences du bois

Nice 16-18 Novembre 2022



UNIVERSITÉ
CÔTE D'AZUR 



GDR

Groupement
de recherche

Sciences du bois

Actes provisoires

version du 11 novembre 2022

Influence de l'étendue de la gamme spectrale sur la performance des modèles de discrimination SPIR : cas de six espèces de *Diospyros* de Madagascar

RAOBELINA Andry Clarel^{1,4}, CHAIX Gilles^{2,3,4}, RAMANANANTOANDRO Tahiana¹

¹Mention Foresterie et Environnement – Université d'Antananarivo, BP 175 Antananarivo
101, Madagascar

² CIRAD, UMR AGAP Institut, F-34398 Montpellier, France

³ UMR AGAP Institut, Univ Montpellier, CIRAD, Institut Agro, Montpellier, France

⁴ ChemHouse Research Group, Montpellier, France

andryclarel@gmail.com

Mots clefs : Spectrométrie Proche Infrarouge, *Diospyros*, Madagascar, gamme spectrale

Introduction

Le genre *Diospyros* de Madagascar regroupe actuellement 255 espèces dont 82 sont considérées comme des grands arbres (Schatz et al. 2021). Plusieurs de ces espèces sont exploitées illégalement, ce qui a entraîné l'insertion de toutes les espèces de *Diospyros* malgaches dans l'annexe II de la Convention sur le Commerce international des espèces de faune et de flore sauvage menacées d'extinction (CITES) en 2013 (Ratsimbazafy et al. 2016). Le développement d'outil d'identification comme la Spectrométrie Proche InfraRouge (SPIR), qui permet d'identifier les ébènes malgaches à partir de leurs bois, est en cours à Madagascar (Raobelina et al 2021) pour mettre en œuvre le plan d'action de la CITES dans l'objectif d'une exploitation et d'un commerce durable de ces espèces.

La SPIR est parmi les techniques utilisées dans l'identification des espèces anatomiquement proches et dont la commerce est réglementée par la CITES (Pastore et al. 2011, Tsuchikawa et Kobori 2015, Snel et al. 2018). Plusieurs modèles de spectromètres ont été utilisés dans la littérature dont les caractéristiques varient selon leur portabilité, leur gamme spectrale, leur résolution spectrale, etc. Plus particulièrement, l'utilisation des spectromètres portatifs s'est beaucoup développée ces dernières années (Yan et Siesler 2018, Zhu et al. 2021, Giussani et al. 2022). Ces modèles sont plus accessibles pour les pays en développement comme Madagascar étant donné qu'ils sont moins onéreux. Mais leurs caractéristiques techniques sont moins performantes que celles des spectromètres de laboratoire en présentant une gamme spectrale limitée et une résolution spectrale moindre. L'étendue de la gamme spectrale peut être également différente d'un modèle à un autre (Zhu et al. 2022). La performance des modèles d'étalonnage pourrait par conséquent varier selon des modèles de spectromètres car les informations chimiques sous-jacentes aux spectres est liée à la gamme spectrale. Peu d'études ont été publiées sur la détermination de la région spectrale idéale pour discriminer les essences à partir des spectres du bois. Ces connaissances sont pourtant importantes pour bien choisir les spécifications techniques lors de l'acquisition d'un spectromètre portable. La question principale adressée dans le cadre de cette étude est alors la suivante. Quelle est la plage idéale de longueurs d'onde dans le proche infrarouge pour discriminer les essences d'ébènes malagasy sur la base des spectres PIR de leurs bois ?

Matériels et méthodes

Collecte et préparation des carottes de bois. Les travaux de terrain ont été effectués par les botanistes de la Missouri Botanical Garden (MBG) de Madagascar dans le cadre du projet Gestion Durable des bois précieux *Dalbergia* et *Diospyros* de Madagascar (G3D) financé par l'Union Européenne et mis en œuvre par l'université d'Antananarivo. Soixante-dix-huit carottes de bois, appartenant à six espèces de *Diospyros*, ont été collectées sur 78 arbres de plusieurs aires protégées réparties dans la région Boeny et Diana (région Nord de Madagascar). Les carottes de 5 mm de diamètre ont été collectées sur toute la largeur du tronc des à 1,30 m à partir du sol, perpendiculairement à l'axe longitudinal de l'arbre. L'identification taxonomique des carottes a été effectuée par les experts taxonomistes de MBG de Madagascar et du Musée National d'Histoire Naturelle de France à travers des analyses morphologiques des organes végétatifs et reproductifs sur les herbiers.

Les carottes de bois ont été ensuite conditionnées à un taux d'humidité théorique de 12% selon la courbe d'isotherme de sorption du bois (AFNOR, 2010), en utilisant une enceinte climatique réglée à une température de 20°C et une humidité relative de 65% afin de limiter les effets de la différence d'humidité du bois sur les spectres.

Mesure des spectres PIR. Les spectres ont été mesurés à l'aide d'un spectromètre Bruker MPA II entre 12500 cm^{-1} à 4000 cm^{-1} de nombre d'ondes, correspondant à 800 – 2500 nm de longueurs d'ondes, et à une résolution de 8 cm^{-1} (Fig. 1). Six spectres ont été acquis sur la partie duramen du bois de chaque carotte dans la direction radiale afin de considérer la variabilité radiale du bois, ce qui fait un total de 468 spectres pour l'ensemble des échantillons.

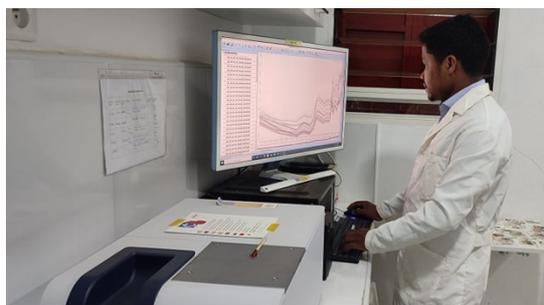


Fig. 1 : Mesure des spectres avec le Bruker MPA II

Analyse et traitement des données. Toutes les analyses et traitements ont été effectués avec le logiciel R version 4.1.2 (R core Team 2021) et le package rchemo version 0.0-17 (Lesnoff, 2021). Une analyse en composantes principales a été d'abord effectuée à partir des données spectrales $X_{(468 \times 2974)}$ prétraitées par lissage suivi d'une dérivation première (SG1 [W = 15 points]) afin d'observer, d'une part, le groupement des spectres en fonction des espèces dans le plan formé par les composantes principales (CPs) et, d'autre part, pour évaluer la contribution des différentes longueurs d'ondes dans la plage 12500 cm^{-1} à 4000 cm^{-1} par les « loadings » des spectres.

Six bases de données ont ensuite été extraites à partir du jeu de données initial. La première est constituée par les données d'origine (spectres et données de références $Y_{(468 \times 6)}$), c'est-à-dire des spectres comprenant les absorbances dans toute la gamme spectrale (12500 cm^{-1} à 4000 cm^{-1}). Les cinq autres bases de données sont extraites des données originales, d'une part, sur des gammes de longueur d'onde différente selon le type de spectromètres portables existants et courants : 950-1650 correspondant aux spectromètres avec détecteur InGas et, d'autre part, sur les autres gammes correspondant aux spectromètres de la marque NIRONe (Photinnov 2021, Tab. 1).

Tab. 1 : Caractéristiques des six jeux de données et résultats de classification en validation indépendante spectres prétraités par SG1 (W = 15 points)

Modèle	Données d'entraînement	Région spectrale	Nombre de carottes/nombre de spectres en [Etalonnage] et [Validation]	DV _s	Acc.(%)	Rec.(%)	Prc.(%)	
PLSDA₀	X _(468 × 2974) Y _(468 × 6)	12495 – 3998 cm ⁻¹	<i>D.analamerensis</i> [10/60][4/24]	4	72,5	75	69	
			<i>D.chitoniophora</i> [14/84][4/24]				56	
			<i>D.clusiifolia</i> [8/48][3/18]				72	
		800 – 2500 nm	<i>D.mapingo</i> [11/66][4/24]				88	100
			<i>D.parifolia</i> [7/42][2/12]				8	17
			<i>D.platycalyx</i> [8/48][3/18]				83	100
PLSDA₁	X _(468 × 475) Y _(468 × 6)	7407 – 6060 cm ⁻¹	<i>D.analamerensis</i> [10/60][4/24]	19	71,7	75	58	
			<i>D.chitoniophora</i> [14/84][4/24]				83	95
			<i>D.clusiifolia</i> [8/48][3/18]				78	64
		1350 – 1650 nm	<i>D.mapingo</i> [11/66][4/24]				83	80
			<i>D.parifolia</i> [7/42][2/12]				75	69
			<i>D.platycalyx</i> [8/48][3/18]				28	63
PLSDA₂	X _(468 × 464) Y _(468 × 6)	6451 – 5128 cm ⁻¹	<i>D.analamerensis</i> [10/60][4/24]	5	40,9	42	47	
			<i>D.chitoniophora</i> [14/84][4/24]				63	65
			<i>D.clusiifolia</i> [8/48][3/18]				17	15
		1550 – 1950 nm	<i>D.mapingo</i> [11/66][4/24]				42	45
			<i>D.parifolia</i> [7/42][2/12]				8	9
			<i>D.platycalyx</i> [8/48][3/18]				17	38
PLSDA₃	X _(468 × 372) Y _(468 × 6)	5714 – 4651 cm ⁻¹	<i>D.analamerensis</i> [10/60][4/24]	11	74,2	61	75	
			<i>D.chitoniophora</i> [14/84][4/24]				100	92
			<i>D.clusiifolia</i> [8/48][3/18]				79	68
		1750 – 2150 nm	<i>D.mapingo</i> [11/66][4/24]				79	68
			<i>D.parifolia</i> [7/42][2/12]				67	57
			<i>D.platycalyx</i> [8/48][3/18]				50	100
PLSDA₄	X _(468 × 323) Y _(468 × 6)	5000 – 4082 cm ⁻¹	<i>D.analamerensis</i> [10/60][4/24]	11	99,9	100	100	
			<i>D.chitoniophora</i> [14/84][4/24]				100	89
			<i>D.clusiifolia</i> [8/48][3/18]				100	95
		2000 – 2450 nm	<i>D.mapingo</i> [11/66][4/24]				100	92
			<i>D.parifolia</i> [7/42][2/12]				42	100
			<i>D.platycalyx</i> [8/48][3/18]				100	100
PLSDA₅	X _(468 × 1563) Y _(468 × 6)	10526 – 4081 cm ⁻¹	<i>D.analamerensis</i> [10/60][4/24]	30	45,9	67	47	
			<i>D.chitoniophora</i> [14/84][4/24]				58	54
			<i>D.clusiifolia</i> [8/48][3/18]				67	67
		950 – 1650 nm	<i>D.mapingo</i> [11/66][4/24]				46	42
			<i>D.parifolia</i> [7/42][2/12]				8	25
			<i>D.platycalyx</i> [8/48][3/18]				17	19

Chaque base de données a été partitionnée aléatoirement en jeux de données d'étalonnage (75%) et de validation (25%), en gardant les six spectres d'un même carotte dans un même jeu pour éviter le surajustement des modèles. Une validation croisée à 4 blocs, répétée 20 fois avec 30 Variables Discriminantes (VDs) au maximum, a été effectuée à partir de chaque jeu d'étalonnage afin d'identifier les meilleurs prétraitements et le nombre optimal de VDs correspondant au minimum d'erreur de classification. Les meilleurs modèles ont ensuite été testés pour classifier les spectres du jeu de validation. La performance des modèles a été évaluée à partir de trois métriques de classification (Lapcharoensuk et al., 2019) dont « Accuracy » qui est le pourcentage global de spectres bien classés par le modèle pour les six espèces, « Recall » qui exprime le nombre de spectres bien classés sur le nombre total de spectre pour une classe déterminée, et « Precision » qui est le nombre de spectres bien classés pour une classe sur le

nombre total de spectre prédites en tant que cette classe, elle exprime la probabilité qu'un spectre inconnu appartient à la classe prédite par le modèle.

Résultats et discussion

Exploration des données. L'allure des spectres originaux est similaire pour les six espèces de *Diospyros* (Fig. 2a). Sur la base des résultats de l'ACP, deux groupes d'espèces séparés par l'axe CP1 peuvent être observés. Les spectres de *D. analamerensis* et *D. chitoniophora* forment un groupe généralement situé dans la partie négative de CP1. Tandis qu'un deuxième groupe, formé par les spectres de *D. mapingo*, *D. parifolia* et *D. platycalyx*, se trouve dans la partie positive. Aucun groupement des spectres de *D. clusiifolia* est observé car ses spectres se chevauchent avec ceux des cinq autres espèces dans le plan CP1-CP2. Les « loadings » issu de l'ACP pour les deux premières CPs mettent en exergue la contribution des longueurs d'onde aux alentours de 4100, 4350, 4500, 4850, 5200 et 7100 cm^{-1} (Fig. 2c).

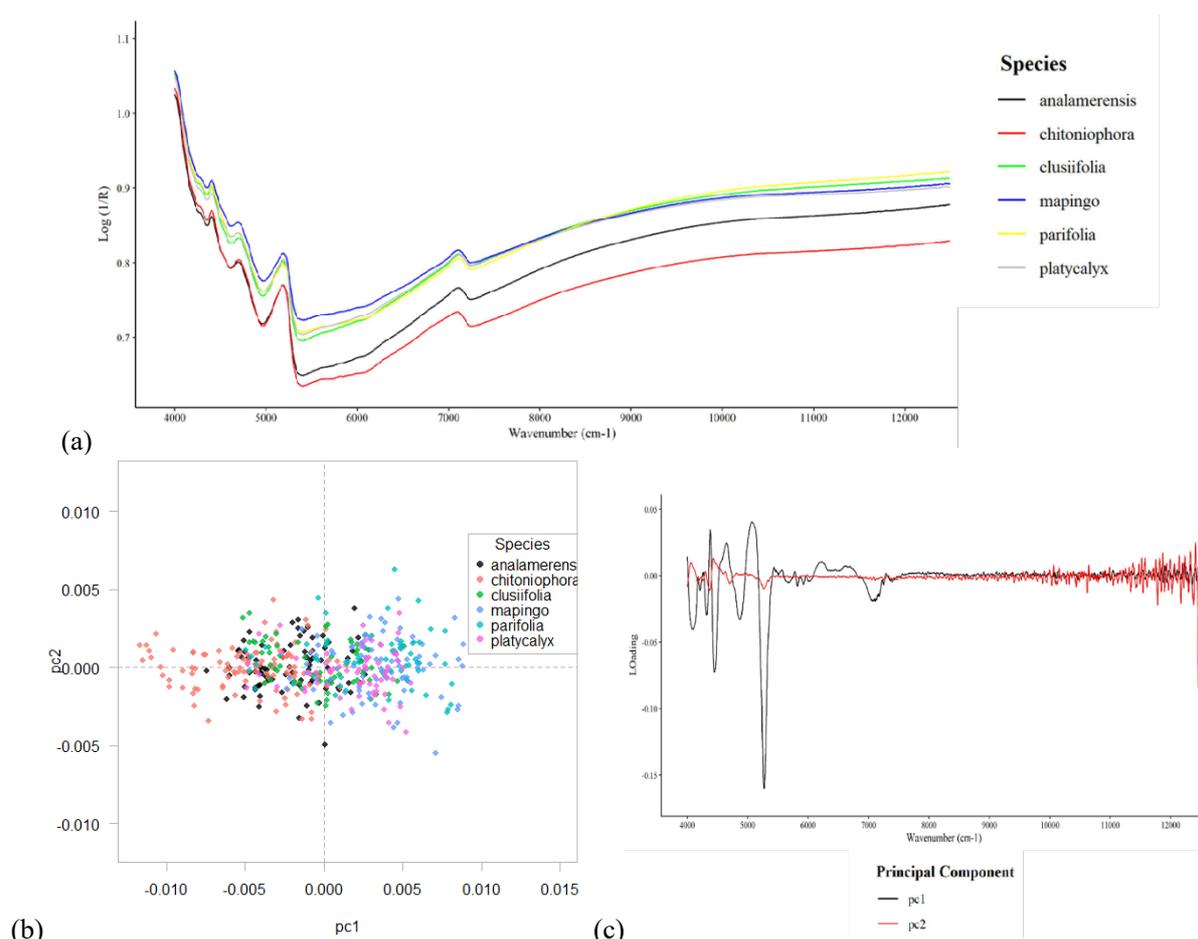


Fig. 2 : Allures des spectres bruts moyens par espèce : (a) Projections des échantillons dans le plan PC1-PC2 de l'ACP ; (b) « loadings » pour PC1 et PC2 issu de l'ACP sur les spectres.

Analyse discriminante. La performance globale des modèles varie significativement en fonction de la gamme spectrale considérée (Tab. 1). La région 5000 – 4082 cm^{-1} permet de mieux discriminer les six espèces de *Diospyros* avec une « Accuracy » de 99,9%). Les modèles basés sur les gammes spectrales supérieures à 5000 cm^{-1} sont moins performants (Tab. 1), notamment pour les modèles PLSDA₂ (Accuracy = 40,9%) et PLSDA₅ (Accuracy = 45,9%) qui ont été étalonnés respectivement dans la région 6451 – 5128 cm^{-1} et 1526 – 6060 cm^{-1} . Ce résultat peut être expliqué par le fait que la région avant 5000 cm^{-1} , spécifique aux bandes de combinaisons (Schwanninger et al., 2011), est la plus informative pour le bois. Cela est confirmé par les

résultats de l'ACP qui ont mis en exergue la présence de la plupart des pics d'informations dans la région 5000 – 4000 cm⁻¹.

La performance en classification au niveau espèce varie d'un modèle à un autre. Le meilleur modèle PLSDA₄ est capable d'identifier correctement les carottes pour *D. analamerensis*, *D. chitoniophora*, *D. clusifolia*, *D. mapingo* et *D. platycalyx* car leurs spectres (dans le jeu de validation) ont été tous bien classés (Recall=100%). Pour ces six espèces, la certitude moyenne des prédictions, pour que le spectre appartienne réellement à la classe prédite par le modèle, est de 95,2%. Les prédictions ne présentent aucun risque de confusion entre elles ou avec les autres espèces pour *D. analamerensis* et *D. platycalyx* (Precision=100%). La capacité du modèle, à identifier les spectres du *D. parifolia* par rapport aux autres espèces, est pourtant plus faible (Recall=42%, Tab. 1, Tab. 2). Mais les prédictions par le modèle sont toujours certaines pour cette espèce (Precision=100%).

Tab. 2 : Matrice de confusion issue de la prédiction des spectres de validation SG1 (W = 15 points) par le modèle PLSDA₄

	<i>D. analamerensis</i>	<i>D. chitoniophora</i>	<i>D. clusifolia</i>	<i>D. mapingo</i>	<i>D. parifolia</i>	<i>D. platycalyx</i>
<i>D. analamerensis</i>	24					
<i>D. chitoniophora</i>		24		4		
<i>D. clusifolia</i>			18	1		
<i>D. mapingo</i>				24	2	
<i>D. parifolia</i>					5	
<i>D. platycalyx</i>						18

Conclusion et perspectives

Le choix de l'étendue de la gamme spectrale, sur laquelle les modèles de discrimination sont étalonnés, a une influence considérable sur la performance de ces derniers. La région avant 5000 cm⁻¹ est la plus importante pour étalonner des modèles pour le cas des six espèces de *Diospyros* de Madagascar à partir des spectres PIR du bois. Le meilleur modèle (dans la plage de longueurs d'onde 5000 – 4082 cm⁻¹) a une performance globale de 99,9%, alors qu'elle diminue jusqu'à 40-45% pour les régions avant 5000 cm⁻¹. Tester un spectromètre PIR portable (qui couvre la région 5000 – 4082 cm⁻¹ tel le NIRone 2.5 (Photinnov, 2021)), pour discriminer les six espèces de *Diospyros*, puis comparer les résultats avec ceux de la présente étude, serait intéressant dans l'avenir. Cela permettrait de renforcer le choix d'un spectromètre couvrant cette gamme pour avoir des modèles performants, et prendre en compte les caractéristiques techniques du spectromètre comme la résolution spectrale, son optique et l'éclairage.

Remerciements

Cette étude a été réalisée sous le financement de la Délégation de l'Union Européenne à Madagascar dans le cadre du programme Rindra et du projet G3D (Gestion Durable des Bois précieux *Dalbergia* et *Diospyros* de Madagascar). Les auteurs adressent leurs gratitude à l'équipe de l'UFR Sciences du Bois de l'ESSA-Forêts, l'équipe des taxonomistes et botanistes de terrain de la MBG-Madagascar et de MNHN Paris pour leurs contributions. Les remerciements s'adressent également au CIRAD pour son appui dans le cadre de l'Actions incitatives – Appui aux doctorants du Sud.

Références

- AFNOR, (2010) NF EN 1995-1-1: Eurocode 5 - Conception et calcul des structures en bois - Partie 1-1 : généralités - Règles communes et règles pour les bâtiments.
- Giussani B., Gorla G., Riu J. (2022) Analytical Chemistry Strategies in the Use of Minaturised NIR instruments: An overview. *Critical reviews in Anal chem* 0(0):1-33.
- Lapcharoensuk R., Malithong A., Thappho D., Phonpho P. (2019) Discrimination of vegetable oil types using Fourier transforms near infrared spectroscopy coupled with pattern recognition techniques, *Earth and Environmental Science* 301(2019): 1-6, doi:10.1088/1755-1315/301/1/012067
- Lesnoff M. (2021). rchemo: Dimension reduction, Regression and Discrimination for Chemometrics. R package version 0.0-17. <https://github.com/mlesnoff/rchemo>
- Pastore T.C.M., Braga J.W.B., Coradin V.T.R., Magalhaes W.L.E., Okino E.Y.A., Camargos J.A.A., Muniz G.I.B., Bressan O.A., Davrieux F. (2011) Near infrared spectroscopy (NIRS) as a potential tool for monitoring trade of similar woods: Discrimination of true mahogany, cedar, andiroba, and curupixá. *Holzforschung* 65(1):73–80. <https://doi.org/10.1515/HF.2011.010>
- R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Raobelina A.C., Chaix G., Razafimahatratra A.R., Ramananantoandro T. (2021), Discrimination of three commercial valued *Dalbergia* species from Madagascar using Near-InfraRed-Spectroscopy, toward the development of an identification tool for supporting CITES enforcement, Virtual meeting of the ATBC; july 21-23, 2021 (oral presentation), https://www.atbc2021.org/_files/ugd/98d29f_00fc300c3a694391907f36ff13cea2ff.pdf
- Ratsimbazafy C., Newton D.J., Ringuet S. (2016) The rosewood and ebony trade of Madagascar. Timber Island, Traffic report. 125 pp.
- Schatz G.E., Lowry II P.P., Rakouth H.N., Randrianaivo R. (2021). Études taxonomiques du genre *Diospyros* (Ebenaceae) de la région malgache. VI. Nouvelles espèces de grands arbres à Madagascar. *Candollea* 76: 201–236. En anglais, résumés anglais et français. DOI: <http://dx.doi.org/10.15553/c2021v762a3>
- Snel F.A., Braga J.W.B., da Silva D., Wiedenhoeft A.C., Costa A., Soares R., Coradin V.T.R., Pastore T.C.M. (2018) Potential field-deployable NIRS identification of seven *Dalbergia* species listed by CITES. *Wood Sci Technol* 52(5):1411–1427. <https://doi.org/10.1007/s00226-018-1027-9>
- Schwanninger M., Rodrigues J.C., Fackler K. (2011) A review of band assignments in near infrared spectra of wood and wood components. *J Near Infrared Spec* 308(19):287-308.
- Yan H., Siesler H.W. (2018) Hand-held near-infrared spectrometers: State-of-the-art instrumentation and practical applications. *NIR news* 29(7):8–12. <https://doi.org/10.1177/0960336018796391>
- Zhu C., Fu X., Zhang J., Qin K., Wu C. (2021) Review of portable near infrared spectrometers: Current status and new techniques; *J Near Infrared Spec* 0(0):1-16.
- Photinnov (2021) Spectromètre Proche Infrarouge à Base de MEMS – NIRONE SENSOR S, <https://photinnov.com/product/spectrometre-infrarouge-a-base-de-mems-nirone-sensor-s/>, [visité le 15 Octobre 2022]