# Exploring the potential of Near Infrared Hyperspectral Imaging and chemometrics to discriminate soil seed bank of two central African timber species: *Erythrophleum suaveolens* (Guill. & Perr.) Brenan and *Erythrophleum ivorense* A. Chev.

**Chauvelin Douh**[1, 4]
**Jean-Louis Doucet**[1]
**Félicien Tosso**[1]
**Kasso Daïnou**[5, 6]
**Anaïs-Pasiphaé Gorel**[1, 2]
**Antoine Deryck**[3]
**Juan Antonio Fernández Pierna**[3]

[1] Forest is life, TERRA Teaching and Research Centre
Gembloux Agro-Bio Tech, University of Liège
Passage des Déportés 2, 5030 Gembloux, Belgium

[2] Ghent University
Faculty of Bioscience Engineering
Department of Plants and Crops
Laboratory of Plant Ecology
Ghent, Belgium

[3] Walloon Agricultural Research Centre (CRA-W)
Valorization of Agricultural Products Department
Chaussée de Namur n°24
5030 Gembloux, Belgium

[4] Université Marien N'gouabi
École nationale supérieure d'Agronomie et de Foresterie (ENSAF)
Laboratoire de Géomatique et d'Écologie tropicale appliquées
BP 69, Brazzaville, République du Congo

[5] Université Nationale d'Agriculture
BP 43, Kétou, Bénin

[6] Nature+ asbl/TERRA Research Centre, Central African Forests,
Gembloux Agro-Bio, Tech, University of Liege
Passage des Déportés 2, 5030 Gembloux, Belgium

**Auteur correspondant / Corresponding author:**
Chauvelin DOUH – douhch382@gmail.com /
chauvelin.douh@alumni.uliege.be

ORCID : https://orcid.org/0000-0003-0656-9831

**Photo 1.**
Near-infrared (NIR) hyperspectral imaging system (courtesy of the Walloon Agricultural Research Center, Belgium). Photo N. Kayoka Mukendi.

**Doi :** 10.19182/bft2024.360.a37138 – Droit d'auteur © 2024, Bois et Forêts des Tropiques – © Cirad – Date de soumission : 27 février 2023 ; date d'acceptation : 22 novembre 2023 ; date de publication : 1er juin 2024.



**Citer l'article / To cite the article**
Douh C., Doucet J.-L., Tosso F., Daïnou K., Gorel A.-P., Deryck A., Fernández Pierna J. A., 2024. Exploring the potential of Near Infrared Hyperspectral Imaging and chemometrics to discriminate soil seed bank of two central African timber species: *Erythrophleum suaveolens* (Guill. & Perr.) Brenan and *Erythrophleum ivorense* A. Chev. Bois et Forêts des Tropiques, 360 : 1-10. Doi : https://doi.org/10.19182/bft2024.360.a37138

Bois et Forêts des Tropiques – ISSN: L-0006-579X
Volume 360 – 2nd quarter – June 2024 – p. 1-10
RESEARCH / *ERYTHROPHLEUM SUAVEOLENS* AND *E. IVORENSE*



## RÉSUMÉ

**Exploration du potentiel de l'imagerie hyperspectrale proche infrarouge et de la chimiométrie pour discriminer la banque de graines du sol de deux espèces de bois d'Afrique centrale : *Erythrophleum suaveolens* (Guill. & Perr.) Brenan et *Erythrophleum ivorense* A. Chev.**

Les graines contenues dans la banque du sol peuvent être trop petites pour être quantifiées visuellement. Les espèces de bois concernées sont généralement identifiées lors d'essais de germination en pépinière, ce qui prend du temps. Cette étude explore une nouvelle approche fondée sur l'imagerie hyperspectrale dans le proche infrarouge (NIR-HSI) couplée à des outils chimiométriques. Elle se concentre sur la banque de graines du sol des forêts denses humides d'Afrique centrale, qui est encore méconnue. Nous avons utilisé quatre-vingt-trois graines de deux espèces sœurs, *Erythrophleum suaveolens* (Guill. & Perr.) Brenan et *Erythrophleum ivorense* A. Chev., collectées dans le sol forestier (profondeur entre 0 et 10 cm) au Gabon, au Cameroun et au Congo. À l'aide d'analyses en composantes principales et d'analyses discriminantes par moindres carrés partiels, nous avons étudié la capacité de l'imagerie hyperspectrale proche infrarouge à identifier les graines de deux espèces. La méthode est rapide, non-destructive et offre de nouvelles perspectives pour l'étude de la banque de graines des sols forestiers.

**Mots-clés** : banque de graines, espèces de bois, imagerie hyperspectrale proche infrarouge, chimiométrie, identification, discrimination.

## ABSTRACT

**Exploring the potential of Near Infrared Hyperspectral Imaging and chemometrics to discriminate soil seed bank of two central African timber species: *Erythrophleum suaveolens* (Guill. & Perr.) Brenan and *Erythrophleum ivorense* A. Chev.**

Seeds contained in the soil bank can be too small to quantify visually. Concerned timber species are usually identified after germination trials in the nursery, which is time-consuming. This study explores a new approach based on near infrared (NIR-HSI) hyperspectral imaging coupled with chemometric tools. It focuses on the soil seed bank of the central African moist forests, which is still unknown. We used eighty-three seeds of two sister species, *Erythrophleum suaveolens* (Guill. & Perr.) Brenan and *Erythrophleum ivorense* A. Chev., collected in the forest soil (between 0 and 10 cm in depth), in Gabon, Cameroon, and Congo. Applying principal component analysis and partial least squares discriminant analysis, we studied the capacity of near-infrared hyperspectral imaging to identify the seeds of the two timber species. This method is fast, non-destructive, and offers new prospects for studies of forest soil seed banks.

**Keywords**: soil seed bank, timber species, near-infrared hyperspectral imaging, chemometrics, identification, discrimination.

## RESUMEN

**Estudio del potencial de la imagen hiperespectral del infrarrojo cercano combinado con la quimiometría para diferenciar dos especies de árboles centroafricanos en la reserva de semillas del suelo: *Erythrophleum suaveolens* (Guill. & Perr.) Brenan, y *Erythrophleum ivorense* A. Chev.**

Las semillas contenidas en la reserva del suelo pueden ser demasiado pequeñas para cuantificarlas visualmente. Habitualmente estas especies se identifican después de intentos de germinación en el vivero, lo que requiere tiempo. Este estudio explora un nuevo enfoque basado en la imagen hiperespectral del infrarrojo cercano (NIR-HSI) combinada con herramientas quimiométricas. Se centra en la reserva de semillas del suelo de los bosques húmedos de África, que todavía es poco conocida. Usamos 83 semillas de dos especies hermanas, *Erythrophleum suaveolens* (Guill. & Perr.) Brenan y *Erythrophleum ivorense* A. Chev., recogidas en la reserva de semillas del suelo (entre 0 y 10 cm de profundidad), en el Gabón, el Camerún y el Congo. Aplicando análisis de componentes principales y análisis discriminante de mínimos cuadrados parciales, estudiamos la capacidad de la imaginería hiperespectral del infrarrojo cercano para identificar las semillas de ambas especies. Este método es rápido, no destructivo y ofrece nuevas perspectivas para estudios de reservas de semillas en suelos forestales.

**Palabras clave**: reserva de semillas del suelo, especies de madera, imágenes hiperespectrales en el infrarrojo cercano, quimiometría, identificación, discriminación.

## Introduction

The soil seed bank designates the stock of viable seeds in the soil (Lipoma et al. 2018). It is considered a significant compartment of the natural regeneration in both temperate and tropical forests (Roberts 1981; Sousa et al. 2017) and represents the memory of plant communities in an ecosystem (Plue et al. 2010; Skowronek et al. 2014). Studies dealing with the soil seed bank could provide information on the structure and functioning of the forest ecosystems (Odum 1969; Hille Ris Lambers et al. 2005). Until recently, most studies on the seed bank of tropical forests were carried out in Asia and America (Martins and Engel 2007; Shen et al. 2014; Sousa et al. 2017). Over the past decade, however, studies in Central Africa have increased in number (Daïnou et al. 2011; Douh et al. 2018a; Zebaze et al. 2021; Padonou et al. 2022; Douh et al. 2023). All of them have been carried out using conventional methods. In order to identify the species present in the soil seed bank, germination tests were performed (Daïnou et al. 2011; Sousa et al. 2017; Douh et al. 2018a). This method is time-consuming, and protocols may vary from one author to another (Plaza-Bonilla et al. 2014). Another difficulty comes from ascertaining the viability of non-germinated seeds at the end of the experiments, especially for the smallest dormant seeds, which are almost unobservable in the soil bank. Consequently, the use of other, more effective techniques would be desirable to avoid underestimation in the quantification of the soil seed bank.

In this context, the use of near-infrared (NIR) spectroscopy technology linked with a microscope (NIRM) to identify seeds present in the soil deserves to be investigated.

In the recent years, new methods based on NIR spectroscopy technology have been developed. Thus, NIR has been linked with a microscope to create the NIR microscopy (NIRM). This spectrometer instrument includes a classical NIR spectrometer coupled with an optical microscope in which the optics have been adapted to NIR radiation.

Near-infrared microscopes allow the spectra to be collected from an extremely small sample (typically, 50 µm × 50 µm or less, depending on the instrument and the configuration) (Yang et al. 2011).

Recent developments in NIR Focal Plane Array (FPA) technology offer a solution to this problem in the form of imaging spectroscopy, which combines the advantages of spectroscopic and microscopic methods, along with much faster sample analysis since the spectral data are acquired in parallel. A NIR hyperspectral imaging spectrometer (NIR-HSI) gathers spectral and spatial data simultaneously by recording sequential images of a pre-defined sample (Fernández Pierna et al. 2004).

For both classical NIR and NIR-HSI, the advantages include simplicity of data acquisition, low cost per analysis, rapid inspection, non-destructive method, and accuracy (Fernández Pierna et al. 2004, 2012; Dale et al. 2012). Nonetheless, in NIR spectroscopy systems, samples usually have to be ground at less than 1 mm. With NIR-HSI systems, sample preparation is not necessary. The samples can be scanned without any grinding and can be subsequently used for other purposes (e.g., for germination trials) (Roggo et al. 2005).

Also, with NIR techniques, one measure gives one average spectrum, while thousands of spectra can be obtained with NIR-HSI, giving a complete picture of the distribution of chemical compounds at the pixel level (Ben-Dor and Banin 1995; Fernández Pierna et al. 2009; Dale et al. 2013; Shahin et al. 2014). Considering the complexity of the spectrum, chemometrics (application of mathematical tools, in particular statistics, to obtain maximum information from chemical data) and multivariate statistical approaches are needed to progress in the exploitation of the data (Massart et al. 1998).

The NIR-HSI has been used to discriminate mung bean seeds into normal and hard groups (Phuangsombut et al. 2018). This technology has also been used to identify the authenticity of maize seed varieties (Cui et al. 2018).

NIR-HSI has been revealed as a promising tool in the discrimination and identification of the quality of cereal grains (Fernández Pierna et al. 2010; Vermeulen et al. 2017; Caporaso et al. 2018). However, to our knowledge, this technology has never been used to discriminate or quantify the seed bank of forest soils.

In this work, a complete procedure based on NIR-HSI coupled with chemometrics has been proposed in order to discriminate the soil seed bank of the two-sister species exhibiting relatively similar morphological characteristics, namely *Erythrophleum suaveolens* (Guill. & Perr.) Brenan and *Erythrophleum ivorense* A. Chev. (Gorel et al. 2015).

## Material and Methods

### Species study and sampling of the soil seed bank

*Erythrophleum suaveolens* and *E. ivorense* are two timber species belonging to the family of Leguminosae-Detarioideae, exploited in central Africa (commonly known as "Tali"). *Erythrophleum suaveolens* is found in semi-deciduous rain tropical forests, forest galleries, and dry forests, while *E. ivorense* is found in evergreen rain forests (Aubréville 1970; Duminil et al. 2010). Both species are parapatric, and their distinction in contact zones is challenging (Gorel et al. 2015; Duminil et al. 2016). They have dormant seeds, and they are found in the soil seed bank at densities reaching 0.15 to 8.55 seed/m² (Douh et al. 2018b).

In this study, seeds of *E. suaveolens* were collected in June 2015 in the forests of Northern Republic of the Congo (Loundoungou) within 4 km²-plots of the DynAfFor[1] project.

[1] *Dynamique des forêts d'Afrique centrale*, http://www.dynafac.org


**Figure 1.**
Location of the collection sites of seeds in Cameroon (Ma'an), in Congo (Loundoungou) and in Gabon (Estuaire).

Seeds of *E. ivorense* were collected in June 2008 and 2015 in the forests of Ma'an community (Cameroon) and of Estuaire province (Gabon) (figure 1). Globally, 83 seeds were collected in the 0-10 cm layer of soil. Table I summarises the characteristics of the seed collection sites (Segalen 1967; Freycon 2014).

Reference numbers have been assigned to each seed (68 seeds of *E. suaveolens* and 15 seeds of *E. ivorense*). Seeds were kept at ambient temperature and humidity for 30 days until posterior analysis using NIR-HSI.

### Near-infrared (NIR) hyperspectral imaging (HSI)

Near-infrared hyperspectral images were acquired with a system combining a NIR hyperspectral line scan instrument and a conveyor belt (BurgerMetrics SIA, Riga, Latvia; photo 1). The camera was a short-wave infrared camera XEVA CL 2.5 320 TE4 (SPECIM Ltd., Oulu, Finland), using an ImSpector N25E spectrograph that includes a cooled, temperature-stabilised mercury-cadmium-telluride detector (XENICS NV, Leuven, Belgium). The images were acquired

**Table I.**
Synthetic description of the collection sites of seeds of *Erythrophleum suaveolens* and *E. ivorense*

| | Cameroon (Ma'an) | Congo (Loundoungou) | Gabon (Estuaire) |
|---|---|---|---|
| **Geographical coordinates** | 02°21' N – 02°35' N<br>08°42' E – 09°42'E | 02°18' N – 02°22'N<br>17°31' E – 17°34'E | 0°23'N – 0°24' N<br>9°15' E – 9°27' E |
| **Altitude (m)** | 600 – 800 | 410 – 460 | 800 – 1,000 |
| **Annual rainfall (mm/an)** | 2,800 | 1,686 | 2,831 |
| **Soil types** | Ferralsols | Acrisols-Arenosols-Gleysols | Acrisols-Arenosols-Gleysols |
| **Geomorphology** | Sandy to sandy-clay | Alluvial deposits in the Cuvette central of the Congo | Shale sandstone-Shale limestone |
| **Forest types** | Evergreen | Semi-deciduous | Evergreen |

in the wavelength range from 1,100 nm to 2,400 nm, with a 6.3-nm spectral resolution (i.e., 209 wavelengths), and a width of 320 pixels using the protocol described by Fernández Pierna et al. (2006, 2012) and Vermeulen et al. (2010). Thirty-two (32) scans per image have been averaged, and each pixel provides an absorbance spectrum at each pixel of the image.

The near-infrared spectra acquired on both faces of each seed (lower and upper) were extracted with HyperSee software (BurgerMetrics SIA, Riga, Latvia), and an average of 1,000 pixels/spectra per seed were selected (figure 2). The faulty spectra (from which no information can be extracted) were removed from all images.

### Analyses and soil seed bank identification

The first step before chemometrics analysis is the building of a spectral library. This is done by selecting representative spectra (the average spectra) of each seed variety from each acquired image. Then, this spectral library was used to assess the capacity of the NIR-HSI to distinguish seeds of *E. suaveolens* and *E. ivorense*. Once the library has been built, different chemometric tools have been used, namely a non-supervised method, principal component analysis (PCA, Legendre and Gallagher 2001; Wise et al. 2006), and a supervised technique, partial least squares discriminant analysis (PLS-DA, Naganathan et al. 2008; Williams et al. 2009; McGoverin et al. 2011). The PCA was used as an exploratory method to investigate the possibility of distinguishing the seeds of the two species on basis of their spectra linked to the chemical differences of the seed's constituents (Janné et al. 2001; Dale et al. 2012; Fernández Pierna et al. 2012).

The PCA loadings obtained from the PCA were investigated to figure out which molecules were responsible for the separation. If any, the seeds of both species (Reeve et al. 1996; Silverstein et al. 2007; ASD Inc. 2005-2013).

Regarding PLS-DA, a first calibration model needs to be built before being validated to check its capacity for discriminating both seed species. For this, 50 and 11 seeds of *E. suaveolens* and *E. ivorense* (photos 2 and 3), respectively, were used for calibration and 18 and 4 seeds (respectively for *E. suaveolens* and *E. ivorense*) for validation (Dale et al. 2013). The accuracy of the PLS-DA classification model was determined in terms of sensitivity and specificity values.

Sensitivity, or true positive rate, is a statistical measure of how the classification model is able to recognise a sample belonging to a given class. And the specificity, or true negative rate, measures the ability of the model to reject all samples not belonging to that given class (Eylenbosch et al. 2017).

All the analyses were performed using the Matlab R2015a software (The MathWorks, Inc., Natick, MA, USA) and the computing environment R (R Core Team, 2013).

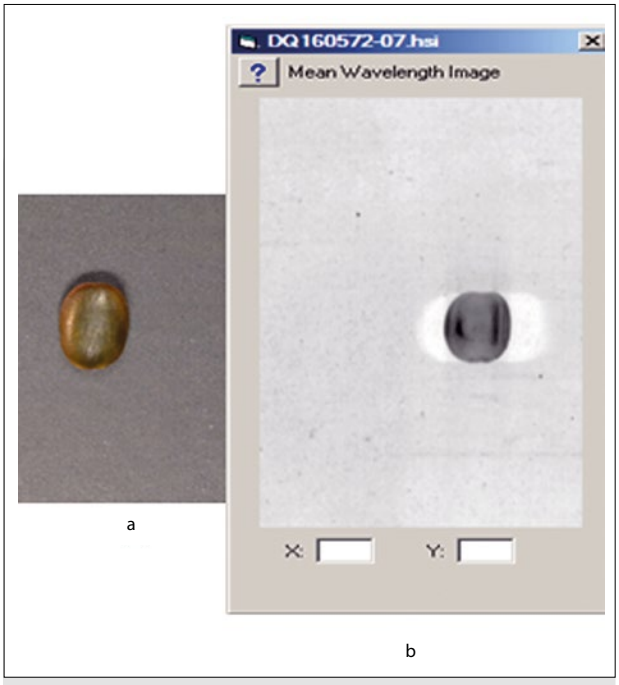


**Figure 2.**
a) Radial Basis Functions (RBF) image of a seed.
b) Near-infrared hyperspectral imaging (NIR-HSI) of the same seed.

Bois et Forêts des Tropiques – ISSN: L-0006-579X
Volume 360 – 2nd quarter – June 2024 – p. 1-10


# Results and discussion

### Spectral signatures of *Erythrophleum suaveolens* and *E. ivorense* seeds

Figure 3 shows the spectra and the raw average spectra NIR of *E. suaveolens* and *E. ivorense* seeds obtained with NIR-HSI. The overall level of absorbance appears to be higher for *E. ivorense* seeds than for *E. suaveolens* seeds.

This trend shows that seeds of *E. suaveolens* would have significant light scattering, which would lead to a high level of reflection and therefore a relatively lower level of absorbance (Dale et al. 2012).

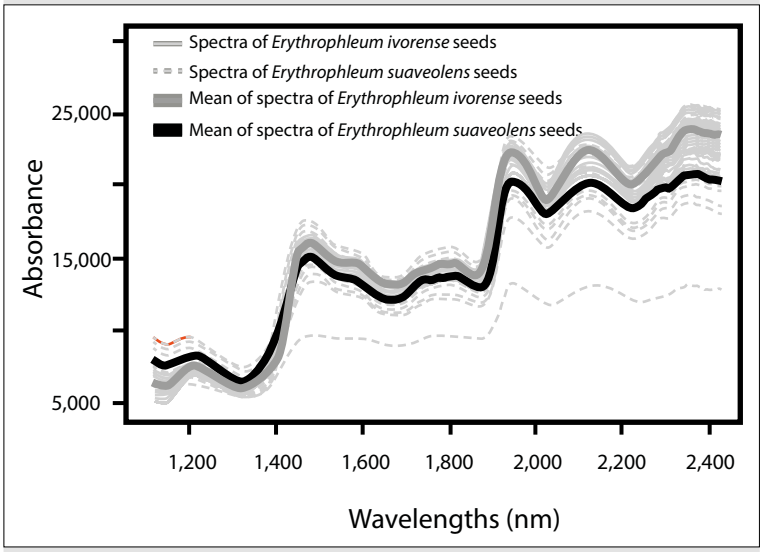


**Figure 3.**
The spectra and the raw average spectra of the 68 seeds of *Erythrophleum suaveolens* and 15 seeds of *E. ivorense*.

### Discrimination of the soil seed bank of two timber species: *Erythrophleum suaveolens* and *E. ivorense*

Overall, a trend of discrimination between the two types of seeds is observed (figure 4). The first two axes of the PCA analysis display 95.01% of the explained variance (76.64% for the PC1 and 18.37% for the PC2). An observation of the PCA loadings (not shown) allowed refusing water absorption bands as responsible for this separation pattern. Using PLS-DA on the calibration set, a specificity and a sensibility of 100% were obtained. On the basis of the confusion matrix (table II), the validation of discrimination of *E. suaveolens* vs. *E. ivorense* displayed a sensibility of 100% and a specificity of 80%.

Consequently, these results confirm the ability of the PLS-DA classification method to discriminate the seeds of the two species.

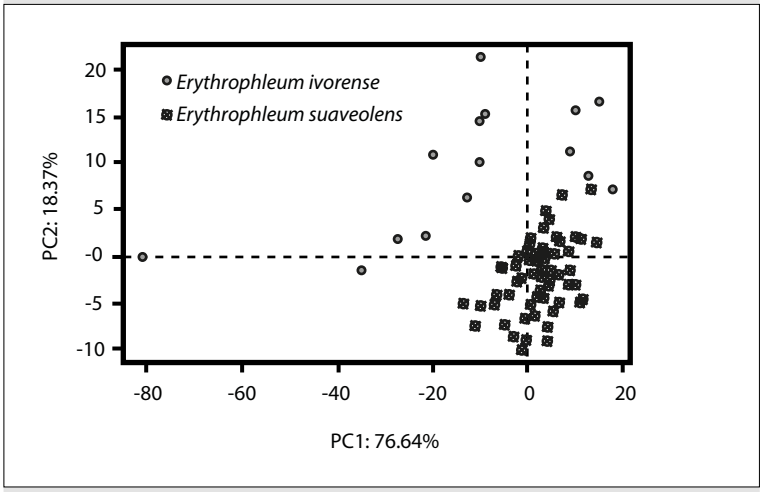


**Figure 4.**
Principal component analysis of seeds of *Erythrophleum suaveolens* and *E. ivorense*.

Compared to *E. ivorense*, PCA loading reveals the presence of amino acids (group N-H) only in *E. suaveolens* seeds (wavelength, 1,500-1,600 nm) (Silverstein et al. 2007; ASD Inc. 2005-2013). Gorel et al. (2015) demonstrated that *E. suaveolens* had larger seeds than *E. ivorense*, and according to Duminil et al. (2016), *E. suaveolens* typically occupies drier climates than *E. ivorense*. Proteins are sources of nitrogen, carbon, and sulphur essential for the future nutritive needs of the embryo (Hacisalihoglu et al. 2010).

A larger amount of proteins in the seeds would ensure its vigour and probably a faster seedling development, which can be an advantage when the rainy season is shorter (Dumas and Rogowsky 2008; Noguero et al. 2011; D'Erfurth et al. 2012).

Conversely, the absence of proteins in the seeds of *E. ivorense* would prevent maintenance and renewal of the embryo tissues, and consequently, the seeds would have a prolonged dormancy compared to the seeds of *E. suaveolens*. This could be an advantage in evergreen forests, which are subject to less seasonality and lower light levels at ground level (Gond et al. 2013). Both *Erythrophleum* species are light-demanding and need a canopy gap to germinate (Duminil et al. 2016). We can therefore assume that this opening must be greater in evergreen forest than in semi-deciduous forest, which requires a longer waiting period in the soil.

**Photo 2.**
*Erythrophleum suaveolens* seed. Photo N. Kayoka Mukendi.

**Photo 3.**
*Erythrophleum ivorense* seed. Photo N. Kayoka Mukendi.

## Conclusion

This study demonstrated that near-infrared (NIR) hyperspectral imaging spectroscopy coupled with chemometrics is an efficient tool to discriminate the seeds of two sister tree species, here between *Erythrophleum suaveolens* (Guill. & Perr.) Brenan and *Erythrophleum ivorense* A. Chev. Seeds (photos 2 to 4), respectively collected in the forests of Loundoungou (Northern Republic of the Congo) for the first, and of Ma'an community (Southern Cameroon) and of Estuaire province (North-Western Gabon) for the second. Respective spectral signatures and the discrimination of the soil seed bank for the seed samples of both species were determined. The difference detected in seed protein content could explain the differences observed between the seeds of each species, but this result should be strengthened in future research using the Kjeldahl method (Kjeldahl 1883; Biancarosa et al. 2017; Sadaiah et al. 2018; Mæhre et al. 2018).

This exploratory research offers new perspectives in qualifying and quantifying the soil seed bank of forests.

But, in order to validate the approach, additional tests should be done on the smallest seeds found in the soil of the rainforests (Douh et al. 2018ab), such as *Musanga cecropioides*, *Nauclea diderrichii*, and *Macaranga* spp., without forgetting the species that are difficult to distinguish.

### 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 article.

### Acknowledgements

The authors thank Marien Ngouabi University (Republic of the Congo), the University of Liège and its faculty Gembloux

**Table II.**
Discriminant Analysis (PLS-DA) confusion matrix of the two types of seeds.

| **Confusion Matrix** | | | | |
|---|---|---|---|---|
| **Class** | **TP** | **FP** | **TN** | **FN** |
| *Erythrophleum ivorense* | 1.00000 | 0.00000 | 1.00000 | 0.00000 |
| *Erythrophleum suaveolens* | 1.00000 | 0.00000 | 1.00000 | 0.00000 |
| **Confusion table** | | | | |
| | | Actual Class | | |
| | | *Erythrophleum ivorense* | *Erythrophleum suaveolens* | |
| Predicted as *E. ivorense* | | 15 | 0 | |
| Predicted as *E. suaveolens* | | 0 | 68 | |
| Key : | | | | |
| *Erythrophleum ivorense* | | | | |
| *Erythrophleum suaveolens* | | | | |
| **CV RESULTS** | | | | |
| **Confusion Matrix (CV)** | | | | |
| **Class** | **TP** | **FP** | **TN** | **FN** |
| *Erythrophleum ivorense* | 1.00000 | 0.00000 | 1.00000 | 0.00000 |
| *Erythrophleum suaveolens* | 1.00000 | 0.00000 | 1.00000 | 0.00000 |
| **Confusion Table (CV)** | | | | |
| | | Actual Class | | |
| | | *Erythrophleum ivorense* | *Erythrophleum suaveolens* | |
| Predicted as *Erythrophleum ivorense* | | 15 | 0 | |
| Predicted as *Erythrophleum suaveolens* | | 0 | 68 | |

TP = True Positive; FP = False Positive; TN = True Negative; FN = False Negative.

Agro-Bio Tech (Belgium), the PEFOGRN-BC projects (support project for the wider programme of training in natural resources management in the Congo Basin), the DynAfFor project (*Dynamique des forêts d'Afrique centrale*) financed by the FFEM (*Fonds français pour l'environnement mondial*) and the AFD (*Agence française de d*éveloppement), led by CIRAD, Gembloux Agro-Bio Tech and Nature+, as well as the logging companies Mokabi-Dzanga (*Groupe Rougier*) and CIB/OLAM for their financial, scientific and technical support, and for their aid with fieldwork. Finally, we would like to thank Dr. Dakis Yaoba Ouedraogo, Dr. Barbara Haurez, Dr. Jean-François Gillet, Dr. Armel Loïc Donkpêgan Sègbédji, Jean-Yves De Vleeschouwer, Fabrice Rapezant, Gilbert Nsongola, Élodie Alberny, Mercier Mayinga, Lambert Imbalo, Angélique Ouadiabantou, Yves Crépin Nzihou, Isaac Dzombo, and Maturin Mazengué for their various contributions and their advice.

**Funding**

This study was carried out in the framework of the DynAfFor project (*Dynamique des forêts d'Afrique centrale*) financed by the FFEM (*Fonds français pour l'environnement mondial*) and the AFD (*Agence française de d*éveloppement).

**Conditions of access to data**

The data can be accessed using the Zenodo open digital repository link: https://doi.org/10.5281/zenodo.13908452
You are invited to cite this dataset and the article when using them and to inform the authors:
1- Chauvelin DOUH (https://orcid.org/0000-0003-0656-9831)
2- Antoine DERYCK (https://orcid.org/0009-0005-2461-1011)

**To cite the dataset:**

DOUH, C. (2024). Exploring the potential of Near Infrared Hyperspectral Imaging and chemometrics to discriminate soil seed bank of two timber species central African : *Erythrophleum suaveolens* (Guill. & Perr.) Brenan, and *Erythrophleum ivorense* A. Chev. [Data set]. In Exploring the potential of Near Infrared Hyperspectral Imaging and chemometrics to discriminate soil seed bank of two timber species central African : *Erythrophleum suaveolens* (Guill. & Perr.) Brenan, and *Erythrophleum ivorense* A. Chev. (Vol. 360, Numéro 2, p. 1-10). Zenodo. https://doi.org/10.5281/zenodo.13908452

**Photo 4.**
Tree trunk of *Erythrophleum* sp.
Photo J.-L Doucet.

## References

Aubréville A., 1970. Légumineuses-Césalpinioidés (Leguminosae-Caesalpinioideae). Flore du Cameroun. Volume 9. Muséum National d'histoire Naturelle, Paris, France, 339 p.

Biancarosa I., Espe M., Bruckner C. G., Heesch S., Liland N., Waagbo R., *et al.*, 2017. Amino acid composition, protein content, and nitrogen-to-protein conversion factors of 21 seaweed species from Norwegian waters. Journal of Applied Phycology, 29: 1001-1009. https://doi.org/10.1007/s10811-016-0984-3

Ben-Dor E., Banin A., 1995. Near infrared analysis as a rapid method to simultaneously evaluate several soil properties. Soil Science Society of America Journal, 59: 364-372. https://doi.org/10.2136/sssaj1995.03615995005900020014x

Caporaso N., Whitworth M. B., Fisk I. D., 2018. Near-Infrared spectroscopy and hyperspectral imaging for non-destructive quality assessment of cereal grains. Applied Spectroscopy Reviews, 1-21. https://doi.org/10.1080/05704928.2018.1425214

Cui Y., Xu L., An D., Liu Z., Gu J., Li S., *et al.*, 2018. Identification of maize seed varieties based on near infrared reflectance spectroscopy and chemometrics. International Journal of Agricultural and Biological Engineering, 11: 177-183. https://doi.org/10.25165/j.ijabe.20181102.2815

Daïnou K., Bauduin A., Bourland N., Gillet J. F., Fétéké F., Doucet J.-L., 2011. Soil seed bank characteristics in Cameroonian rainforests and implications for post-logging forest recovery. Ecological Engineering, 37: 1499-1506. https://doi.org/10.1016/j.ecoleng.2011.05.004

Dale L. M., Thewis A., Boudry C., Rotar I., Dardenne P., Baeten V., *et al.*, 2013. Hyperspectral imaging applications in agriculture and agrofood product quality and safety control: a review. Applied Spectroscopy Reviews, 48: 142-159. https://doi.org/10.1080/05704928.2012.705800

Dale L. M., Thewis A., Rotar I., Fernández Piema J. A., Boudry C., Vidican R. M., *et al.*, 2012. Chemometric tools for NIRS and NIR hyperspectral imaging. Bulletin of Usavm Cluj-Napoca Agriculture, 69: 70-76. http://hdl.handle.net/2268/132855

D'Erfurth I., Le Signor C., Aubert G., Sanchez M., Vernoud V., Darchy B., *et al.*, 2012. A role for an endosperm-localized subtilase in the control of seed size in legumes. New Phytologist, 196: 738-751. https://doi.org/10.1111/j.1469-8137.2012.04296.x

Douh C., Daïnou K., Loumeto J. J., Moutsambote J.-M., Fayolle A., Tosso F., *et al.*, 2018a. Soil seed bank characteristics in two central African forest types and implications for forest restoration. Forest Ecology and Management, 409: 766-776. https://doi.org/10.1016/j.foreco.2017.12.012

Douh C., Gorel A. P., Daïnou K., Fonteyn D., Bustillo E., Obsomer L., *et al.*, 2018b. Banque de graines du sol et déterminants de la germination du tali, *Erythrophleum suaveolens* (Guill. & Perr.) Brenan. Bois et Forêts des Tropiques, 338 : 43-55. https://doi.org/10.19182/bft2018.338.a31681

Douh C., Mabengo B. C., Nzila J. D., Malonga Mbouchi L., Nsongola G., Loumeto J. J., *et al.*, 2023. Soil Seed Bank Characteristics in Congolese Rainforests and Implications for Post-Logging Plots Reforestation. Open Journal of Forestry, 13: 294-314. https://doi.org/10.4236/ojf.2023.133018

Dumas C., Rogowsky P., 2008. Fertilization and early seed formation. Académie des Sciences Institut de France, Comptes Rendus Biologies, 331: 715-725. https://doi.org/10.1016/j.crvi.2008.07.013

Duminil J., Daïnou K., Kombi Kaviriri D., Gillet P., Loo J., Doucet J.-L., *et al.*, 2016. Relationships between population density, fine-scale genetic structure, mating system and pollen dispersal in a timber tree from African rainforests. Heredity (Edinb), 116: 295-303. https://doi.org/10.1038/hdy.2015.101

Duminil J., Heuertz M., Doucet J.-L., Bourland N., Cruaud C., Gavory F., *et al.*, 2010. CpDNA-based species identification and phylogeography: application to African tropical tree species. Molecular Ecology, 19: 5469-5483. https://doi.org/10.1111/j.1365-294X.2010.04917.x

Eylenbosch D., Bodson B., Baeten V., Fernández Pierna J. A., 2017. NIR hyperspectral imaging spectroscopy and chemometrics for the discrimination of roots and crop residues extracted from soil samples. Journal of Chemometrics, 32: e2982. https://doi.org/10.1002/cem.2982

Fernández Pierna J. A., Vermeulen P., Amand O., Tossens A., Dardenne P., Baeten V., 2012. NIR Hyperspectral Imaging spectroscopy and chemometrics for the detection of undesirable substances in food and feed. Chemometrics and intelligent laboratory systems, 117: 233-239. https://doi.org/10.1016/j.chemolab.2012.02.004

Fernández Pierna J. A., Vermeulen P., Stilmant D., Dupuis B., Dardenne P., Baeten V., 2010. Characterisation of fonio millet by near infrared hyperspectral imaging, Proceedings of the 14th International Conference on Near Infrared Spectroscopy (ICNIRS): Breaking the Dawn. Bangkok, Thailand, 997-1002.

Fernández Pierna J. A., Baeten V., Dardenne P., Dubois J., Lewis E. N., Burger J., 2009. Spectroscopic Imaging. In: Comprehensive Chemometrics – 1st Edition. Brown S., Tauler R., Walczak B. (eds). Elsevier, 4: 173-196. https://doi.org/10.1016/B978-044452701-1.00004-1

Fernández Pierna J. A., Baeten V., Dardenne P., 2006. Screening of compound feeds using NIR hyperspectral data. Chemometrics and intelligent laboratory systems, 84: 114-118. https://doi.org/10.1016/j.chemolab.2006.03.012

Fernández Pierna J. A., Baeten V., Michotte Renier A., Cogdill R. P., Dardenne P., 2004. Combination of support vector machines (SVM) and near-infrared (NIR) imaging spectroscopy for the detection of meat and bone meal (MBM) in compound feeds. Journal of Chemometrics, 18: 341-349. https://doi.org/10.1002/cem.877

Freycon V., 2014. Caractérisation des sols de Loundoungou et de Mokabi (Congo). Rapport de mission DynAfFor, 2 au 24 avril 2014. CIRAD, Montpelier. https://agritrop.cirad.fr/573670/

Gond V., Fayolle A., Pennec A., Cornu G., Mayaux P., Camberlin P., et al., 2013. Vegetation structure and greenness in Central Africa from Modis multi-temporal data. Philosophical Transactions of the Royal Society: Biological Sciences, 368: 20120309. http://dx.doi.org/10.1098/rstb.2012.0309

Gorel A. P., Fayolle A., Doucet J.-L., 2015. Écologie et gestion des espèces multi-usages du genre *Erythrophleum* (Fabaceae-Caesapinioideae) en Afrique (synthèse bibliographique). Biotechnologie, Agronomie, Société et Environnement, 19 : 415-429. https://popups.uliege.be/1780-4507/index.php?id=12455

Hacisalihoglu G., Larbi B., Settles A., 2010. Near-Infrared reflectance spectroscopy predicts protein, starch, and seed weight in intact seeds of Common Bean (*Phaseolus vulgaris* L.). Journal of Agricultural and Food Chemistry, 58: 702-706. https://doi.org/10.1021/jf9019294

Hille Ris Lambers J., Clark J. S., Lavine M., 2005. Implications of seed banking for recruitment of southern Appalachian woody species. Ecology, 86: 85-95. https://doi.org/10.1890/03-0685

Janné K., Pettersen J., Lindberg N. O., Lundstedt T., 2001. Hierarchical principal component analysis (PCA) and projection to latent structure (PLS) technique on spectroscopic data as a data pre-treatment for calibration. Journal of Chemometrics, 15: 203-213. https://doi.org/10.1002/cem.677

Kjeldahl J., 1883. Neue Methode zur Bestimmung des Stickstoffs in organischen Körpern. Zeitschrift fur Anal Chemie, 22: 366-382. https://doi.org/10.1007/BF01338151

Legendre P., Gallagher E. D., 2001. Ecologically meaningful transformations for ordination of species data. Oecologia, 129: 271-280. https://doi.org/10.1007/s004420100716

Lipoma M. L., Funes G., Díaz S., 2018. Fire effects on the soil seed bank and post-fire resilience of a semi-arid shrubland in central Argentina. Austral Ecology, 43: 46-55. https://doi.org/10.1111/aec.12533

Martins A. M., Engel V. L., 2007. Soil seed banks in tropical forest fragments with different disturbance histories in south-eastern Brazil. Ecological Engineering, 31: 165-174. https://doi.org/10.1016/j.ecoleng.2007.05.008

Massart D. L., Vandeginste B. G. M., Buydens L. M. C., De Jong S., Lewi P. J., Smeyers-Verbeke J., 1998. Handbook of chemometrics and qualimetrics, Part A. NIR Publications, West Sussex, United Kingdom, 867 p. https://www.sciencedirect.com/bookseries/data-handling-in-science-and-technology/vol/20/part/PA

McGoverin C. M., Engelbrecht P., Geladi P., Manley M., 2011. Characterisation of nonviable whole barley, wheat and sorghum grains using near-infrared hyperspectral data and chemometrics. Analytical Bioanalytical. Chemistry, 401: 2283-2289. https://doi.org/10.1007/s00216-011-5291-x

Mæhre H. K., Dalheim L., Edvinsen G. K., Elvevoll E. O., Jensen I.-J., 2018. Protein determination-method matters. Foods, 7: 5. https://doi.org/10.3390/foods7010005

Naganathan G. K., Grimes L. M., Subbiah J., Calkins C. R., Samal A., Meyer G. E., 2008. Visible/near-infrared hyperspectral imaging for beef tenderness prediction. Computers and Electronics in Agriculture, 64: 225-233. https://doi.org/10.1016/j.compag.2008.05.020

Noguero M., Le Signor C., Verdier J., Aubert G., Udvardi M. K., Buitink J., *et al.*, 2011. Rôle de l'albumen dans le développement de la graine chez les légumineuses. *In* : Graine 2011 (p. S4). Colloque national du Réseau français de Biologie des Graines, Nantes, France, 27-28 octobre 2011. https://prodinra.inra.fr/record/48575

Odum E. P., 1969. The strategy of ecosystem development. Science, 164: 262-270.

Padonou E. A., Akakpo B. A., Tchigossou B., Djossa B., 2022. Methods of soil seed bank estimation: a literature review proposing further work in Africa. iForest, 15: 121-127. https://doi.org/10.3832/ifor3850-015

Phuangsombut K., Ma T., Inagaki T., Tsuchikawa S., Terdwongworakul A., 2018. Near-infrared hyperspectral imaging for classification of mung bean seeds. International Journal of Food Properties, 21: 799-807. https://doi.org/10.1080/10942912.2018.1476378

Plaza-Bonilla D., Álvaro-Fuentes J., Hansen N. C., Lampurlanés J., Cantero-Martínez C., 2014. Winter cereal root growth and aboveground-belowground biomass ratios as affected by site and tillage system in dryland Mediterranean conditions. Plant and soil, 374: 925-939. https://doi.org/10.1007/s11104-013-1926-3

Plue J., Verheyen K., Calster H. V., Marage D., Thompson K., Kalamees R., *et al.*, 2010. Seed banks of temperate deciduous forests during secondary succession. Journal of Vegetation Science, 21: 965-978. https://doi.org/10.1111/j.1654-1103.2010.01203.x

R Core Team, 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.r-project.org/

Reeve A. S., Siegel D., Glaser P. H., 1996. Geochemical controls on peatland pore water from the Hudson Bay lowland: a multivariate statistical approach. Journal of Hydrology, 181: 285-304. https://doi.org/10.1016/0022-1694(95)02900-1

Roberts H. A., 1981. Seed banks in soils. In: Advances in Applied Biology. Coaker T. H. (ed.). Academic Press, 6: 1-55.

Roggo Y., Edmond A., Chalus P., Ulmschneider M., 2005. Infrared Hyperspectral Imaging for qualitative analysis of pharmaceutical solid forms. Analytica Chimica Acta, 535: 79-87. https://doi.org/10.1016/j.aca.2004.12.037

Sadaiah K., Veronica N., Nagendra V., Niharika G., Neeraja C. N., Surekha K., *et al.*, 2018. Methods of Protein Estimation and the Influence of Heat Stress on Rice Grain Protein. International Journal of Pure & Applied Bioscience, 6: 159-168. http://doi.org/10.18782/2320-7051.5733

Segalen P., 1967. Les sols et la géomorphologie du Cameroun. Cahier de l'ORSTOM, Série Pédologie, 5 (2) : 137-188.

Shahin M. A., Symons S. J., Hatcher D. W., 2014. Quantification of Mildew Damage in Soft Red Winter Wheat Based on Spectral Characteristics of Bulk Samples: A Comparison of Visible-Near-Infrared Imaging and Near-Infrared Spectroscopy. Food and Bioprocess Technology, 7: 224-234. https://doi.org/10.1007/s11947-012-1046-8

Shen Y. X., Liu W. L., Li Y. H., Guan H. L., 2014. Large sample area and size are needed for forest soil seed bank studies to ensure low discrepancy with standing vegetation. PLoS ONE, 9 (8): e105235. https://doi.org/10.1371/journal.pone.0105235

Silverstein R. M., Webster F. X., Kiemle D. J., 2007. Spectrometric identification of organic compounds. Seventh edition. State University of New York. College of Environmental Science & Forestry, 502 p. https://doi.org/10.1021/ed039p546

Skowronek S., Terwei A., Zerbe S., Mölder I., Annighöfer P., Kawaletz H., *et al.*, 2014. Regeneration potential of floodplain forests under the influence of non-native tree species: soil seed bank analysis in Northern Italy. Restoration Ecology, 22: 22-30. https://doi.org/10.1111/rec.12027

Sousa T. R., Costa F. R. C., Bentos T. V., Leal Filho N., Mesquita R. C. G., Ribeiro I. O., 2017. The effect of forest fragmentation on the soil seed bank of Central Amazonia. Forest Ecology and Management, 393: 105-112. https://doi.org/10.1016/j.foreco.2017.03.020

Vermeulen P., Ebene M. B., Orlando B., Fernández Pierna J. A., Baeten V., 2017. Online detection and quantification of particles of ergot bodies in cereal flour using near infrared hyperspectral imaging. Food Additives & Contaminants: Part A, 34 (8): 1312-1319. https://doi.org/10.1080/19440049.2017.1336798

Vermeulen P., Fernández Pierna J. A., Abbas O., Dardenne P., Baeten V., 2010. Authentication and traceability of agricultural and food products using vibrational spectroscopy. In: Applications of vibrational spectroscopy in food science. Li-Chan E.C.Y., Griffiths P.R., Chalmers J. M. (Eds.). United Kingdom, John Wiley & Sons, 609-630. https://doi.org/10.1002/0470027320.s8969

Williams P., Geladi P., Fox G., Manley M., 2009. Maize kernel hardness classification by Near Infrared (NIR) Hyperspectral Imaging and multivariate data analysis. Analytica Chimica Acta, 653: 121-130. https://doi.org/10.1016/j.aca.2009.09.005

Wise B. M., Shaver J. M., Gallagher N. B., Windig W., Bro R., Koch R. S., 2006. PLSJToolbox Version 4.0 for use with Matlab™. USA, Eigenvector Research Inc., 420 p.

Yang Z., Hana L., Fernandez Pierna J. A., Dardenne P., Baeten V., 2011. The potential of near infrared microscopy to detect, identify and quantify processed animal by-products. Journal of Near Infrared Spectroscopy, 19: 211-231. https://doi.org/10.1255/jnirs.935

Zebaze D., Fayolle A., Daïnou K., Libalah M., Droissart V., Sonké B., *et al.*, 2021. Land Use Has Little Influence on the Soil Seed Bank in a Central African Moist Forest. Biotropica, 54: 100-112. https://doi.org/10.1111/btp.13032

**Douh et al. – Credit authorship contribution statement**

| Contributor role | Contributor names |
|---|---|
| Conceptualization | C. Douh, J.-L Doucet |
| Formal Analysis | C. Douh, F. Tosso,K. Daïnou, J. A. Fernández Pierna |
| Investigation | C. Douh |
| Project Administration | J.-L Doucet |
| Resources | C. Douh, J.-L Doucet, J. A. Fernández Pierna |
| Supervision | J.-L Doucet, J. A. Fernández Pierna |
| Validation | C. Douh, J.-L Doucet, A. Deryck, J. A. Fernández Pierna |
| Visualization | C. Douh, J.-L Doucet |
| Writing – Original Draft Preparation | C. Douh, F. Tosso, J.-L Doucet |
| Writing – Review & Editing | C. Douh, J.-L Doucet, A.-P. Gorel, J. A. Fernández Pierna |

Bois et Forêts des Tropiques - Revue scientifique du Cirad -


Cirad - Campus international de Baillarguet,
34398 Montpellier Cedex 5, France
Contact : bft@cirad.fr - ISSN : L-0006-579X