



26^{èmes} rencontres HélioSPIR
Montpellier, 24-25 juin 2025

Résumés des communications





HélioSPIR est l'association francophone dédiée à la spectrométrie dans le proche infrarouge.

HélioSPIR a vocation à fédérer les scientifiques et les utilisateurs de la technologie SPIR au sein d'un réseau et à promouvoir l'utilisation de la spectroscopie proche infrarouge. Fondée en 2004 autour de la communauté scientifique d'Agropolis à Montpellier, l'association dépasse maintenant les contours de la région Occitanie et de l'hexagone. C'est un pôle de compétences à dimension internationale dans le domaine de la spectroscopie proche infrarouge.

HélioSPIR organise chaque année une ou deux sessions de rencontres scientifiques. C'est un moment privilégié d'échanges autour de diverses thématiques autour de la spectroscopie proche infrarouge et de découverte des derniers travaux de la communauté. C'est également l'occasion de découvrir ou redécouvrir les équipements de spectroscopie et d'imagerie hyperspectrales des principaux fabricants du secteur.

Président : G. Chaix ; adjoint : J.-M. Roger

Secrétaire : S. Beaumont ; adjointe : A. Cambou

Trésorier : C. Fontange ; adjoint : R. Cinier

Conseil d'administration : V. Baeten, D. Bastianelli, S. Beaumont, A. Cambou, G. Chaix, R. Cinier, M. Ecarnot, C. Fontange, M. Loudiyi, S. Mas-Garcia, T. Ricour, J.M. Roger, V. Rossard, S. Roussel, E. Ziemons

Comment citer ce document

HélioSPIR, 2025. Résumés des communications présentées aux 26èmes rencontres HélioSPIR, Montpellier (France), 24-25 juin 2025. D. Bastianelli, G. Chaix, Eds. Association HélioSPIR, Montpellier (France), 53p. DOI : 10.19182/agritrop/00240

Comment citer un résumé particulier

Auteur 1, Auteur 2... Auteur n, 2025. Titre du résumé. *In* : HélioSPIR, 2025. Résumés des communications présentées aux 26èmes rencontres HélioSPIR, Montpellier (France), 24-25 juin 2024. D Bastianelli, G Chaix, Eds. (DOI : 10.19182/agritrop/00240), Association HélioSPIR, Montpellier (France). Numéros de page.



Publié sous licence [Creative Commons CC-BY](https://creativecommons.org/licenses/by/4.0/)

Communications orales **Pages**

Mardi 24 juin 2025

| | |
|--|----|
| M. Lesnoff. Weighted PLSR-DA with unbalanced data – Study case of blood BHB predicted from milk MIR spectra for ketosis detection in dairy herds | 5 |
| O. Aoultit et al. Caractérisation des sols par SPIR : connaissance des erreurs de mesures et qualité de la prédiction des concentrations en carbone organique et en azote total | 9 |
| M. Brandolini-Bunlon et B. Jaillais. Apport de la SO-PLS à l'approche multiblocs exploratoire | 11 |
| A. Niemöller et al. Are you prepared? Rediscovering wavelets and a new local regression approach | 12 |
| A Malechaux et al. Machine learning interpretability methods applied to calibration models developed on near infrared spectroscopic data | 13 |
| P-Y. Werrie et al. Tri optique infrarouge grain à grain comme outil de phénotypage haut débit technologique et sanitaire | 15 |

Mercredi 25 juin 2025

| | |
|---|----|
| R. Nikzad-Langerodi. Near Infrared Spectroscopy @KremsChem | 16 |
| M. Watiez et Z Fayeulle. Applications industrielles des NIR portatifs | 17 |
| S. Pinault et al. A physics-based data simulation and augmentation method for enhancing calibration model development on a visible spectrometer | 18 |
| F. Abdelghafour et M. Metz. Évaluation de méthodes d'apprentissage profond pour les données spectroscopiques et la chimométrie : un cas d'étude pour les applications agro-environnementales | 19 |
| V. Larat. Développement et intégration à une plateforme utilisée en milieu industriel d'un modèle NIR de classification des matières premières d'origine agricole | 21 |
| C. Biennu et al. La sélection phénotypique : Quand les généticiens s'essaient à la NIRS | 22 |
| O. Parthenay et al. Utilisation de la spectroscopie NIR pour évaluer la qualité des graines de hêtre selon leur environnement | 24 |
| A. Barreau et al. Authentification et prédiction de paramètres physico-chimiques dans la cannelle | 25 |
| I. Tumoine et al. Comparaison de l'augmentation de données et du prétraitement pour l'apprentissage profond | 26 |
| E. Deschamps et al. Caractérisation de la diversité des ressources pastorales consommées par des petits ruminants sur parcours méditerranéens : apports de la spectroscopie infrarouge | 28 |
| M. Amrani Idrissi. Détection d'anomalies dans les images hyperspectrales de feuilles de blé en utilisant Matrix Profile | 31 |
| A.M. Dieng et al. Suivi en temps réel de l'évolution de la qualité organoleptique des pommes au cours d'un procédé thermomécanique par spectroscopie proche infrarouge | 32 |
| J. Do Souto Soeiro et al. Prédiction du potentiel de minéralisation du carbone au sol (C_{min91}) sur des digestats de méthanisation grâce à la spectroscopie proche infra-rouge | 34 |

| | |
|--|----|
| R. Ouhammadou et al. Authentification des mélanges farineux blé-légumineuses par fusion multiblocs des spectres proche et moyen infrarouge | 36 |
| I. Abdelmeguid et al. Real-Time Assessment of PEG-Induced Hydric Stress in Sunflowers Using Spectral Data and ASCA | 37 |
| K. Boughalmi et al. Potentiel et limites de l'utilisation de la SPIR foliaire pour l'identification taxonomique de spécimens d'herbier : étude de cas sur les <i>Annonaceae</i> (résumé non disponible) | nd |
| G. Chaix. La spectroscopie proche infrarouge et la biodiversité : illustrations pour l'aide à l'identification taxonomique | 40 |
| N. Belmokhtar et al. La Spectroscopie proche infrarouge au service de la détermination des essences forestières | 43 |
| T.L.P. Couvreur et al. Perspectives sur l'évolution des données de réflectance foliaire à partir de spécimens d'herbier dans une grande famille de plantes tropicales (<i>Annonaceae</i>) (résumé non disponible) | nd |
| S. Bazan et al. Utiliser la SPIR sur des échantillons d'herbiers pour l'identification taxonomique ? | 45 |

Posters

Pages

| | |
|---|----|
| S. Bazan et al. A new light in Herbaria: How can NIRS be applied to Herbarium specimens for identification? | 47 |
| J. Boyer et al. Machine learning methods for sugar quantification in grapes based on near-infrared hyperspectral imaging | 48 |
| M. Lambert et al. Étude de la Diversité Génétique et de la Phénologie du Ray-Grass Anglais (<i>Lolium perenne</i> L.) par Spectroscopie Proche Infrarouge | 50 |

Communications orales

Weighted PLSR-DA with unbalanced data – Study case of blood BHB predicted from milk MIR spectra for ketosis detection in dairy herds

¹ **Matthieu Lesnoff**

¹ CIRAD/UMR SELMET, 34398 Montpellier – France

matthieu.lesnoff@cirad.fr

Keywords : Discrimination, PLSR-DA, Unbalanced data, Jchemo

Context

PLSR-DA is very popular within the “PLSDA” methods. It considers a discrete variable \mathbf{y} (class membership) to be predicted from variables \mathbf{X} . A dummy table ($\mathbf{Y}_{\text{dummy}}$) is built from \mathbf{y} , consisting in one binary column for each class (0: the observation belongs to the class, 1: the observation does not belong to the class) and PLS2 scores (\mathbf{T}) are computed from the data $\{\mathbf{X}, \mathbf{Y}_{\text{dummy}}\}$. Then, an MLR is run on the data $\{\mathbf{T}, \mathbf{Y}_{\text{dummy}}\}$, giving predictions $\hat{\mathbf{Y}}_{\text{dummy}}$. For a given observation, the final predicted class corresponds to the $\hat{\mathbf{Y}}_{\text{dummy}}$ column showing the maximal value.

PLSR-DA has the advantage to be simple and fast even on large datasets (actually this is a PLSR2 algorithm). Nevertheless, a drawback is that it can lose performance when the number of classes is large (e.g. > 5 classes). This is due to possible masking effects between classes in the PLS2 score space (Hastie et al. 2009). Another drawback is its high sensitivity to unbalanced classes, generating biased predictions. For instance, when \mathbf{y} contains a predominant class, the method will favor systematically the prediction of this class. This presentation tackles this last point.

A common approach to manage unbalanced classes is to subsample the training data $\{\mathbf{X}, \mathbf{y}\}$ to equalize the class importance. For instance, if a two-class training dataset (classes A and B) consists in $n_A = 1000$ obs. and $n_B = 200$ obs. ($n = n_A + n_B$), the approach will be to sample $m_A = 200$ obs. of class A within the 1000 observations available, and to calibrate the model on the $m_A + n_B = 400$ obs. This approach can be too aggressive for the training set, especially in presence of rare classes (e.g. for classes A and B above, if $n_B = 10$ obs., the training set would decrease from 1010 obs. to 20 observations!). More generally, the general limitation is that the model does not use all the available information for the model calibration.

Theory

This presentation describes an alternative simple approach, based on weighted PLS (WPLS). In WPLS, a vector of observation weights $\mathbf{w} = (w_1, \dots, w_n)$ is embedded in the PLS equations, in the scores and regression computations, as well as in the preliminary data centering and eventual scaling (e.g. Schaal et al. 2002). WPLS is a versatile method that can be used to build robust approaches, locally weighted approaches, etc. In the PLSR-DA context, the principle is to define \mathbf{w} such as the sums of the weights by class are equal.

Following the two-class example above, and noting p_A and p_B the sum of the weights w_i of the observations in class A and B, respectively, vector \mathbf{w} would be defined such as $p_A = p_B = 1/2$.

Material and methods

WPLSR-DA is demonstrated on a dataset collected by a European research consortium (see the Acknowledgments section) targeting the ketosis disease in cattle dairy herds. This disease affects milking cows especially in early lactation and produce important economic losses. The beta-hydroxybutyrate (BHB) in blood has been identified as a successful biomarker of ketosis but its measure implies blood sampling. One of the research objectives of the consortium is to assess the interest of the MIR spectrometry from milk samples, because they are routinely available, to detect high BHB concentration in the blood (disease symptom). MIR detection from milk samples would be an interesting alternative to the usual blood tests that are costly, invasive and induces pain to the cows.

The available dataset for this presentation (\mathbf{X} = milk MIR spectra, \mathbf{y} = BHB level with two classes: ≤ 1.2 mmol/L \Rightarrow *low*, > 1.2 mmol/L \Rightarrow *high*) was split to a training set ($n = 3,195$ obs.) and a test set ($n = 1,025$ obs.) constituted by other herds (and cows) than in training set. Classes were highly unbalanced both in the training and test sets (Tab1), with almost 92% of the observations belonging to class *low*.

Table 1: Nb. samples in the training a test sets.

| BHB level | Training | Test |
|-----------|----------|------|
| Low | 2,929 | 952 |
| High | 266 | 73 |

Milk samples (the observations) have been analyzed with 34 spectrometers, in particular FT2, FT6000, FT+, and FT7 (Foss, Hillerød, Denmark), and standard lactoscopes FT-MIR automatic (Delta Instruments, Drachten, the Netherlands). The MIR spectra from the different instruments were standardized to be merged into a common dataset. The selected spectral area consisted of 212 wavenumbers from 968.1–1,577.5 cm^{-1} , 1,731.8–1,762.6 cm^{-1} , 1,781.9–1,808.9 cm^{-1} and 2,831.0–2,966.0 cm^{-1} to exclude areas not reproducible between instruments. The spectra were pretreated by a second derivative.

A replicated K-fold cross-validation (CV; $K = 3$ and nb. rep. = 30) was implemented on the naïve and weighted PLSR-DA models. Then, the selected models were used to predict the test set and compute an estimate of generalization error. The mean within-class error rate (MERRP) was preferred to the usual global error rate to calibrate the models since classes were unbalanced in the validation folds of the CV.

Results

CV results (Fig1) clearly show that the naïve PLSR-DA fails to discriminate the two classes (MERRP > 40%). In contrast, a more relevant CV curve was obtained from WPLSR-DA, reaching final MERPPs lower than 20%.

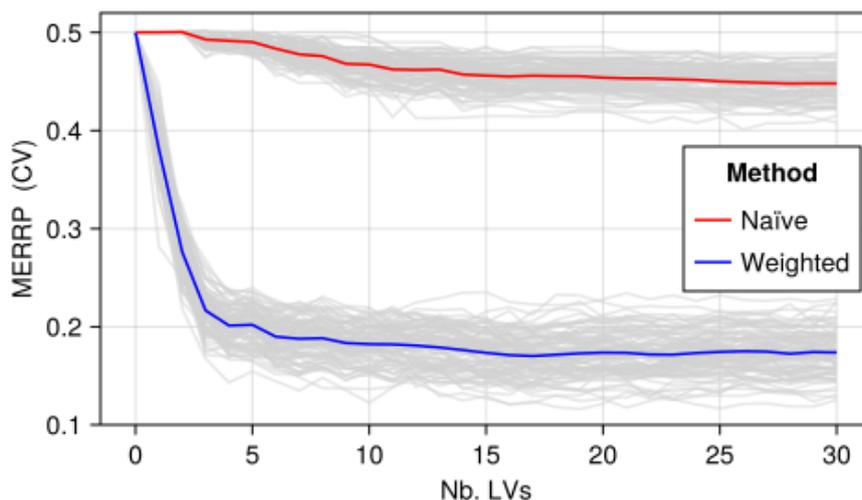


Fig.1: MERRP results for the repeated K-fold CV (fold \times replication curves are in grey).

Predictions on the test set (models with 17 LVs) are presented in Tab.2. As expected, the naïve PLSR-DA favors systematically the predominant class (100% and 93% of samples *low* and *high*, respectively, are predicted as *low*) and cannot be used as predictive method. W-PLSRDA improves a lot the predictions, with only 12% of false positives and 16% of false negatives. These performances seem acceptable as a first detection filter for farmers, and encourage future developments of the MIR spectrometry as an operational methodology to detect ketosis from milk sampling.

Tab2: Confusion matrix (row %) for the predictions of the BHB level on the test set (n = 1,025).

| True values | (a) Naïve | | (b) Weighted | |
|-------------|-------------|----------|--------------|-----------|
| | Predictions | | Predictions | |
| | Low | High | Low | High |
| Low | 100 | <u>0</u> | 88 | <u>12</u> |
| High | <u>93</u> | 7 | <u>16</u> | 84 |

Software

WPLSR-DA can easily be implemented on data, either in Julia using the *Jchemo* package (Lesnoff 2021) or in R using the *rchemo* package (Brandolini et al. 2024).

- *Jchemo*: WPLSR-DA is implemented by default in function `plsrd`. The syntax is very simple:

```
model = plsrd(nlv = 20)           # defines the model
fit!(model, X, y)                 # fits the model on the data
pred = predict(model, Xnew).pred  # makes and store the predictions
```

The naïve un-weighted model can be fitted as follows:

```
model = plsrd(nlv = 20, prior = :prop)
fit!(model, X, y)
```

- *rchemo*: A weight vector has to be specified before the model fitting

```
weights <- [...]                 # a vector of n weights to build
fm <- plsrd(X, y, weights, nlv = 20) # defines/fits the model
pred <- predict(model, Xnew)$pred  # predictions
```

References

Brandolini-Bunlon, M., Jallais, B., Roger, J.M., Lesnoff, M. 2023. rchemo: Dimension reduction, regression and discrimination for chemometrics. <https://cran.r-project.org/web/packages/rchemo>

Hastie, T., Tibshirani, R., Friedman, J., 2009. The elements of statistical learning: data mining, inference, and prediction, 2nd ed. Springer, New York.

Lesnoff, M. 2021. Jchemo: Chemometrics and machine learning on high-dimensional data with Julia. <https://github.com/mlesnoff/Jchemo>. UMR SELMET, Univ Montpellier, CIRAD, INRA, Institut Agro, Montpellier, France

Schaal S, Atkeson CG, Vijayakumar S. Scalable techniques from nonparametric statistics for real time robot learning. *Appl Intell.* 2002; 17(1): 49-60. <https://doi.org/10.1023/A:1015727715131>

Acknowledgments

I thank the European blood biomarkers consortium (Crowe M., UCD, Ireland; Koeck A., ZuchtData, Austria; Hummel J., University of Gottingen, Germany; Bapst B., Qualitas, Switzerland; Wolf V., CEL25-90, France; Leblois J., Elévéo, Belgium) and Clément Grelet (CRA-Walonie, Belgium) to have provided the BHB data.

Caractérisation des sols par SPIR : connaissance des erreurs de mesures et qualité de la prédiction des concentrations en carbone organique et en azote total

Othmane Aoultit¹, Cyril Girardin², Eric Latrille³, Marie-Noël Mistou⁴, Virginie Rossard³, Maud Seger⁵

¹ LAS, INRAE

² ECOSYS, Paris-Saclay, INRAE

² LBE, Université Montpellier, INRAE

² Agronomie, Paris-Saclay, INRAE

² InfoSol, INRAE

Email : Eric.latrille@inrae.fr

Mots-clefs : carbone organique, azote total, sol, prédiction SPIR

Les services écosystémiques principalement attendus par les agrosystèmes ou services agrosystémiques se basent en partie sur les fonctions remplies par les sols agricoles. Depuis 20 ans, plusieurs systèmes de culture innovants en rupture ont été conçus et testés à l'INRAE dans des « expérimentations systèmes ». Nous nous intéressons à caractériser, par spectrométrie proche infrarouge, les sols agricoles ayant opérés une transition agroécologique. Ainsi, le réseau INRAE-CAREX s'intéresse à la caractérisation des sols des unités expérimentales agricoles de l'INRAE et effectue des analyses spectrales infrarouge et des analyses physico-chimiques. Les mesures les plus importantes au niveau agronomique sont les teneurs en carbone et en azote dans le sol. La teneur en carbone dans le sol est composée de deux grandeurs : la teneur en carbone organique, utile pour le développement des cultures et la teneur inorganique ou minérale, utile à la structure physique du sol.

La spectrométrie est une méthode d'analyse indirecte qui nécessite une phase d'étalonnage préalable à la prédiction des concentrations des éléments étudiés. En effet, la granulométrie de l'échantillon et sa teneur en eau modifient fortement l'allure du spectre infrarouge et masque l'information sur les concentrations en carbone et en azote. Des méthodes de traitement du signal et de prétraitements adaptés des spectres sont nécessaires avant la phase d'étalonnage. L'étalonnage consiste à corrélérer les spectres à la variable d'intérêt, en utilisant des analyses de référence de cette variable, effectuées en laboratoire. L'étalonnage, une fois réalisé, est ensuite utilisé pour prédire la propriété cible sur de nouveaux échantillons en fonction de leur spectre seul. Le développement de modèles de prédiction utilisables en routine à partir des mesures spectrales permet d'augmenter le nombre de mesures de teneurs en carbone organique et azote total sur les parcelles car elles sont aisées à mettre en œuvre et moins coûteuses que les analyses chimiques.

Les échantillons ont été collectés sur 6 sites expérimentaux différents du dispositif CAREX en 2013- 2015 et en 2024 et correspondent à un large panel de sols. Les spectres et toutes les analyses de sol ont été acquis sur le même échantillon. Les échantillons ont été analysés au laboratoire LAS (Laboratoire d'Analyse des Sols, INRAE Arras, France) selon les normes internationales pour 6 composants chimiques : contenu_en_eau_du_sol_g.kg⁻¹ (105°C, (MO-015)), carbone organique du sol (SOC, g.kg⁻¹, ISO 14235, NF ISO 10694 (1995)), azote_total_g.kg⁻¹ (NF ISO 13878 (1998)). Les incertitudes de mesure du Corg sont dépendantes de la teneur en carbone total et de la teneur en carbone inorganique des échantillons selon les formules suivantes réactualisées chaque année. Avant 2020, Incertitude_Corg = 0.050 * Corg + 0.01800 * CaCO3 + 0.1360 et en 2024 Incertitude_Corg = 0.028 * Corg + 0.00590 * CaCO3 + 0.2784 , où Corg est la

concentration en carbone organique exprimée en g.kg^{-1} et CaCO_3 est la concentration en carbone inorganique exprimée en g.kg^{-1} . Les échantillons de sol ont été séchés à l'air à 40 °C maximum pendant 12 h puis broyés et tamisés (maille de 2 mm) (NF ISO 11464). Les échantillons ont été à nouveau séchés à l'air (à 40 °C max) juste avant la mesure des spectres. Tous les échantillons de sol ont été scannés par un spectromètre NIRFLEX N-500 Büchi. 50 g de sol tamisé séché ont été placés dans une boîte de 8,6 cm de diamètre. Chaque mesure est la moyenne de 32 balayages et est enregistrée en réflectance et en nombre d'onde et transformée en absorbance. La gamme spectrale est comprise entre 10 000 et 4 000 cm^{-1} et la résolution spectrale est de 8 cm^{-1} .

Le prétraitement et le traitement des données ont été réalisés avec le logiciel Chemflow [1]. L'ensemble des données spectrales et d'analyses physico-chimiques sont disponibles sur recherche.data.gouv.fr [2], [3]. La calibration des modèles de prédiction a été réalisée à l'aide de l'algorithme de régression PLS. Une validation croisée k-fold ($k=4$) a été utilisée pour sélectionner le nombre optimal de composantes de régression PLS dans les modèles d'étalonnage afin d'éviter un surajustement. Le nombre de composantes PLS retenu est celui qui a donné le RMSECV le plus bas conjointement à une valeur du critère de Durbin-Watson inférieure à 0,03.

Pour la calibration des modèles de prédiction, une pré-sélection des données de référence de teneur en carbone organique ayant une incertitude de mesure inférieure à 1 g.kg^{-1} a conduit à des erreurs de prédiction en validation-croisée de l'ordre de 2 g.kg^{-1} alors qu'elles étaient supérieures à 3 g.kg^{-1} pour les modèles obtenus sans pré-sélection des données. Ainsi, sur les 957 échantillons de sol retenus et analysés, seuls 705 échantillons ont été utilisés pour la prédiction du Corg. Il a été ainsi possible de prédire des concentrations en carbone organique des sols avec des précisions égales à 2,12 g.kg^{-1} pour des nouveaux prélèvements de sols de l'unité expérimentale d'Epoisses analysée en 2024 à partir d'un modèle global obtenu sur les 6 sites expérimentaux analysés en 2013 ou à partir d'un modèle local, c'est-à-dire à partir des données obtenues sur ce même site 10 ans plus tôt. A titre de comparaison, les erreurs de prédiction du carbone organique reportées dans la littérature sont de 4,4 g.kg^{-1} [4].

Nous montrons ainsi toute l'importance de sélectionner a priori les données servant à la calibration des modèles PLS sur la base des incertitudes de mesures des valeurs de référence.

References

- [1] Rossard Virgini et al. ChemFlow, chemometrics using Galaxy. F1000 Research 2016, 5:1671 (doi: 10.7490/f1000research.1112573.1)
- [2] Thoisy, Jeanne, et al. "French soil samples near infrared spectroscopy measurements and associated physico-chemical reference analysis.", <https://doi.org/10.15454/9RDHIN>, Recherche Data Gouv, V1 (2022).
- [3] Thoisy, Jeanne, et al. "physico_chemical_measurements_CAREX.tabular", French soil samples near infrared spectroscopy measurements and associated physico-chemical reference analysis., <https://doi.org/10.57745/QUCYI9>, Recherche Data Gouv, V1 (2022).
- [4] Clairotte, Michaël, et al. "National calibration of soil organic carbon concentration using diffuse infrared reflectance spectroscopy." Geoderma 276 (2016): 41-52.

Apport de la SO-PLS à l'approche multiblocs exploratoire

1 Marion Brandolini-Bunlon & 2* Benoît Jaillais

¹ Université Clermont Auvergne, INRAE, UNH, PFEM, MetaboHUB Clermont, Clermont-Ferrand, France. ² INRAE, TRANSFORM, Nantes, France

Email : benoit.jaillais@inrae.fr

Mots-clefs : proche infrarouge, sequential and orthogonalised-PLS (SO-PLS), Consensus-ACP, chimiométrie

De nombreuses études consistent à caractériser les mêmes échantillons avec différents procédés - chimiques, physiques, structuraux ou combinant plusieurs techniques spectroscopiques - pour prendre en compte le mieux possible la complexité des échantillons et la complémentarité des techniques analytiques. Ces données multivariées sont généralement consignées dans différents tableaux appariés par les échantillons ayant des variables différentes en nombre et en nature. Pour traiter ces données, il convient généralement d'analyser les blocs séparément puis ensemble, pour identifier si certains blocs sont spécifiques d'une source variabilité.

Dans notre travail, 3 blocs de données spectroscopiques (visible, proche infrarouge (NIR) et moyen infrarouge (MIR)) acquis sur des échantillons de corps gras sont considérés. Nous présentons l'ACP séparée des 3 blocs et l'apport de la consensus-ACP. Pour identifier la part de variabilité inhérente à chaque bloc, liée à la teneur en matière grasse (MG), nous avons utilisé la SO-PLS qui permet d'estimer la variabilité expliquée par un bloc en s'affranchissant de celle déjà expliquée par le(s) bloc(s) déjà présent(s) dans le modèle.

L'ACP appliquée à chaque bloc montre un léger effet lié à la teneur en matière grasse dans les blocs NIR et MIR. La consensus-ACP apporte une explication sur la variabilité commune aux blocs. Pour la SO-PLS, l'évolution de l'erreur de prédiction en fonction de la combinaison des variables latentes issues de chaque bloc est présentée dans le Mageplot (Fig.1a). La Fig.1b permet de visualiser les individus, colorés en fonction du pourcentage de MG, dans le premier plan factoriel. La SO-PLS souligne donc que le bloc NIR est celui qui permet de prédire au mieux la quantité de matière grasse des échantillons.

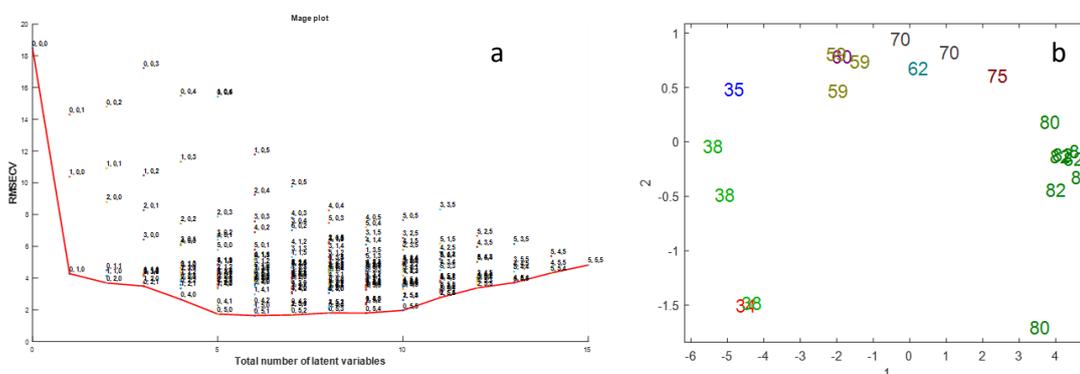


Figure 1: SOPLS Mageplot (a) et plan factoriel (b) pour les deux premières dimensions du modèle MIR&NIR&VIS

Are you prepared? Rediscovering wavelets and a new local regression approach

Andreas Niemöller^a, Antoine Deryck^b and Juan Antonio Fernández Pierna^b

^a Bruker Optics GmbH & Co. KG, Ettlingen, Germany,

^b CRAW-W, Gembloux, Belgium

Email : Andreas.Niemoeller@bruker.com

The pre-treatment, processing and, for example, targeted projection of data are important aspects in chemometrics, especially in the age of big data and ML through to AI. Data compression with good representation of the relevant information with fewer latent variables plays an especially important role. This is why data pretreatment and subsequent application of PCA and PLS are widely used in the field of chemometrics and, for example, NIR spectroscopy.

The wavelet transformation, which found further applications in chemometrics in the 1990s [1, 2], offers some possibilities for data treatment as well as compression with very good representation of the signal/data. One application is, for example, the smoothing of spectra and signals, which is achieved by eliminating signal components of higher frequencies, followed by signal reconstruction.

In this work, however, the focus is on the option of using the wavelet coefficients directly after the transformation to perform filtering and other operations. Examples are shown and corresponding approaches are discussed.

Ultimately, these considerations have been incorporated into a new approach to local regression which will be presented and has been implemented in a new software package, QUANT 3 in OPUS. The kNN step as well as the local PLS calibration are based on wavelet coefficients, which brings some simplifications to the modeling flow and provides good results [3].

References

[1] Alsberg, B. K.; Woodward, A. M.; Kell, D. B., An introduction to wavelet transforms for chemometricians - A time-frequency approach, 1997, *Chemometrics and Intelligent Laboratory Systems*, 37, 215-239, [https://doi.org/10.1016/S0169-7439\(97\)00029-4](https://doi.org/10.1016/S0169-7439(97)00029-4)

[2] Jouan-Rimbaud, D.; Walczak, B.; Poppi, R. J.; de Noord, O. E.; Massart, D. L., Application of wavelet transform to extract the relevant component from spectral data for multivariate calibration, 1997, *Anal Chem*, 69, 4317-4323, <https://doi.org/10.1021/ac970293n>

[3] Deryck, A.; Niemöller, A.; Baeten, V.; Fernández Pierna, J. A., Another pipeline in local Partial Least Squares Regression (LPLS) methods: Assessing the impact of wavelet transform integration, 2024, *Microchemical Journal*, 207, 112144, <https://doi.org/10.1016/j.microc.2024.112144>

Machine learning interpretability methods applied to calibration models developed on near infrared spectroscopic data

Astrid Malechaux¹, Jordane Poulain¹, Sylvie Roussel¹

¹Ondalys, 4 rue Georges Besse, 34830 Clapiers, France

Email: amalechaux@ondalys.fr – jpoulain@ondalys.fr – sroussel@ondalys.fr

Keywords: Machine Learning, Explainable ML, Interpretability, SVM, ANN

In the past decades, Machine Learning (ML) models have become more and more complex, leading to improvements in their predictive performance. However, these models can often be described as “black boxes”, in the sense that it is very difficult to explain how results are obtained by a model from the input data. As complex Machine Learning models are increasingly used to make decisions, for instance in industrial or medical applications, there is a growing need to improve their interpretability in order to have greater confidence in their results [1], providing the so-called Explainable AI (Artificial Intelligence).

This study is focused on the interpretability of Machine Learning models after model calibration on near-infrared spectroscopic data, a.k.a. the “post-hoc explanation” of ML models. Several regression models such as Support Vector Machines (SVM) and Artificial Neural Networks (ANN) have been trained and compared to the classical PLS calibration models.

The study is applied to the freely available Benchmark dataset consisting of Near Infrared spectra of minced meat samples recorded with the FOSS Tecator Infratec Food and Feed Analyzer in the 850-1050nm wavelength range [2].

Various ML model interpretability algorithms have been tested and compared: Local Interpretable Model-agnostic Explanations [3], Shapley Additive Explanations [4], pseudo-samples prediction [5], and some sensitivity tests developed by our team. These methods have also been compared to the regression coefficients of an intrinsically interpretable Partial Least Squares model. While the applicability to spectroscopic data of pseudo-samples prediction has been demonstrated [5], Local Interpretable Model-agnostic Explanations and Shapley Additive Explanations are theoretically better suited to uncorrelated variables [6].

The results indicate that the outputs of the interpretability methods tested appear to be in good agreement with the Partial Least Squares regression coefficients. Thus, they allow the identification of wavelengths ranges used by “black box” Machine Learning models. Moreover, they could be used as indicators of the risk of overfitting for complex models developed on spectroscopic data.

References

- [1] P. Linardatos, V. Papastefanopoulos, S. Kotsiantis, Explainable AI: A Review of Machine Learning Interpretability Methods, *Entropy*, 2021, 23, 18. (<https://doi.org/10.3390/e23010018>)
- [2] H. H. Thodberg, Tecator meat sample dataset, *StatLib Datasets Archive*, 1995.
- [3] M. T. Ribeiro, S. Singh, C. Guestrin, "Why should I trust you?" Explaining the predictions of any classifier, *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 2016, 1135-1144. (<https://doi.org/10.1145/2939672.293977>)

- [4] S. M. Lundberg, S. Lee, A Unified Approach to Interpreting Model Predictions, *Advances in Neural Information Processing Systems*, 2017, 30, 4765-4774.
- [5] G. J. Postma, P. W. T. Krooshof, L. M. C. Buydens, Opening the kernel of kernel partial least squares and support vector machines, *Analytica Chimica Acta*, 2011, 705(1-2), 123-134. (<https://doi.org/10.1016/j.aca.2011.04.025>)
- [6] A. M. Salih, Z. Raisi-Estabragh, I. B. Galazzo, P. Radeva, S. E. Petersen, K. Lekadir, G. Menegaz, A perspective on explainable artificial intelligence methods: SHAP and LIME, *Advanced Intelligent Systems*, 2024, 2400304. (<https://doi.org/10.1002/aisy.202400304>)

Tri optique infrarouge grain à grain comme outil de phénotypage haut débit technologique et sanitaire

Pierre-Yves Werrie¹, Corentin Demoitie², Philippe Vermeulen², Bruno Godin¹

¹ Unité Valorisation des produits, de la biomasse et du bois, Centre wallon de recherche agronomique (CRA-W), Gembloux, Belgique

² Unité Qualité et authentification des produits agricoles, Centre wallon de recherche agronomique (CRA-W), Gembloux, Belgique

Email : p.werrie@cra.wallonie.be

Mots-clefs : Céréales, phénotypage, qualité technologique, SK-NIR, tri optique

Dans le cadre du projet VALCERWAL, le Centre wallon de recherches agronomiques (CRA-W) développe une approche innovante pour valoriser les céréales produites en Wallonie, en conciliant exigences de qualité et durabilité des pratiques agricoles. Ce projet ambitionne de renforcer la résilience de la filière céréalière régionale en augmentant la valeur ajoutée des productions, en diversifiant leurs usages et en répondant aux normes technologiques et sanitaires en vigueur. L'outil central de cette démarche est une plateforme de tri optique proche infrarouge (NIR) opérant à l'échelle du grain individuel, combinant des technologies de spectroscopie et d'intelligence artificielle pour un tri précis, rapide et automatisé.

Le principe de cette technologie repose sur trois étapes clés : la singularisation des grains, leur analyse spectroscopique (par réflectance ou transmittance dans la gamme 900–1700 nm), et l'éjection ciblée de grains non conformes par air comprimé. Le système utilise notamment un spectromètre Hamamatsu C11482GA et un trieur optique Qsorter Explorer développé par Qualysense, capable de traiter un grain toutes les 20 à 30 millisecondes. Ce dernier intègre également des capteurs RGB et des algorithmes d'apprentissage automatique pour la classification en temps réel.

Les applications de cette technologie sont multiples pour l'ensemble des acteurs de la chaîne céréalières : agriculteurs, sélectionneurs, transformateurs, et distributeurs. En matière de contrôle qualité, elle permet de détecter automatiquement la présence de contaminants tels que l'ergot ou les grains fusariés, assurant ainsi la conformité des lots aux normes européennes strictes (par exemple, une limite de 0,2 g/kg pour les sclérotés). Elle permet également l'amélioration technologique des lots par un tri fondé sur des critères qualitatifs comme la teneur en protéines, la dureté ou l'homogénéité, garantissant de meilleures performances à la transformation (mouture, panification, etc.).

Enfin, cette technologie offre un outil précieux pour la sélection variétale. Grâce à l'utilisation de modèles statistiques (PLS-DA, PLS-R) et de techniques d'apprentissage machine, le système peut prédire des traits complexes tels que la vitrosité ou le mitadinage, facilitant l'identification de lignées de blé dur adaptées aux conditions agro climatiques belges. Cette capacité de caractérisation rapide et précise en fait un atout majeur pour des stratégies de phénotypage à haut débit. À terme, cette plateforme pourrait être étendue à d'autres environnements agricoles et couplée à des données génomiques, renforçant encore son utilité en sélection végétale.

Cette démarche illustre comment l'intégration de technologies avancées au sein d'une filière agricole peut répondre à des enjeux agronomiques, économiques et environnementaux cruciaux pour l'agriculture régionale.

Near Infrared Spectroscopy @KremsChem

Ramin Nikzad-Langerodi¹

¹ Senior Researcher Data Science
Research Team Lead Sustainable Process Cognition
Software Competence Center Hagenberg GmbH
A 4232 Hagenberg, Softwarepark 32a

Email : Ramin.Nikzad-Langerodi@scch.at

This keynote aims at highlighting the interdisciplinary effort between the Software Competence Center Hagenberg (GmbH) and associated RTOs to integrate different near-infrared (NIR) technologies at KremsChem Austria, the regional market and technology leader for resins, as well as a specialist in fine chemicals, surfacing materials and flame retardants. The talk will cover aspects related to sensor selection and integration, data acquisition and management, as well as the development of tailored software solutions for process monitoring, control and optimization. A special emphasis will be on chemometrics and machine learning, in particular on how different challenges encountered during integration and operation of NIR technology in production prompted original research in these fields.

Applications industrielles des NIR portatifs

Mickael Watiez¹, Zoé Fayeulle¹,

¹ Avril, Paris, France

Email : mickael.watiez@avril.com

Mots-clefs : NIR portatifs, analyse in situ, applications industrielles

La technologie NIR est utilisée au sein du groupe Avril depuis 2009 avec les premières calibrations développées sur les huiles de graines et d'olive. Elle se propage sur tous les procédés du groupe comme le raffinage, l'estérification ou le conditionnement dans les domaines de la nutrition humaine, animale, du biocarburant ou encore de l'oléochimie. Les analyses sont réalisées en laboratoire sur les matières premières et les produits finis, en production sur les produits dits « en cours » par les opérateurs de fabrication ou directement « on line » sur certaines de nos opérations unitaires.

Ce parc instrumental constitué aujourd'hui d'une trentaine d'instruments s'enrichit d'une technologie complémentaire : le NIR ultraportatif. Comme son nom l'indique, sa principale qualité est la portativité. Elle permet ainsi l'analyse « in situ » et offre ainsi un nouveau champ d'application chez Avril : de la livraison des matières premières jusqu'à l'aide à la conduite de process. Le projet « NIR 4.0 » est né, il a été réalisé en plusieurs phases : étude du marché fournisseur, mesure de l'attractivité de la technologie au sein des business units d'Avril, sélection d'un fournisseur, création des premières calibrations et déploiement.

Les POC (preuves de concept) ont été réalisés sur graines de colza et de tournesol et ont rapidement confirmé l'intérêt de la technique mais aussi ses limites notamment le domaine de longueur d'onde, logiquement plus restreint que les instruments NIR classiques. Actuellement, les premiers instruments sont en cours de déploiement notamment utilisés pour la libération des matières premières lors de la livraison par camions. En parallèle, de nombreuses applications ont été identifiées pour compléter le catalogue actuel : sélection des graines de colza selon leur teneur en protéine et matière grasse, maturité des olives, libération de citernes d'huile ou mesures on line (NIR-PAT).

A physics-based data simulation and augmentation method for enhancing calibration model development on a visible spectrometer

^{1,2}Simon PINAULT, ²Daphné HERAN, ¹Davinia BROUCKAERT, ¹Fabien CHAUCHARD, ²Ryad BENDOULA

¹ Indatech (Chauvin-Arnoux Group)

² Inrae, ITAP-comic, Montpellier

Keywords: Simulation, Chemometrics, Data augmentation, PAT, Spectroscopy

Spectrometer calibration is a hot topic for PAT industrials and researchers because of the difficulties and challenges it brings [1]. For industrial applications, a low number of calibration samples and/or limited variability often compromise the chemometric calibration model development and hence its performances. Data augmentation is gaining popularity for overcoming this issue on visible or near-infrared spectrometers, for instance by using Deep Learning [2], pure components estimation [3] or interpolation [4].

This work presents a new data augmentation method based on a mechanistic physics-based approach. The method uses an optics model to simulate realistic synthetic spectra. Those simulated spectra are then used to augment measured calibration sets. The study focuses on a data set made out of a small number of samples with a limited variability. Moreover, we study the extreme case where the independent test data is at the edge of the variability range.

The data set consists of 55 phantoms samples displaying varying absorption and scattering properties, inspired by those made by Watté and al. [4]. They were inspected in reflectance and transmittance mode simultaneously, thanks to a double-integration sphere and visible spectrometer setup. Synthetic spectra were computed according to the Kubelka-Munk model [5] with Saunderson correction. Two data augmentation modes were investigated, namely completion and superposition. The augmented data sets are compared with a baseline set, in terms of correlation (R^2) and prediction (SEP) accuracy, by training partial least squares (PLS) regression models. This study investigates the benefits of augmenting the calibration data with synthetic data generated from various calibration sets.

The proposed method shows encouraging results in the extreme case where the test data are at the edge of the variability range (extrapolation). The strengths and weaknesses of the method will be addressed.

This work contributes to the recent trends in development of intelligent instruments for the PAT industry.

References

- [1] W. Chew, P. Sharatt, *Analytical Methods*, 2010, 2, 1412-1438
- [2] C. Jiang, J. Zhao, *Sensors*, 2023, 23(7), 3686
- [3] R. Schiemer, M. Rüdte, *Frontiers in Bioengineering and Biotechnology*, 2024, 12, 1228846
- [4] R. Watté, N. Do Trong, *Optics Express*, 2013, 21(26), 32630
- [5] P. Kubelka, *Journal of the Optical Society of America*, 38(5), 448

Évaluation de méthodes d'apprentissage profond pour les données spectroscopiques et la chimiométrie : un cas d'étude pour les applications agro-environnementales

Florent Abdelghafour^{1,3}, Maxime Metz^{2,3}

¹ ITAP, Univ. Montpellier, INRAE, Institut Agro, 34196 Montpellier, France

² Pellenc Selective Technologie, 125 Rue François Gernelle, BP124, 84120 Pertuis, France

³ Labcom AIOLY,

Florent.abdelghafour@inrae.fr m.metz@pellencst.com

Mots-clefs: Apprentissage profond, Chimiométrie, spectroscopie

Avec le déploiement croissant des capteurs optiques en laboratoire, au champ ou en serre, les bases de données spectrales augmentent en taille, deviennent plus diversifiées et intègrent de plus en plus de multi-modalité. Des approches ont été développées pour traiter des modèles complexes et non linéaires dans un contexte de big data [1]. Parallèlement, l'Apprentissage profond (AP) ouvre une nouvelle voie de recherche pour faire face aux nouveaux défis de la chimiométrie. Toutefois, malgré sa popularité croissante et son adoption massive dans d'autres domaines tels que la vision par ordinateur et le traitement du langage naturel, il en est encore à ses débuts dans le domaine des données spectrales.

L'adoption du AP en chimiométrie fait face à de nombreuses limitations, à commencer par la disponibilité de codes open source et d'architectures AP standards. Il n'existe pas de consensus concernant des étapes fondamentales comme l'augmentation de données, le prétraitement ou les stratégies d'apprentissage. L'un des atouts du AP est sa capacité à fournir des modèles génériques, applicables à une diversité de bases de données grâce aux mécanismes d'apprentissage par transfert. Dans le cas des données spectrales, il reste à déterminer comment exploiter les caractéristiques génériques apprises à partir de données diverses pour ajuster les modèles à des applications spécifiques.

Dans ce travail, nous proposons de présenter des architectures de réseaux de neurones standards, adaptées de la vision par ordinateur à la chimiométrie. L'objectif est de mettre en lumière les limitations actuelles du AP au regard de l'état de l'art en chimiométrie, et de dégager des perspectives pour lever les verrous scientifiques. Dans cette optique, une solution open source permettant d'expérimenter des modèles de « deep chemometrics » est présentée. Nous proposons une étude de cas sur des bases de données spectrales ouvertes : OSSL [2], Anderson Mango's [3], Wheat Kernel [4], qui fournissent une base de référence pour l'évaluation des performances. Nous avons comparé les résultats en régression et en classification obtenus avec nos modèles AP exploratoires de base, incluant une stratégie d'apprentissage par transfert, avec des approches standards en chimiométrie telles que la PLS.

Parmi les architectures, nous avons testé certaines spécifiquement développées pour la chimiométrie, en plus de quelques approches emblématiques de la vision par ordinateur que nous avons adaptées aux données spectrales :

- **CuiNet [5]** : un CNN 1D composé d'une seule couche de convolution et d'un MLP à trois couches
- **DarioNet [6]** : une version 1D optimisée d'un CNN léger pour les données NIRS
- **ResNet-1D**, versions 18, 34, 50 et 101 couches, réseaux résiduels 1D adaptés de [7]
- **DeepSpectra [8]** : un CNN 1D incluant des modules inception
- **ViT-1D** : une adaptation du transformeur visuel emblématique [9]

Les résultats obtenus montrent que même les grands modèles peuvent atteindre de bonnes performances sur des bases de données de taille intermédiaire. Toutefois, il a été observé que ces approches, bien que nettement plus coûteuses en temps d'entraînement et plus complexes à optimiser en raison du grand nombre de paramètres, ne surpassent pas actuellement les méthodes traditionnelles telles que la PLSR (régression) ou la PLS-DA (analyse discriminante).

L'apprentissage par transfert n'a pas amélioré les performances, probablement en raison de l'ultra-spécialisation des bases de données spectrales. Contrairement à la vision par ordinateur, où de grandes bases de données génériques sont largement disponibles, les bases de données spectrales accessibles sont généralement très spécialisées. Par conséquent, le transfert d'un modèle entraîné sur une base spécialisée vers une autre base spécialisée ne fournit pas encore de performances satisfaisantes.

Ce travail met en évidence la principale limitation de l'Apprentissage profond : le manque d'intégration des connaissances spécifiques à la chimométrie dans les contributions en Apprentissage profond. Pour y remédier, la communauté de la chimométrie doit rendre accessibles ses codes, modèles et données afin d'accélérer le développement du AP en chimométrie et d'ouvrir de nouvelles pistes de recherche pour les propriétés physico-chimiques et la simulation spectrale.

REMERCIEMENTS

Ce travail a été réalisé en utilisant les ressources HPC de GENCI-IDRIS (Projet 2025-AD010114820R1). Projet financé par l'ANR-23-LCV2-0015 de l'Agence Nationale de la Recherche.

References

- [1] M. Metz, et al. A "big-data" algorithm for KNN-PLS, *Chemometrics and Intelligent Laboratory Systems*, Volume 203, 2020, 104076, ISSN 0169-7439, <https://doi.org/10.1016/j.chemolab.2020.104076>.
- [2] Safanelli, J. et al. 2023. Open Soil Spectral Library (OSSL): Building reproducible soil calibration models through open development and community engagement. *bioRxiv preprint*. DOI:10.1101/2023.12.16.572011.
- [3] N.T. Anderson, et al. Achieving robustness across season, location and cultivar for a NIRS model for intact mango fruit dry matter content, *Postharvest Biology and Technology*, Volume 168, 2020, 111202, ISSN 0925-5214, <https://doi.org/10.1016/j.postharvbio.2020.111202>.
- [4] Zhou L, et al. Wheat Kernel Variety Identification Based on a Large Near-Infrared Spectral Dataset and a Novel Deep Learning-Based Feature Selection Method. *Front Plant Sci*. 2020 Nov 10;11:575810. doi: 10.3389/fpls.2020.575810. PMID: 33240294; PMCID: PMC7683420.
- [5] Chenhao Cui, Tom Fearn, Modern practical convolutional neural networks for multivariate regression: Applications to NIR calibration, *Chemometrics and Intelligent Laboratory Systems*, Volume 182, 2018, Pages 9-20, ISSN 0169-7439, <https://doi.org/10.1016/j.chemolab.2018.07.008>.
- [6] Dário Passos, Puneet Mishra, Deep Tutti Frutti: Exploring CNN architectures for dry matter prediction in fruit from multi-fruit near-infrared spectra, *Chemometrics and Intelligent Laboratory Systems*, Volume 243, 2023, 105023, ISSN 0169-7439, <https://doi.org/10.1016/j.chemolab.2023.105023>.
- [7] He, K. et al. 2015. Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770-778.
- [8] X. Zhang et al. DeepSpectra: An end-to-end Deep Learning approach for quantitative spectral analysis, *Analytica Chimica Acta*, Volume 1058, 2019, Pages 48-57, ISSN 0003-2670, <https://doi.org/10.1016/j.aca.2019.01.002>.
- [9] A. Dosovitskiy, et al. "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale." *ArXiv abs/2010.11929* (2020)

Développement et intégration à une plateforme utilisée en milieu industriel d'un modèle NIR de classification des matières premières d'origine agricole

Vincent Larat

Adisseo, Centre d'Expertise et de Recherche en Nutrition (CERN), 03600 Malicorne

Email: vincent.larat@adisseo.com

Adisseo, acteur dans la filière de l'alimentation animale, offre à ses clients en accompagnement de ses produits un certain nombre de services. Le service PNE (Precise Nutrition Evaluation) est une application web permettant de caractériser par spectrométrie Proche IR la valeur nutritionnelle des matières premières d'origine agricoles utilisées pour la fabrication d'aliments.

Le client utilisateur charge sur la plateforme PNE le spectre de l'échantillon à tester, mesuré au préalable sur son appareil NIRS (Foss ou Bruker). Il sélectionne ensuite la matière première correspondante ainsi que les analyses souhaitées pour obtenir, via les modèles adéquats, l'évaluation du profil nutritionnel.

Il arrive cependant que des erreurs soient commises par l'utilisateur du service soit lors du scan de l'échantillon, soit lors du chargement des spectres ou de la sélection de la matière première, ayant pour conséquences des prédictions erronées (outliers). Le traitement de ces outliers demande une intervention humaine et donc des ressources internes non négligeables.

Afin de limiter l'occurrence de ces outliers dus à une inadéquation entre les spectres et les modèles utilisés, une reconnaissance automatique de la matière première lors du chargement du spectre est incluse à la plateforme PNE. Celle-ci repose sur un modèle de classification basé sur les spectres NIR : différents algorithmes de classification ont été testés (PLS-DA, SVM) et un système de validation robuste a été mis en place afin d'optimiser la reconnaissance de l'échantillon. Le modèle de classification retenu est intégré à la plateforme PNE, permettant ainsi de détecter automatiquement ces inadéquations entre le spectre et la matière première sélectionnée et de réduire significativement le nombre d'outliers dus à ces erreurs.

La sélection phénotypique : Quand les généticiens s'essaient à la NIRS

Clément Bienvenu^{1,2}, David Pot^{1,2}, Vincent Segura^{2,3}, Hugues de Verdal^{1,2}

¹ CIRAD, UMR AGAP Institut, Montpellier, F-34398, France

² UMR AGAP Institut, Univ Montpellier, CIRAD, INRAE, Institut Agro, Montpellier, F-34398, France

³ Geno-Vigne, IFV-INRAE-Institut Agro, F-34398, Montpellier, France

Email : clement.bienvenu@cirad.fr

Mots-clefs : Sélection variétale, prédiction, génétique quantitative

La sélection variétale permet le développement de nouvelles variétés de plantes adaptées aux changements climatique et est cruciale pour garantir la sécurité alimentaire [1]. La sélection génomique (SG) [2] est une méthode qui permet de prédire la performance des variétés (rendement, durée du cycle, taille etc...) à partir d'information génétique obtenue par séquençage de l'ADN des plantes. La SG a déjà permis d'augmenter l'efficacité de la sélection en réduisant le temps et en augmentant la précision de l'évaluation des variétés. La sélection phénotypique (SP) [3] est une adaptation de la SG où les prédictions sont réalisées avec les mêmes modèles, mais à partir de spectres infrarouges pris sur différents tissus des plantes au lieu de l'information génétique. Ces modèles (SG et SP) permettent de prédire la valeur génétique des variétés et non leurs valeurs phénotypiques. Le passage de l'information génétique à l'information spectrale est motivé par le moindre coût et le haut débit d'acquisition des spectres qui pourraient permettre d'augmenter encore l'efficacité de la sélection. Contrairement aux données de séquençage qui ne contiennent que de l'information génétique, les spectres contiennent aussi en théorie de l'information sur l'environnement de croissance de la plante dont on pourrait tirer parti pour prédire plus précisément la performance de variétés dans des environnements donnés.

Plusieurs études montrent que la SP peut avoir des performances similaires à la SG (la SG étant une référence) mais réagit différemment à certains paramètres importants comme la taille du jeu d'entraînement, ou l'apparentement génétique entre population d'entraînement et population cible [4], [5]. Les mécanismes expliquant ces performances n'ont pas été élucidés, notamment, nous ne savons pas à quel point les prédictions se font par un mécanisme de type "proxy" en tirant parti des corrélations phénotypiques entre les spectres et les traits étudiés, ou se font en tirant parti de l'information génétique présente dans les spectres pour prédire les valeurs génétiques des variétés.

Certains résultats sur le nombre de longueurs d'ondes nécessaires pour obtenir des prédictions [6] montrent que les méthodes de génétique exploitent probablement mal l'information spectrale. Une piste intéressante pour mieux comprendre comment la SP fonctionne et améliorer sa capacité prédictive serait de réussir à décomposer les spectres en leurs parties génétiques et environnementales pour mieux utiliser ces parties séparément dans les modèles de prédiction. Cette décomposition pourrait se faire à l'aide de méthodes chimiométriques plus adaptées pour analyser les spectres.

References

[1] M. Tester et P. Langridge, « Breeding Technologies to Increase Crop Production in a Changing World », *Science*, vol. 327, n° 5967, p. 818-822, févr. 2010, doi: 10.1126/science.1183700.

[2] T. H. E. Meuwissen, B. J. Hayes, et M. E. Goddard, « Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps », *Genetics*, vol. 157, n° 4, p. 1819-1829, avr. 2001, doi: 10.1093/genetics/157.4.1819.

- [3] R. Rincent *et al.*, « Phenomic Selection Is a Low-Cost and High-Throughput Method Based on Indirect Predictions: Proof of Concept on Wheat and Poplar », *G3 GenesGenomesGenetics*, vol. 8, n° 12, p. 3961-3972, déc. 2018, doi: 10.1534/g3.118.200760.
- [4] X. Zhu *et al.*, « The performance of phenomic selection depends on the genetic architecture of the target trait », *TAG Theor. Appl. Genet. Theor. Angew. Genet.*, vol. 135, n° 2, p. 653-665, févr. 2022, doi: 10.1007/s00122-021-03997-7.
- [5] X. Zhu, W. L. Leiser, V. Hahn, et T. Würschum, « Phenomic selection is competitive with genomic selection for breeding of complex traits », *Plant Phenome J.*, vol. 4, n° 1, p. e20027, 2021, doi: 10.1002/ppj2.20027.
- [6] T. Mróz, S. Shafiee, J. Crossa, O. A. Montesinos-Lopez, et M. Lillemo, « Multispectral-derived genotypic similarities from budget cameras allow grain yield prediction and genomic selection augmentation in single and multi-environment scenarios in spring wheat », *Mol. Breed.*, vol. 44, n° 1, p. 5, janv. 2024, doi: 10.1007/s11032-024-01449-w.

Utilisation de la spectroscopie NIR pour évaluer la qualité des graines de hêtre selon leur environnement

Océane Parthenay, Brigitte Musch, Caroline Teyssier

BioForA, INRAE, ONF, UMR 0588, Orléans, France

Dans le contexte du changement climatique, les forêts européennes subissent des stress environnementaux, pouvant entraver la régénération naturelle. À l'échelle du peuplement, le hêtre (*Fagus sylvatica*), en raison de son rôle essentiel dans la biodiversité et l'industrie forestière, est particulièrement affecté dans son processus d'intermittence de productions de graines (phénomène de masting) et par le stress hydrique. Déjà une plus faible survie des semis a été constatée. La régénération des forêts dépend directement de la qualité des graines, déterminée par leur composition biochimique et leur diversité génétique, qui conditionne leur adaptabilité aux nouvelles conditions climatiques. Toutefois, évaluer rapidement et de manière non destructive cette qualité en milieu forestier reste un défi.

Ma thèse étudie l'impact des contraintes environnementales sur la qualité des graines de hêtre, en utilisant la spectroscopie NIR pour une analyse rapide, non destructive et in situ de leur composition biochimique et de leur potentiel de germination. L'objectif est de développer des modèles prédictifs afin d'évaluer la qualité des graines.

Deux études complémentaires sont menées :

Étude 1 : Évaluer l'influence de l'altitude et des variations interannuelles de production sur la qualité biochimique des graines. Pour cela, des analyses biochimiques, associées à la spectroscopie NIR, sont réalisées sur des graines récoltées dans le massif de Chamrousse.

Étude 2 : Analyser l'impact du stress hydrique sur la qualité des graines et leur descendance. Cette analyse repose sur des études biochimiques et génétiques des graines, combinant spectroscopie NIR, tests de germination et suivi des semis, avec un génotypage des parents et des semis.

L'innovation de cette approche réside dans l'application de la spectroscopie NIR aux graines récoltées en forêt, permettant de capturer la variabilité interannuelle et inter-sites, et de développer un modèle prédictif pour aider les gestionnaires forestiers à anticiper la régénération des hêtres face aux changements climatiques. Les résultats attendus incluent l'identification de signatures biochimiques et génétiques associées à la germination et à la résilience, l'optimisation des critères de sélection des peuplements producteurs de graines et la fourniture d'un outil rapide et non destructif pour améliorer les stratégies de renouvellement forestier.

Authentification et prédiction de paramètres physico-chimiques dans la cannelle

A. Barreau^{1,2}, S. Charvet², E. Mezzatesta¹, P. Faury², I. Bombarda¹, N. Dupuy¹

¹ IMBE, Aix Marseille Univ, Avignon Univ, CNRS, IRD, Marseille, France

² Vanalyse sarl, 14 Boulevard du Travail, 13013 Marseille, France

Email: anthony.barreau@imbe.fr

Mots-clefs : Cannelle, PIR, Chimiométrie, Authenticité

Les épices constituent une valeur ajoutée dans l'industrie alimentaire, notamment en raison de leur impact sur la saveur des produits. Avec la hausse des fraudes et des erreurs sur les étiquettes, il devient plus que jamais nécessaire de pouvoir vérifier l'authenticité des produits. Pour cela, certaines techniques d'analyse permettent d'identifier précisément leur origine géographique, leur espèce, leur variété ou encore les méthodes employées pour les produire. Mais pour que ces outils soient réellement fiables, il faut s'appuyer sur l'étude d'un grand nombre d'échantillons, en tenant compte de toutes les variations possibles : la fabrication, le stockage, le conditionnement, l'origine, la partie utilisée du produit, etc. En combinant l'efficacité des analyses chimométriques, des techniques spectroscopiques rapides et des données de référence de l'entreprise, il est possible d'élaborer des outils efficaces de contrôle qualité et de traçabilité des épices. Cette étude se concentre sur la cannelle, en s'appuyant sur l'analyse chimique pour construire des modèles non ciblés d'authentification et de prédiction. Grâce à un échantillonnage important et à l'acquisition de spectres PIR (Proche InfraRouge) sur différentes catégories de cannelle (bâtons, brisures, poudres), il est possible de distinguer les échantillons selon leur catégorie et d'identifier les échantillons potentiellement adultérés. Les données utilisées pour le contrôle qualité incluent les teneurs en humidité, en huile essentielle et en cinnamaldéhyde. Les méthodes employées, telles que la spectroscopie PIR, la chromatographie liquide haute performance (CLHP) et la chimiométrie, permettent d'extraire des informations pertinentes pour modéliser et prédire les propriétés physico-chimiques de l'épice. Il est possible d'élaborer des modèles prédictifs fiables. Cette approche non destructive optimise le contrôle qualité et l'authentification de la cannelle tout en réduisant les besoins en analyses longues et coûteuses. Un autre aperçu des résultats issus de l'association entre le PIR et la chimiométrie est par exemple la capacité à identifier les échantillons présentant une forte teneur en coumarine, qui est une substance naturelle réglementée. Il est donc essentiel de pouvoir la déterminer afin de garantir la sécurité alimentaire.

Comparaison de l'augmentation de données et du prétraitement pour l'apprentissage profond

Ivy Tumoine^{1,3,4}, Florent Abdelghafour^{1,2,3}, Maxime Metz^{3,4}, Nicolas Grotus⁴, Ryad Bendoula^{1,3}, Jean-Michel Roger^{1,2,3}

¹INRAE, Institut agro Montpellier, ITAP

²ChemHouse research group

³Labcom Aioly

⁴Pellenc Selective Technologies

Email : i.tumoine@pellencst.com

Mots-clefs : augmentation de données, apprentissage profond, chimiométrie

Les spectres proche infrarouge sont souvent entachés de perturbations, telles que les effets de la diffusion et les artefacts instrumentaux. Ces perturbations peuvent masquer l'information utile et compromettre la robustesse des modèles, elles sont d'autant plus nombreuses que la diversité des dispositifs et des échantillons est grande. L'approche classique en chimiométrie consiste à minimiser ces perturbations par un prétraitement approprié, puis, à calibrer des modèles linéaires pour la prédiction de variables d'intérêt. Cependant, le choix d'un prétraitement, ou d'un ensemble de prétraitements et leur combinaison, est une limitation importante pour le déploiement de modèle chimiométriques. Parallèlement, l'apprentissage profond (deep learning) s'est imposé comme un outil avancé standard dans de nombreux domaines et gagne en popularité pour l'analyse des spectres. En apprentissage profond, l'approche usuelle de prise en compte des perturbations est l'augmentation de données [1]. Contrairement aux prétraitements qui visent à retirer les perturbations nuisibles, l'augmentation de données consiste à enrichir l'ensemble d'apprentissage en perturbant les données. En forçant le réseau de neurones à devenir invariant aux perturbations introduites, les augmentations améliorent la qualité des modèles. Relativement peu d'études ont proposé des méthodes d'augmentation de données spectrales pour l'entraînement de réseaux de neurones, les plus pertinentes étant [2], [3], [4]. Notamment, l'approche présentée dans [3] consiste à estimer des distributions gaussiennes pour chacun des effets (pente polynomiale et effet multiplicatif) du prétraitement EMSC (Extended Multiplicative Signal Correction [5]), puis, à échantillonner indépendamment à partir de ces distributions pour générer des spectres avec des effets légèrement différents (EMS Augmentation). Cependant, dans la plupart des cas réels, ces effets ne sont pas indépendants : ils sont structurés. L'absence de prise en compte de cette structure risque de générer des données non-réalistes, masquant ainsi la relation entre x et y . De plus, si l'approche présentée dans [3] semble montrer que les augmentations de données sont un meilleur paradigme que les prétraitements pour l'entraînement de modèles profonds en chimiométrie, le consensus n'est pas encore atteint.

Dans cette présentation, nous proposons d'abord d'étudier le potentiel des augmentations de données en se plaçant dans le cas le plus favorable, où les données sont simulées. Nous proposons ensuite une nouvelle méthode d'augmentation de données. Comme l'EMSA, notre méthode repose sur le prétraitement pour identifier les effets non liés à la variable cible, puis utilise ces effets pour perturber les spectres. Toutefois, l'augmentation proposée fait l'hypothèse que les perturbations introduites doivent respecter la structure des effets identifiés. Nous proposons de comparer nos augmentations à l'EMSA [3] et aux prétraitements standards sur un jeu de données réel de mangues. Les résultats obtenus suggèrent que l'hypothèse de structure est valable, et que pour l'apprentissage profond, l'augmentation de données est un meilleur paradigme que le prétraitement traditionnel.

References

- [1] C. Shorten et T. M. Khoshgoftaar, « A survey on Image Data Augmentation for Deep Learning », *J. Big Data*, vol. 6, no 1, p. 60, juill. 2019, doi: 10.1186/s40537-019-0197-0.
- [2] E. J. Bjerrum, M. Glahder, et T. Skov, « Data Augmentation of Spectral Data for Convolutional Neural Network (CNN) Based Deep Chemometrics », 5 octobre 2017, arXiv: arXiv:1710.01927. doi: 10.48550/arXiv.1710.01927.
- [3] U. Blazhko, V. Shapaval, V. Kovalev, et A. Kohler, « Comparison of augmentation and pre-processing for deep learning and chemometric classification of infrared spectra », *Chemom. Intell. Lab. Syst.*, vol. 215, p. 104367, août 2021, doi: 10.1016/j.chemolab.2021.104367.
- [4] M. Wohlers, A. McGlone, E. Frank, et G. Holmes, « Augmenting NIR Spectra in deep regression to improve calibration », *Chemom. Intell. Lab. Syst.*, vol. 240, p. 104924, sept. 2023, doi: 10.1016/j.chemolab.2023.104924.
- [5] N. K. Afseth et A. Kohler, « Extended multiplicative signal correction in vibrational spectroscopy, a tutorial », *Chemom. Intell. Lab. Syst.*, vol. 117, p. 92-99, août 2012, doi: 10.1016/j.chemolab.2012.03.004.

Caractérisation de la diversité des ressources pastorales consommées par des petits ruminants sur parcours méditerranéens : apports de la spectroscopie infrarouge

Elisa Deschamps¹, Magali Jouven¹, Lucile Sautot², Denis Bastianelli¹

¹ SELMET, Univ Montpellier, CIRAD, INRAE, Institut Agro, Montpellier, France

² TETIS, Univ Montpellier, AgroParisTech, CIRAD, CNRS, INRAE

Email : elisa.deschamps@supagro.fr

Mots-clefs : ovins, comportement alimentaire, fourrage, classification, SPIR

Introduction

Dans les systèmes pastoraux méditerranéens, les parcours offrent une ressource fourragère essentielle, à faible coût. Toutefois, leur forte hétérogénéité spatiale et saisonnière, accentuée par le changement climatique, complique l'évaluation rapide et fiable de leur potentiel alimentaire. Chaque milieu génère une grande diversité de bouchées, composées d'assemblages complexes de parties végétales issues de multiples espèces. La composition chimique seule ne permet ni de saisir cette complexité, ni de refléter les choix alimentaires opérés par les animaux. Les outils développés pour les prairies homogènes s'avèrent peu adaptés à ces contextes complexes (Daget & Poissonet, 1971). La spectroscopie dans le proche infrarouge (SPIR), combinée à des observations directes du comportement alimentaire pour identifier les parties de plantes consommées par les troupeaux selon les contextes, émerge comme une solution prometteuse pour décrire la diversité des ressources alimentaires à moindre coût (Silué et al., 2016). Ce travail vise à développer une base de données des aliments consommés par les petits ruminants sur parcours méditerranéens, et à en tirer une classification des ressources pastorales opérationnelle dans différents milieux et pour différentes saisons, en vue d'accompagner la conduite adaptative du pâturage sur parcours.

Matériel et méthodes

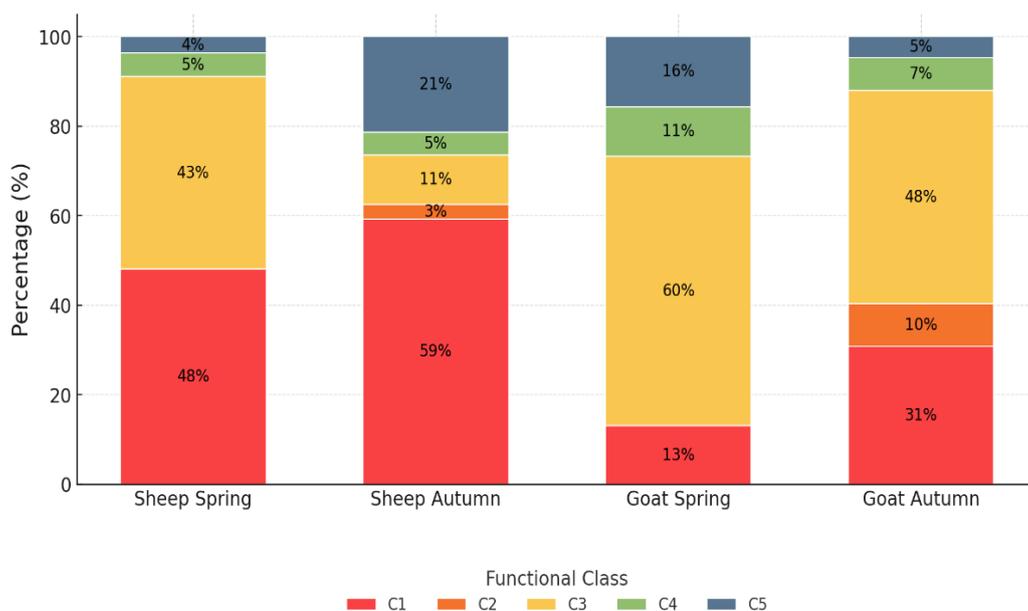
Des observations du comportement alimentaire d'ovins et de caprins ont été réalisées au printemps et à l'automne 2024 sur trois exploitations pastorales (un site en Ardèche, deux dans l'Hérault). Sur chaque site, quatre animaux ont été suivis pendant quatre jours (6 à 8 heures par jour) après une semaine d'adaptation. Chaque bouchée consommée a été codée (espèce végétale, partie de plante, état phénologique). Au total 165 échantillons mimant les bouchées principales ont été prélevés. Les échantillons ont été séchés (48h à 55°C) et broyés (1 mm) avant d'être analysés au spectromètre NIRSystems 5000 (1100–2500 nm, 2 nm). Les spectres ont été prétraités par normalisation SNV, detrend et dérivée seconde (Savitzky-Golay). Des classifications non supervisées (CAH avec critère de Ward) ont permis d'identifier des regroupements basés sur les caractéristiques spectrales. Contrairement aux approches SPIR classiquement basées sur la prédiction de constituants chimiques, nous utilisons ici le spectre complet comme une carte d'identité globale de l'échantillon, prenant en compte non seulement les constituants majeurs mais également certains composés secondaires (p.ex. tanins) et certains paramètres physiques (paroi cellulaire).

Résultats et discussion

L'analyse des spectres a permis d'identifier cinq classes fonctionnelles distinctes (Tab. 1), chacune regroupant plusieurs espèces collectées au printemps et à l'automne. Ces classes se différencient principalement par la partie de plante consommée (feuille, tige, rameau, fruit) plutôt que par l'espèce

botanique (Dochier et al., 2025). Elles capturent à la fois la variabilité saisonnière des ressources et leurs caractéristiques nutritionnelles, offrant une lecture simplifiée mais pertinente de la végétation pâturée. Le croisement avec les observations comportementales (Fig. 1) révèle que les caprins privilégient les classes lignifiées (C3), tandis que les ovins consomment davantage d'herbacées et de graminées (C1), traduisant des stratégies alimentaires différenciées. En automne, une consommation accrue de ligneux (classe C2) est observée pour les deux espèces, soulignant leur capacité d'adaptation saisonnière. Ainsi, la classification SPIR met clairement en évidence des différences de régimes alimentaires selon la saison et l'espèce, en accord avec les connaissances scientifiques existantes.

Figure 1 : Représentation simplifiée, à partir des 5 classes fonctionnelles, des régimes alimentaires des ovins et des caprins suivis en Ardèche au printemps et à l'automne 2024



L'analyse SPIR se révèle prometteuse pour caractériser de manière opérationnelle et standardisée les régimes alimentaires des petits ruminants sur parcours. Les premiers résultats indiquent que les classes identifiées regroupent des bouchées provenant de sites et de saisons variés, ce qui suggère un potentiel de généralisation et de transférabilité. Toutefois, des analyses complémentaires seront nécessaires pour tester la robustesse de cette classification, notamment en évaluant sa sensibilité aux variations spatiales et saisonnières. En complément des typologies botaniques classiques (Hanisch et al., 2020) ou fondées sur les caractéristiques morphologiques des bouchées (Agreil et al., 2004), cette approche spectrale offre une base innovante pour construire des références directement mobilisables par les éleveurs. Les prochaines étapes viseront à consolider et affiner cette classification fonctionnelle à travers plusieurs leviers : une analyse fine de la sensibilité des classes, leur confrontation aux savoirs empiriques des éleveurs et bergers, ainsi que l'évaluation de leur pertinence opérationnelle dans des contextes de gestion concrète. L'ambition est d'intégrer ces classes dans des outils d'aide à la décision, afin d'optimiser la valorisation des ressources végétales spontanées en milieux méditerranéens.

Tableau 1 : Description et composition chimique moyenne des 5 classes fonctionnelles, basés sur les prédictions à partir des spectres infrarouges sur échantillons secs (P : printemps et A : automne).

| Classes | Type de plantes prises et espèces | Nombre d'échantillons | CRPOT | NDF | ADF | ADL | IVOMD |
|---------|--|------------------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
| 1 | Feuilles <i>Graminées</i> mixtes et vertes + <i>Aphyllante</i> | TOTAL = 15 (6 P et 10 A) | 10,61 ^a | 64,92 ^b | 36,07 ^a | 8,50 ^a | 42,33 ^b |
| 2 | Tiges feuillées et tiges (<i>Filaire, olivier, fragon, genêt, salsepareille, genévrier...</i>) | TOTAL = 25 (11 P et 14 A) | 11,06 ^a | 47,29 ^a | 32,14 ^a | 13,73 ^a | 56,02 ^a |
| 3 | Jeunes feuilles, fleurs (<i>frêne, amélanchier, genêt, hêtre, myrtille, psoralée, sorbier, sureau, figuier, plantain, globulaire, sain foin</i>) | TOTAL = 52 (30 P et 22 A) | 14,77 ^b | 39,93 ^c | 25,17 ^a | 10,23 ^a | 67,85 ^c |
| 4 | Feuilles vertes et sénescents (<i>chêne blanc, chêne vert, pistachier, ronces, prunelier, cornouiller...</i>) | TOTAL = 36 (20 P et 16 A) | 11,61 ^a | 46,55 ^a | 30,23 ^a | 13,79 ^a | 54,38 ^a |
| 5 | Gousses, fruits, inflorescences (<i>ronces, sorbier...</i>) | TOTAL = 28 (23 P et 5 A) | 9,15 ^a | 46,74 ^a | 30,98 ^a | 11,58 ^a | 56,70 ^a |

CRPOT: protéines brutes, NDF: neutral detergent fibre, ADF: acid detergent fibre, ADL: acid detergent lignin, IVOMD: digestibilité enzymatique in vitro de la matière organique. P,A : printemps, automne.

References

- Agreil et al., 2004, Fourrages, 180 : 467–481
 Azambuja Filho et al., 2020, Rangeland Ecology & Management, 73 : 358–367.
 Daget et al., 1971, Annales Agronomiques, 5 : 5–41.
 Dochier et al., 2025, Book of Abstracts, Agrosilvopastoral Futures, 182–184.
 Hanisch et al., 2020, Journal of Applied Ecology, 57: 1535–1550.
 Silué et al., 2016, Option méditerranéenne, 114 : 71-74.

Détection d'anomalies dans les images hyperspectrales de feuilles de blé en utilisant Matrix Profile

Meryem Amrani Idrissi ^{1,2}

¹INRAE Montpellier – Équipe COMiC – UMR ITAP

²Université Sidi Mohamed Ben Abdellah – FST Fès – Filière ISACQ

Email : meryemamrani205@gmail.com

Mots-clefs : *Imagerie hyperspectrale, Séries temporelles, Matrix Profile, Anomalies spectrales, Agriculture de précision*

L'imagerie hyperspectrale constitue un outil puissant pour le suivi non destructif de l'état physiologique des plantes. Ce projet, réalisé dans le cadre d'un stage de fin d'études de six mois au sein de l'équipe COMiC (INRAE Montpellier), en collaboration avec le CIRAD et INRIA, avait pour objectif de détecter automatiquement des anomalies dans les images hyperspectrales de feuilles de blé, dans un contexte agronomique.

Pour cela, une approche originale a été adoptée, consistant à considérer chaque spectre hyperspectral de pixel comme une fausse série temporelle. Cette transformation permet d'appliquer des méthodes issues de l'analyse de séries temporelles à des données purement spectrales. En particulier, l'algorithme Matrix Profile a été utilisé pour mesurer les similarités locales entre ces "séries" spectrales et détecter des motifs discordants (discords), révélant des anomalies.

Les résultats ont montré que cette méthode permet de localiser efficacement des pixels atypiques, potentiellement liés à des stress biotiques ou abiotiques. Des analyses statistiques complémentaires (Q-Q plots, tests de normalité, méthodes robustes) ont permis de mieux caractériser les distances minimales et d'identifier les points aberrants. Cette stratégie ouvre des perspectives intéressantes pour la détection automatisée d'anomalies dans des données agronomiques, et pourrait être étendue à d'autres cultures ou dispositifs hyperspectraux.

Suivi en temps réel de l'évolution de la qualité organoleptique des pommes au cours d'un procédé thermomécanique par spectroscopie proche infrarouge

Ahmadou Moustapha DIENG¹, David PAGE¹, Patrice RELING, Christian GINIES¹, Raphaël PLASSON¹, Isabelle SOUCHON¹, Alexandre LECA¹, Sylvie BUREAU¹

¹ INRAE. Avignon Université. UMR SQPOV. F-84000 Avignon. France

Email : ahmadou-moustapha.dieng@inrae.fr

Mots clef : SPIR, Purée de pomme, prédiction, analyse non-destructive, chimiométrie

La forte variabilité et hétérogénéité des matières premières causée principalement par le changement climatique et les systèmes de culture, complique la prédiction de la qualité finale de la purée de pomme. Lors du traitement thermomécanique de la purée de pomme, les conditions de température et de broyage influencent la qualité finale via des modifications physico-chimiques complexes (changement de taille de particules et de viscosité^{1,2}, la destruction de composés tels que les pigments³ et le changement des odeurs^{4,5}) se produisant à une échelle de temps courte. Un contrôle de la qualité du produit est donc nécessaire, et parmi les outils et méthodes disponibles, la spectroscopie PIR (SPIR) est une solution prometteuse pour le suivi en temps réel grâce aux mesures à haut débit, non destructives, au contact de la matrice, avec la possibilité d'identifier des corrélations avec plusieurs de ces modifications de la qualité. C'est dans ce sens que nous avons exploré la SPIR comme moyen d'estimer des paramètres de qualité de la purée de pomme au cours de la cuisson.

Des purées de pommes (Golden delicious, récoltées en 2023) ont été préparées dans un procédé thermomécanique pilote semi-fermé (RoboQbo, Bentivoglio, Italie), suivant un plan d'expérience factoriel complet avec deux niveaux de température cible (70, 95 °C) et deux niveaux de vitesse de broyage (300, 1000 rpm) ainsi qu'un point central (82,5 °C/650 rpm). Les échantillonnages de référence ont été réalisés à quatre étapes clés : après 5 minutes à 25 °C (T₂₅), après 3 minutes à 45 °C (T₄₅), à l'atteinte de la température cible (T_s) et après 30 minutes de cuisson à la température cible (T_f). Le spectromètre (AvaSpec-NIR512-1.7-HSC-EVO, Avantes BV, Apeldoorn, Pays Bas) en contact avec la matrice via une fibre optique, a permis d'obtenir des spectres toutes les 10 secondes pendant la cuisson. Les spectres correspondant au temps de prélèvement (T₂₅, T₄₅, T_s et T_f) ont été sélectionnés pour étudier les corrélations avec les mesures de référence : viscosité, tailles de particules, composés d'arôme, et les paramètres biochimiques (teneur en eau, pH, acidité titrable et matière sèche soluble) . Les prétraitements des spectres (Savitzky-Golay (SG) et Standard Normal Variate (SNV)) et les régressions linéaires (PLS, leave-one-out) ont été réalisés avec les logiciels Chemflow et R Studio® (Version 2024.12.0.4676, pls package7).

Une Analyse en composantes principales (ACP) appliquée aux spectres SPIR a permis de discriminer les 5 conditions de procédé et chaque étape de prélèvement. Ceci permet de déduire que la SPIR est suffisamment sensible pour détecter les variations de composition et de structure des purées au cours de la transformation. De plus, les résultats obtenus par PLS sont encourageants notamment pour la prédiction de la viscosité (RMSECV% : 16%, R² : 0.81) et de la taille des particules (RMSECV% : 17%, R² : 0.75). La SPIR pourrait être ainsi utilisée pour suivre ces paramètres physiques, très largement utilisés pour définir la qualité finale d'une purée par des suivis en ligne à intégrer aux lignes de transformation industrielles.

Remerciements :

Bourse de thèse octroyée par INRAE Avignon et Avignon Université. Remerciements aux collègues du laboratoire : Marielle Boge et Éric Piétri

References

- (1) Lan, W.; Renard, C. M. G. C.; Jaillais, B.; Buergy, A.; Leca, A.; Chen, S.; Bureau, S. Mid-Infrared Technique to Forecast Cooked Puree Properties from Raw Apples: A Potential Strategy towards Sustainability and Precision Processing. *Food Chemistry* 2021, 355, 129636. <https://doi.org/10.1016/j.foodchem.2021.129636>.
- (2) Buergy, A.; Rolland-Sabaté, A.; Leca, A.; Falourd, X.; Foucat, L.; Renard, C. M. G. C. Pectin Degradation Accounts for Apple Tissue Fragmentation during Thermomechanical-Mediated Puree Production. *Food Hydrocolloids* 2021, 120, 106885. <https://doi.org/10.1016/j.foodhyd.2021.106885>.
- (3) Rios, J. J.; Fernández-García, E.; Mínguez-Mosquera, M. I.; Pérez-Gálvez, A. Description of Volatile Compounds Generated by the Degradation of Carotenoids in Paprika, Tomato and Marigold Oleoresins. *Food Chemistry* 2008, 106 (3), 1145–1153. <https://doi.org/10.1016/j.foodchem.2007.07.045>.
- (4) Bonneau, A.; Boulanger, R.; Lebrun, M.; Maraval, I.; Valette, J.; Guichard, É.; Gunata, Z. Impact of Fruit Texture on the Release and Perception of Aroma Compounds during in Vivo Consumption Using Fresh and Processed Mango Fruits. *Food Chemistry* 2018, 239, 806–815. <https://doi.org/10.1016/j.foodchem.2017.07.017>.
- (5) Lara-Valderrama, G.; Nagaya, M.; Uemura, K. Quality Characterization of Apple Puree Processed by High Frequency Pulse Continuous Heating. *Food Chemistry Advances* 2023, 2, 100228. <https://doi.org/10.1016/j.focha.2023.100228>.
- (6) Posit team (2024). RStudio: Integrated Development Environment for R. Posit Software, PBC, Boston, MA. URL <http://www.posit.co/>.
- (7) Liland K, Mevik B, Wehrens R (2024). `_pls: Partial Least Squares and Principal Component Regression_`. R package version 2.8-4, <<https://CRAN.R-project.org/package=pls>>.

Prédiction du potentiel de minéralisation du carbone au sol ($C_{\text{min}91}$) sur des digestats de méthanisation grâce à la spectroscopie proche infra-rouge

Jessica Do Souto Soeiro¹, Eric Latrille¹, Bastien Zennaro², Julie Jimenez¹

¹ INRAE LBE

² INRAE METYS

Email : jessica.do-souto-soeiro@inrae.fr ; julie.jimenez@inrae.fr

Mots-clefs : SPIR, méthanisation, digestats, carbone minéralisable au sol, indicateur

Contexte et objectifs

La méthanisation est une filière en plein essor répondant à un double enjeu de transition énergétique et agroécologique, puisqu'elle permet à la fois de produire une énergie renouvelable via la production de méthane et un fertilisant organique via la production de digestats. Cependant son optimisation est nécessaire pour permettre sa durabilité via l'étude des synergies et antagonismes existant sur l'utilisation de la matière organique pour produire à la fois énergie et fertilisant organique.

Des indicateurs existent pour caractériser les performances liées à la méthanisation comme le potentiel méthanogène des intrants (BMP)^[1] et leur potentiel de minéralisation de l'azote au sol (BNP)^[2] ainsi que la mesure du potentiel de minéralisation du carbone ($C_{\text{min}91}$) (*norme AFNOR XPU 44-163 – Caractérisation de la matière organique par la minéralisation potentielle du carbone*) et de l'azote organique au sol, qui sont obtenus en laboratoire par des méthodes en conditions contrôlées. Toutefois ces méthodes sont longues et coûteuses, empêchant ainsi un suivi optimal du procédé en temps réel. C'est dans ce cadre que le projet MethasolCN (3BCar, 2024-2025) se propose d'utiliser des techniques de caractérisation fine et rapide de la matière organique (C et N), telles que la spectroscopie proche infra-rouge, pour prédire ces indicateurs et *in fine* développer des méthodes prédictives rapides et peu coûteuses, en adéquation avec les attentes de la filière de méthanisation, comme le FlashBMP développé pour la prédiction du BMP^[3].

Lors de cette étude, un focus est réalisé sur la prédiction du potentiel de minéralisation du carbone au sol ($C_{\text{min}91}$) dans les digestats grâce à la spectroscopie proche infra-rouge. Les sols, plus important réservoir de carbone de la planète, sont capables de stocker du carbone sous la forme de matière organique et le rejeter sous forme de CO_2 dans l'atmosphère. Ils sont donc capables de jouer un rôle important dans l'atténuation du changement climatique. Le carbone minéralisable au sol est donc la fraction de carbone organique pouvant être transformée en CO_2 grâce à l'activité microbienne. Cette matière organique est également très importante puisqu'elle représente l'indicateur principal de la qualité des sols, à la fois pour des fonctions agricoles mais aussi pour des fonctions environnementales. Dans le cas des digestats, analyser le $C_{\text{min}91}$ permet d'évaluer la qualité du carbone résiduel qu'il soit labile (carbone facilement biodégradable) ou plus stable (carbone difficilement biodégradable). En d'autres termes, sa détermination permet de comprendre si un digestat permet de nourrir les micro-organismes à court terme et à atténuer la dégradation des sols à moyen/long terme.

Matériel et méthodes

Pour cette étude les valeurs de $C_{\text{min}91}$ de référence ont été obtenues en laboratoire selon la norme XPU 44-163. La base de données recense 131 digestats au total, dont 110 digestats faisant partie du jeu de calibration et 21 faisant partie du jeu de validation externe. Chaque digestat est alors analysé grâce à la

spectroscopie proche infra-rouge via le dispositif NIRFlex N-500 solids (Büchi) qui a l'avantage d'être rapide et non destructive. L'échantillon, préalablement broyé et lyophilisé, peut être analysé grâce à l'excitation des différentes liaisons entre molécules par une source qui émet entre 1000 et 2500 nm. Les spectres qui sont acquis grâce à ces mesures peuvent être liés ensuite à des valeurs de référence, dans notre cas, le potentiel de minéralisation du carbone de chaque digestat étudié^[4]. Les données spectrales sont alors rassemblées pour subir un traitement chimiométrique sur le logiciel Chemflow^[5] et grâce à une régression Partial Least Squares (PLS), il est possible alors de composer un jeu de calibration et un jeu de prédiction, et développer un modèle de prédiction de $C_{\min 91}$, soit une méthode plus rapide pour qualifier les caractéristiques analytiques d'un digestat^{[6],[3]}.

Résultats

Une grande variabilité des types de digestats définis par leurs intrants et fractions constitue la base de données, avec des valeurs de $C_{\min 91}$, comprises entre 1,1 % C_{org} et 68,27 % C_{org} , permettant ainsi au modèle de se construire sur une assez large gamme de valeurs de référence. Les spectres ont été prétraités premièrement avec Savitsky-Golay dérivée 1 et degré de polynôme 2 et normalisés grâce à un prétraitement SNV, des prétraitements également proposés par Zennaro et al.^[4] et qui permet d'obtenir les meilleurs résultats. Ces spectres prétraités ont ensuite subi une régression PLS, permettant d'obtenir un RMSEC de 8,26 % C_{org} et un RMSECV de 9,48 % C_{org} pour le jeu de calibration pour 5 variables latentes. Pour le jeu de validation, un RMSEP de 10,13 % C_{org} , une erreur jugée acceptable car assez proche de RMSEC et du SEL de 6,49 % C_{org} . De plus, les RPD des jeux de cross-validation et validation ont été calculés, ayant chacun un RPD de 1,45 et 1 respectivement, mais les données de calibration ne suivant pas une loi normale, le RPIQ^[7] a en plus été calculé (0,96 et 1,50 respectivement), donnant un aperçu de la qualité des modèles plus cohérent avec les données.

Perspectives

Par la suite, l'objectif serait d'appliquer la méthode utilisée sur la prédiction de $C_{\min 91}$ à la prédiction de l'azote minéralisable au sol, un indicateur moins étudié dans la littérature par rapport au carbone. *In fine*, le but serait d'aller vers la prédiction sur des digestats frais afin de s'affranchir de la préparation de ces derniers lors de la prédiction sur du broyé et lyophilisé.

References

- [1] <https://doi.org/10.1007/s11157-004-2502-3>
- [2] <https://doi.org/10.1016/j.biortech.2018.04.085>
- [3] <http://dx.doi.org/10.1016/j.biortech.2010.10.044>
- [4] <https://doi.org/10.1016/j.jenvman.2022.115393>
- [5] <https://vm-chemflow-francegrille.eu/root/login?redirect=%2F>
- [6] <http://dx.doi.org/10.1016/j.soilbio.2010.09.036>
- [7] <https://doi.org/10.1016/j.trac.2010.05.006>

Authentification des mélanges farineux blé-légumineuses par fusion multiblocs des spectres proche et moyen infrarouge

^{1,3} OUHADDOU Rachid, ¹ MEGHAR Karima, ¹ CHOCHOIS Vincent, ² LESNOFF Matthieu

¹ CIRAD, PERSYST/UMR QUALISUD, 34090 Montpellier – France

² CIRAD, UMR SELMET, 34980 Montferrier-sur-Lez – France

³ USMBA, FST, 30 000 FES – Maroc

Email : rachid.ouhaddou@cirad.fr

Mots-clefs : Chimie métrique, analyse multibloc, spectroscopies infrarouge, PLS.

Au cœur des stratégies agroécologiques et nutritionnelles actuelles, les mélanges de farine de blé et de légumineuses occupent une place de choix. La demande des filières et des consommateurs pour des produits à haute valeur ajoutée a favorisé l'émergence de formulations intégrant notamment la farine de lentille. Ces mélanges répondent à une double exigence : diversifier les profils protéiques et offrir des farines plus saines, tout en valorisant des cultures à faible empreinte environnementale.

Cette étude s'inscrit dans le projet européen IntercropValues, qui vise à exploiter les bénéfices de la culture intercalaire. Plus précisément, l'objectif de cette recherche est de développer des modèles chimométriques permettant d'estimer la proportion de lentilles dans des mélanges farineux blé-lentilles, à partir de données spectrales Proche Infrarouge (SPIR) et Moyen Infrarouge (SMIR).

Cent cinq échantillons (5 variétés de lentilles × 21 proportions) ont été analysés avec des spectromètres SPIR ASD LabSpec® Pro (350-2500 nm) et FTIR Nicolet™ iS50R (4000–400 cm⁻¹). Après prétraitements spectraux (Savitzky-Golay, Detrend, SNV), des régressions PLS monobloc (sur SPIR ou SMIR) et multibloc (MB-PLS et SO-PLS) ont été calibrées. Les performances ont été évaluées selon deux stratégies de validation (répartition aléatoire et leave-one-variety-out), en retenant le RMSEP sur Test comme critère principal. Le modèle PLS sur SPIR seul atteint un RMSEP d'environ 1.89, tandis que SMIR seul atteint 3.18. Les approches multiblocs, notamment MB-PLS et SO-PLS (SMIR→SPIR), réduisent le RMSEP à 1.56 (validation aléatoire) et 2.55 (leave-one-variety-out), illustrant la complémentarité des spectres et la meilleure robustesse. La concordance au sein de la validation interne et externe (RMSEP_cv VS RMSEP_Test) atteste de l'absence de sur-ajustement. Ces résultats démontrent que la fusion multiblocs offre un outil fiable et rapide pour le contrôle qualité non destructif en agroalimentaire.

Real-Time Assessment of PEG-Induced Hydric Stress in Sunflowers Using Spectral Data and ASCA

Ingi Abdelmeguid^{1,2}, Daphné Heran², Maxime Ryckewaert³, Ryad Bendoula²

¹ Department of Botany and Microbiology, Faculty of Science, Helwan University, Cairo, Egypt

² ITAP, Univ Montpellier, INRAE, Institut Agro, Montpellier, France

³ AGAP Institut, CIRAD, INRAE, Institut Agro, Univ de Montpellier, Montpellier, France

Email: ingi.abdelmeguid@science.helwan.edu.eg

Keywords: Spectroscopy, Hydric stress, ASCA, Sunflowers

Spectroscopy and hyperspectral data are widely recognized non-destructive tools for monitoring plant health, enabling real-time detection of both biotic and abiotic stresses [1]. From a decision-making perspective, real-time assessment represents a transformative approach that saves time and resources while promoting more sustainable and environmentally friendly agricultural practices. However, a key challenge lies in selecting and applying the right analytical method capable of effectively unraveling the complexity of such data and delivering actionable insights for rapid decision support. Hydric stress (drought stress) is one of the most important abiotic factors limiting crop growth and development worldwide. It disrupts key physiological processes such as photosynthesis, transpiration, and nutrient uptake, ultimately compromising plant performance. In this study, we implemented a controlled environment experimental setup to better understand plant responses to drought-like conditions. To this end, polyethylene glycol (PEG) was used in our experimental setup, as it is commonly employed to simulate water stress under reproducible conditions.

The experiment was conducted over a five-day period to assess the effects of PEG-induced hydric stress and subsequent recovery in sunflower plants. Three out of six plants were treated with PEG, while the others served as untreated controls for the entire duration of the study, providing a baseline for comparison. The PEG treatment was applied over a period of two days, after which it was discontinued to allow for the observation of recovery dynamics. Throughout the five-day experiment, Visible-Near Infrared (VNIR) spectral data were collected from all plants at nine time points to monitor changes in their spectral signatures associated with the pre-stress, stress, and recovery phases. To analyze the resulting spectral data, we applied multivariate statistical approaches to extract patterns over time. Principal Component Analysis (PCA) was first used to explore the main sources of variance and visualize the plant responses to stress [2]. While PCA provided initial insight into the temporal dynamics of the stress response, it lacked the ability to isolate the effects of treatment and time. Therefore, we applied ANOVA–Simultaneous Component Analysis (ASCA), which is a method specifically designed to separate structured variation in experimental data [3]. By separating the contributions of PEG treatment, measurement time, and their interaction, ASCA enabled a clearer interpretation of the spectral responses associated with hydric stress progression and recovery.

These results demonstrate the potential of ASCA to effectively separate and interpret treatment and temporal effects in spectral data under controlled hydric stress conditions. The clear distinction observed between stressed and control plants, along with the indication of recovery pattern following PEG removal, highlights the potential of this approach for real-time stress assessment.

References

- [1] Mahlein, A.K., Steiner, U., Hillnhütter, C. et al. Hyperspectral imaging for small-scale analysis of symptoms caused by different sugar beet diseases. *Plant Methods* 8, 3 (2012). <https://doi.org/10.1186/1746-4811-8-3>
- [2] Wold, S., Esbensen, K., & Geladi, P. Principal Component Analysis. *Chemometrics and Intelligent Laboratory Systems* 2(1–3), 37–52 (1987). [https://doi.org/10.1016/0169-7439\(87\)80084-9](https://doi.org/10.1016/0169-7439(87)80084-9)
- [3] Smilde, A. K., Jansen, J. J., Hoefsloot, H. C. J., Lamers, R. J. A. N., van der Greef, J., & Timmerman, M. E. ANOVA–Simultaneous Component Analysis (ASCA): A new tool for analyzing designed metabolomics data. *Bioinformatics* 21(13), 3043–3048 (2005). <https://doi.org/10.1093/bioinformatics/bti476>

Potentiel et limites de l'utilisation de la SPIR foliaire pour l'identification taxonomique de spécimens d'herbier : étude de cas sur les *Annonaceae*

Khalil Boughalmi¹, Paola Santacruz¹, Lily Bennett¹, Denis Bastianelli², Laurent Bonnal², Samantha Bazan², Thomas L.P. Couvreur¹

¹*DIADE, Université de Montpellier, CIRAD, IRD, Montpellier, France*

²*SELMET, CIRAD, INRAE, Institut Agro, Montpellier, France*

Email : khalil.boughalmi@etu.univ-grenoble-alpes.fr

Résumé non disponible

La spectroscopie proche infrarouge et la biodiversité : illustrations pour l'aide à l'identification taxonomique

Gilles Chaix^{1,2,3}

¹ CIRAD - UMR AGAP Institut, Montpellier, France

² UMR AGAP Institut, Univ Montpellier, CIRAD, INRAE, Institut Agro, Montpellier, France

³ ChemHouse, Research Group, Montpellier, France¹

Email : gilles.chaix@cirad.fr

Mots-clefs : Taxonomie, classification, SPIR, anatomie, identification, herbiers

La taxonomie est la science de la classification des êtres vivants. Fondée par Carl von Linné au XVIII^e siècle, elle organise les organismes selon une hiérarchie de catégories emboîtées : domaine, règne, embranchement, classe, ordre, famille, genre et espèce. Cette classification repose sur des caractères morphologiques, génétiques et/ou comportementaux. La taxonomie permet de nommer, décrire et relier les espèces entre elles, facilitant ainsi l'étude de la biodiversité. Elle se base sur des clefs d'identification qui requièrent l'expertise anatomique et les échantillons des organismes entiers ou de leurs différentes parties.

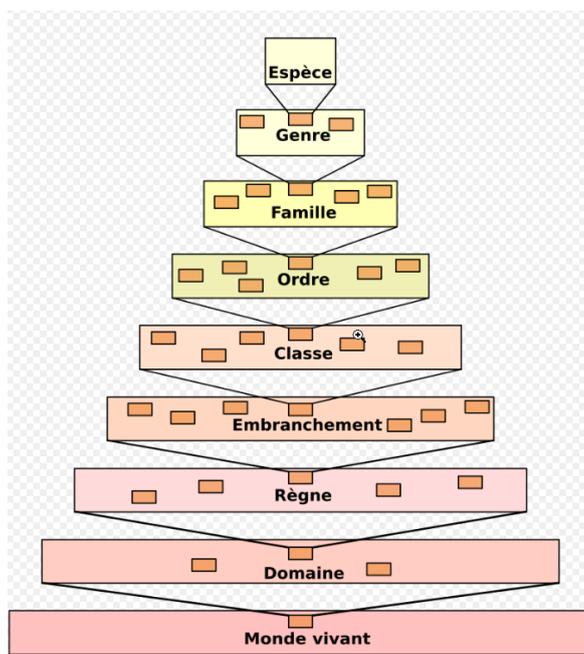


Figure 1. Organisation de la hiérarchie des êtres vivants (<https://www.wikipedia.fr>)

Chez les plantes, il s'agit généralement des tiges, feuilles, inflorescences, fleurs, fruits, pollen. Les échantillons sont verts (fraichement récoltés ou observés sur sites) ou secs, généralement fixés sur un support, parfois stockés en solution. Une fois identifiés et nommés, un ou plusieurs spécimens sont conservés dans des collections spécifiques (herbiers, xylothèque, insectarium, ...).



Figure 2. Exemple de planche d'herbier de deux types de plantes (<https://www.mnhn.fr/fr>)

Le manque de spécialistes (botanistes, taxonomistes, ...), l'ampleur de la biodiversité (tout n'est pas connu !), la nature évoluant constamment, les révisions taxonomiques, ... conduisent à rechercher des méthodes complémentaires de différentes disciplines ou pouvant se substituer, du moins partiellement, aux approches morphologiques qui restent essentielle. Pour les principales, on peut citer les groupes de méthodes :

- biochimiques pour l'analyse de composés secondaires (chimiotaxonomie : alcaloïdes, flavonoïdes,
- moléculaires (Barcoding séquençage ciblés, marqueurs moléculaires, séquençage complet, ...),
- écologiques : écologie comparée (climat, altitude, sol, niche écologique), biogéographie (distribution spatiale, versant, habitat...).

Parmi les méthodes biochimiques, on peut inclure les méthodes spectroscopiques. La spectroscopie proche infrarouge (SPIR) en fait partie, et est donc à l'origine de cette présentation introductive à la session « Nirs et Taxonomie » de ces Rencontres, qui va traiter essentiellement le sujet pour les végétaux, en l'absence de présentation sur d'autres organismes.

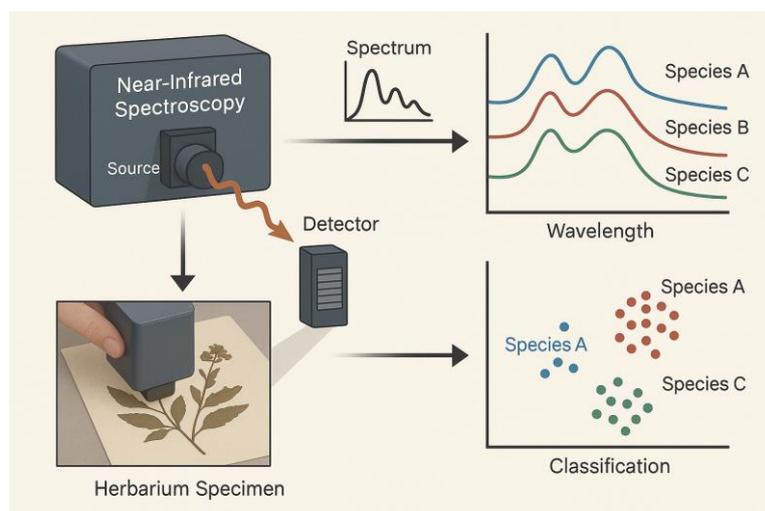


Figure 3. Illustration synthétique de l'utilisation de la SPIR (générée par ChatGPT)

Pour ses mesures rapides, non destructives et reproductibles ne nécessitant que très peu de préparation des échantillons, la SPIR est de plus en plus utilisée pour l'identification taxonomique à partir de spécimens d'herbiers. En effet, chaque espèce possède un profil biochimique unique, qui se traduit par une signature spectrale spécifique qui peut être exploitée pour comparer, classifier des spécimens. Cette approche est particulièrement pertinente pour les spécimens anciens, chez lesquels l'ADN est souvent dégradé et les caractères morphologiques altérés. De plus, la SPIR permet de collecter des données fiables sans endommager les échantillons, ce qui est essentiel pour préserver l'intégrité des collections patrimoniales.

L'utilisation de la SPIR dans les herbiers ouvre ainsi de nouvelles perspectives pour la taxonomie, la révision systématique et la valorisation des collections botaniques. Elle contribue également à améliorer la connaissance de la biodiversité végétale conservée dans les archives scientifiques parfois uniques.

Couplée à des analyses multivariées ou à des algorithmes d'apprentissage automatique, la SPIR est aussi particulièrement utile dans la discrimination d'espèces morphologiquement proches, pour détecter et corriger des erreurs taxonomiques. En tant que méthode complémentaire à la morphologie classique et à la biologie moléculaire, la SPIR contribue au renforcement des capacités d'inventaire, de surveillance et de conservation de la biodiversité.

En conséquence, elle constitue un outil complémentaire à la taxonomie traditionnelle, et favorise une meilleure exploitation des herbiers dans les études de systématique. Les herbiers deviennent des ressources encore plus précieuses pour la recherche en biodiversité.

En dépit de ces perspectives très prometteuses qui valorisent encore la SPIR dans un nouveau domaine, il existe toutefois quelques freins tant sur la mesure que sur l'analyse des données et l'interprétation des résultats. Quid du vieillissement des échantillons et donc de l'effet de l'âge sur les signaux ? Combien de mesures spectrales par échantillons ? Spectroscopie vs imagerie hyperspectrale ? Des questions qui accaparent actuellement les chercheurs et les gestionnaires d'herbiers. De même, au niveau des analyses de données et approches, les recherches actuelles portant sur l'apprentissage automatique rapprochent les communautés, mais à mon sens et en parcourant la bibliographie, encore trop peu.

En conclusion, la SPIR s'intègre comme un outil complémentaire aux méthodes classiques d'identification taxonomique, l'authentification des spécimens et la valorisation des collections botaniques. Elle contribue à améliorer l'intérêt et l'exploitation scientifique des herbiers, véritables archives de la biodiversité végétale. Mais c'est aussi le cas, de manière générale, dans le monde vivant.

La Spectroscopie proche infrarouge au service de la détermination des essences forestières

^{12*}Nassim BELMOKHTAR, ¹²Nathalie BOIZOT, ¹²Orlane TOUZET, ¹Arnaud DOWKIW, ¹Remy GOBIN

¹ INRAE, UMR INRAE-ONF BioForA 0588, 45075 Orléans – France

² INRAE, Plateforme Phénobois, 45075 Orléans – France

Email : nassim.belmokhtar@inrae.fr

Mots-cléf : SPIR, forêt, arbre, résineux, feuillus, PLS-DA

L'identification des espèces d'arbres au sein d'un même genre est un enjeu à la fois pour la gestion forestière courante et l'adaptation des forêts aux changements globaux. L'identification des espèces d'arbres forestiers repose essentiellement sur des critères morphologiques observés sur le terrain. Cette identification peut s'avérer complexe, même pour les experts, notamment hors saison végétative, en l'absence de feuilles ou de fruits.

La spectroscopie proche infrarouge (SPIR) a révélé un fort potentiel pour identifier diverses espèces forestières, aussi bien en laboratoire que sur le terrain avec des appareils portables. A travers plusieurs projets de recherche, la plateforme Phénobois a déployé cette technologie avec succès pour identifier des espèces d'intérêt appartenant à plusieurs genres, tels que les chênes, frênes, peupliers, ormes, pin laricio et mélèzes. Nous avons étudié les capacités prédictives de cette technologie sur différents organes des arbres : feuilles ou aiguilles, bourgeons, ainsi que le bois (carottes et rondelles).

Nos études ont permis de valider l'utilisation de la SPIR pour la détermination des espèces forestières et ouvrent la voie pour l'identification de leurs provenances.

Perspectives sur l'évolution des données de réflectance foliaire à partir de spécimens d'herbier dans une grande famille de plantes tropicales (*Annonaceae*)

Thomas L.P. Couvreur¹, Carlos-Rodrigues-Vaz^{1,2}, Khalil Boughalmi¹

¹ *DIADE, Université de Montpellier, CIRAD, IRD, Montpellier, France*

² *Institut de Systématique, Evolution, Biodiversité (ISYEB), Muséum National d'Histoire Naturelle-CNRS-SU-EPHE-UA, 57 rue Cuvier, CP 39, 75231 Paris, Cedex 05, France*

Email: thomas.couvreur@ird.fr

Résumé non disponible

Utiliser la SPIR sur des échantillons d'herbiers pour l'identification taxonomique ?

Samantha Bazan ^{1,2}, Elodie Baby ^{1,2}, Denis Bastianelli ^{1,2}, Alexia Bennert ³, Laurent Bonnal ^{1,2}, Matthieu Lesnoff ^{1,2}, Jennifer Mersni ^{1,2}, Morgane Dendoncker ³

¹ CIRAD, UMR SELMET, F-34398 Montpellier, France

² SELMET, Univ Montpellier, CIRAD, INRAE, Institut Agro, Montpellier, France

³ Université catholique de Louvain, Earth and Life Institute, Environmental Sciences, 1348, Louvain-la-Neuve, Belgium

Email: samantha.bazan@cirad.fr

Mots-clefs : Botanique, Herbar, Taxonomie, Analyse discriminante, SPIR

Les collections d'herbiers constituent des sources inestimables de matériel et de données sur la biodiversité végétale. Mais ce matériel à la fois scientifique et patrimonial reste fragile. La SPIR est une méthode déjà largement employée pour effectuer des mesures rapides et non destructives sur des échantillons végétaux. Cet outil semble donc particulièrement adapté pour obtenir de nouveaux types de données à partir d'échantillons d'herbiers.

Les applications de la SPIR sur les herbiers sont encore récentes et exploratoires. La spectrométrie peut apporter des éléments intéressants sur différents aspects : chimie, écologie, physiologie, taxonomie, etc. L'une des possibilités qui semble particulièrement intéressante est de distinguer des espèces voisines sur le plan taxonomique. Pour évaluer le potentiel de la SPIR dans cette perspective, nous avons mené deux projets parallèles sur des taxons très différents.

Le premier se concentre sur le genre *Pistacia* et plus particulièrement sur les espèces méditerranéennes (*P. lentiscus*, *P. terebinthus*, *P. vera*, *P. atlantica*). C'est un modèle d'étude intéressant pour travailler sur des espèces proches qui peuvent même s'hybrider (*P. x saportae*). Facile à trouver et à identifier à partir des feuilles, il nous a permis de travailler à la fois sur un grand nombre d'herbiers anciens et sur des collectes récentes réalisées pour notre étude. Les premiers résultats montrent de très bonnes performances en matière de discrimination, à tel point que certaines erreurs d'identification sur des herbiers historiques ont pu être détectées. Les 4 espèces de *Pistacia* sont discriminées avec un taux de réussite de 98%. Entre les 2 espèces les plus proches et leur hybride, la discrimination est aussi très bonne et les erreurs sont toujours entre l'hybride et une des deux espèces. Enfin la prédiction sur les nouveaux échantillons à partir d'une calibration faite sur les herbiers historiques donne un taux de succès de 87%.

Afin de permettre à nos partenaires scientifiques de tester eux-mêmes ces questions, nous avons mis à disposition librement notre jeu de données *Pistacia* : près de 7400 spectres issus de 160 spécimens d'herbiers historiques et de 51 nouvelles collectes. Il contient les données spectrales brutes ainsi qu'une version pré-traitée prête à l'emploi, le tout accompagné des métadonnées liées aux herbiers.

Ce jeu de données est disponible dans le dataverse du Cirad :

<https://doi.org/10.18167/DVN1/J1TZZN>



En parallèle, nous avons également mené une étude sur le genre *Combretum* à l'aide d'un jeu de données très complet d'échantillons d'Afrique de l'Ouest et d'Afrique centrale sur 13 espèces couvrant 275 spécimens d'herbiers. Nous avons analysé les relations entre la taxonomie, les données spectrales, les traits fonctionnels mesurés (Specific Leaf Area, Leaf thickness) et le gradient pluviométrique. Lorsqu'on analyse la discrimination taxonomique sur l'ensemble des données on obtient un succès du modèle de 73%, cependant en distinguant les espèces par zone bioclimatique, les taux de réussite augmentent fortement de 89 à 94%.

Ces premiers résultats très encourageants nous permettent d'envisager des applications concrètes de la SPIR en botanique et en écologie, en valorisant les collections d'herbiers mais aussi en les enrichissant de nouvelles données, caractérisant la composition chimique de l'échantillon.

Posters

A new light in Herbaria: How can NIRS be applied to Herbarium specimens for identification?

Samantha Bazan^{1,2}, Morgane Dendoncker³, Elodie Baby^{1,2}, Denis Bastianelli^{1,2}, Jennifer Mersni^{1,2}, Simon Taugourdeau^{1,2}, Laurent Bonnal^{1,2}

¹ CIRAD, UMR SELMET, F-34398 Montpellier, France

² SELMET, Univ Montpellier, CIRAD, INRAE, Institut Agro, Montpellier, France

³ Université catholique de Louvain, Earth and Life Institute, Environmental Sciences, 1348, Louvain-la-Neuve, Belgium

Email : samantha.bazan@cirad.fr

Keywords: Herbarium, Identification, near infrared spectroscopy, analytical methods, taxonomy

Near infrared spectroscopy (NIRS) is a promising tool for valorizing herbarium specimens as it allows the gathering information on plants without damaging valuable and historical samples. The CIRAD herbarium is currently testing the use of NIRS with various approaches. One project is experimenting with using NIRS as a tool for systematic discrimination. Assuming that intraspecific variations in the spectrum are smaller than interspecific variations, it may be possible to identify specimens in herbarium within some genus, using appropriate calibration.

A test was carried out on the 4 Mediterranean species of *Pistacia* : *P. terebinthus*, *P. lentiscus*, *P. vera* and *P. atlantica* and the hybrid *P. saportae* (*P. terebinthus* x *P. lentiscus*). A first dataset involved specimens dating back 150 years and a second dataset gathered specimens collected recently. Using these two datasets, we evaluate the ability of NIR models built on older collections to discriminate recent samples or, in opposite, the possibility of discriminating older samples with recent collections.

The 4 species of *Pistacia* were distinguished with a success rate of 98%. *P. lentiscus* and *P. terebinthus* were completely separated, while the hybrid appeared between them as expected. Some samples were misclassified, as *P. saportae* has a high phenotypic variability. During this study, a mis-identified sample from 1864 was also spotted by NIRS. When the calibration made from herbarium specimens was applied to new samples, the classification success was 68% only, partly due to the fact that new collections were designed to include non-typical samples such as very young or senescent leaves, etc.

In conclusion, NIRS can be used on herbarium specimens without damaging them and provides valuable information on the samples. In this *Pistacia* example, species can be well distinguished by their spectrum. The spectra of the hybrid are intermediate between those of the two parent species. NIRS seems to account for the taxonomic link and proximity. Finally, new samples of *Pistacia* can be identified with a sample of old herbarium references.

(ce poster a été initialement présenté à l'International Botany Congress IBC Madrid, 21-27/07/2024)

Machine learning methods for sugar quantification in grapes based on near-infrared hyperspectral imaging

Julien BOYER¹, Jordane POULAIN¹, Sylvie ROUSSEL¹, Nicolas SAURIN², Maxime RYCKWAERT³, Ryad BENDOULA³, Carole FEILHES⁴ and Eric SERRANO⁴

¹ Ondalys, 4 rue Georges Besse, 34830 Clapiers, France

² ITAP, Univ Montpellier, INRAE, Institut Agro, Montpellier, France

³ UE Pech Rouge, Univ Montpellier, INRAE, Gruissan, France

⁴ IFV Sud-Ouest, V'innopôle, 1920, route de Lisle/Tarn, 81310 Peyrole, France

Email : jpoulain@ondalys.fr ; sroussel@ondalys.fr

Keywords: Hyperspectral Imaging (HSI), Machine Learning, quantification, sugar, grape bunch

Hyperspectral imaging [1] has applications in many fields, including agriculture, environment, medicine and industry. It can be used to classify objects according to their composition, or to quantify compounds present on the surface and show their spatial distribution. This last case is often challenging as the reference is not known for each pixel. Image processing is even more important in this case to extract the Region of Interest (ROI) to build the model but also to be able to apply it on new images.

This scientific study presents the methodology for analyzing hyperspectral images for grape bunch maturity prediction directly in vineyards. In this example, the parameter to be predicted is the sugar content (in °Brix) on each whole bunch of grapes, providing an average measurement of sugar content per image.

The database includes 131 hyperspectral images of 2 red grape varieties (Fer Servadou N. and Syrah N.) acquired with the SPECIM IQ (SPECIM - Konica Minolta) portable visible – near-infrared hyperspectral camera, with a spectral range of 400 to 1000 nm. Grape bunches are measured in the vineyard with the natural illumination of sunlight and without a uniform background. As a result, the image bank is quite heterogeneous. Extracting the object of interest (single bunch) is therefore a complex matter and a specific methodology and algorithm had to be created.

A methodology to automatically process the images has been developed. It consisted of different steps: 1) spectral preprocessing with row normalization, 2) extraction of the Region of Interest (ROI) by a Soft Independent Modeling of Class Analogy (SIMCA) [2] approach, combined with classical thresholding, and 3) average of the pixels.

The results of the quantitative models built on the average spectra shown that a Support Vector Machines (SVM) [3] regression model is significantly outperforming Partial Least Squares (PLS) regression [4].

The automatic tool for image processing has been successfully applied to predict the sugar content of the whole image ROI to visualize the pixel spatial distribution of the sugar in the bunch.

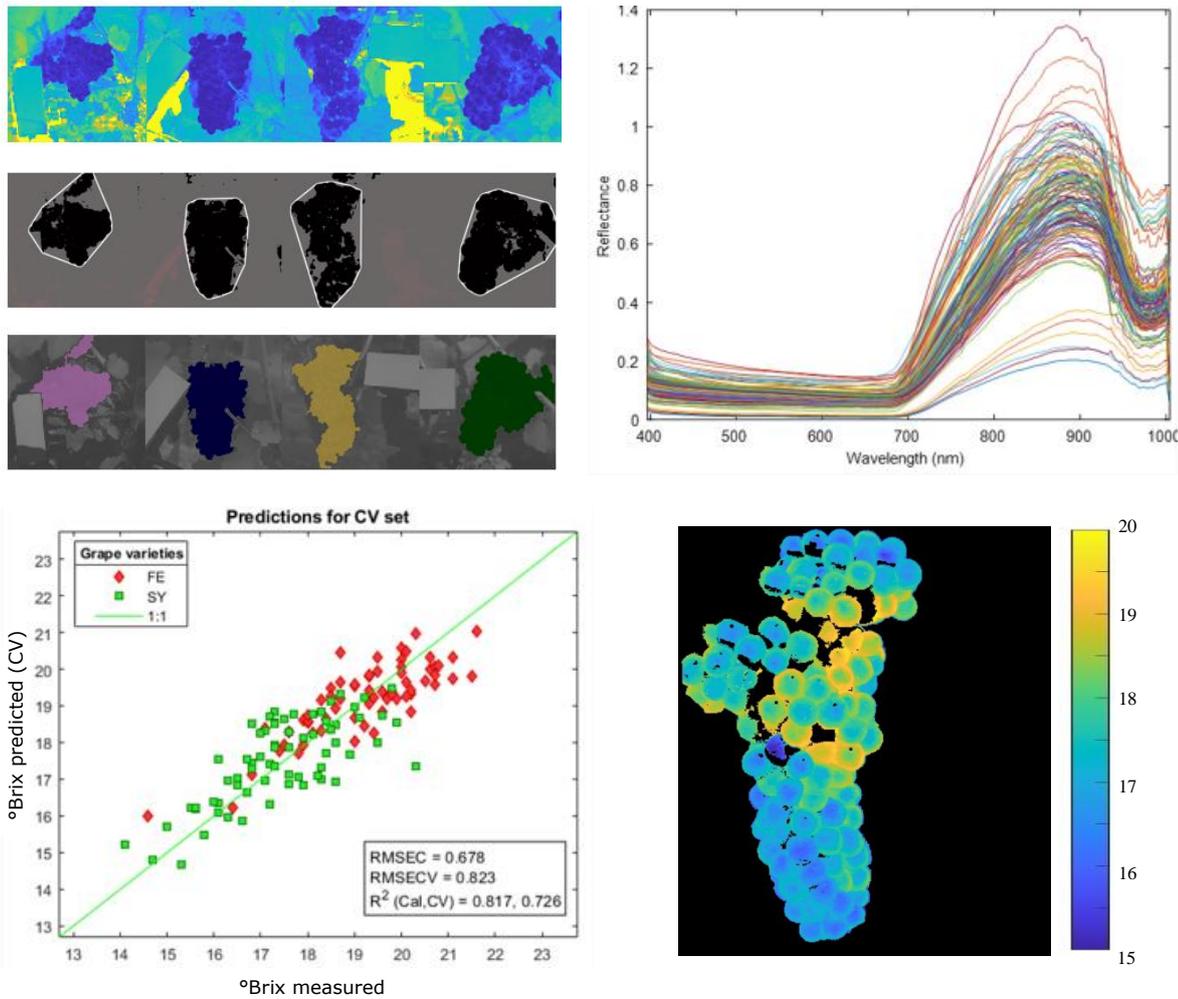


Figure 1: Example of extraction of the ROI on 4 images (top-left), resulting raw averaged spectra for all the images of the whole database (top-right), predictions in cross-validation of the average sugar content for each image of the SVM model (bottom-left) and predictions for each pixel for one of the images (bottom-right).

This HSI data analysis study was carried out by Ondalys in collaboration with INRAE (UMR ITAP, Team COMIC and UE Pech Rouge) and IFV (IFV South West) as part of the VINIoT project, an Interreg SUDOE project funded by the European Regional Council.

Reference

- [1] Amigo, José Manuel, éd. *Hyperspectral imaging. Data handling in science and technology* 32. Amsterdam, Netherlands; Cambridge, MA, United States: Elsevier, 2020.
- [2] S. Wold, M. Sjostrom, SIMCA: A method for analyzing chemical data in terms of similarity and analogy. In Kowalski, B.R., ed., *Chemometrics Theory and Application*, American Chemical Society Symposium Series 52, 1977, p. 243-282.
- [3] C. Cortes, V. Vapnik, Support-vector networks. *Mach. Learn.* 20, 273–297 (1995).
- [4] H. Wold. Estimation of principal components and related models by iterative least squares. In P.R. Krishnaiah (Ed.), *Multivariate Analysis*. 1966. pp.391-420.

Étude de la Diversité Génétique et de la Phénologie du Ray-Grass Anglais (*Lolium perenne* L.) par Spectroscopie Proche Infrarouge

Maiwen Lambert¹, Charlène Barotin¹, Fabien Surault¹ et Marie Pégard¹

¹ INRAE P3F, 86600 Lusignan, France

Email : lambertmaiwen@gmail.com - marie.pegard@inrae.fr

Mots-clefs : Prédiction phénomique, NIRS, Phénologie, Interaction Genotype Environment

Dans cette étude, on s'intéressera à la diversité génétique du ray-grass anglais (*Lolium perenne* L.), plante emblématique des prairies, ainsi qu'à la phénologie et aux facteurs pouvant l'influencer.

Dans notre étude, nous nous servons de la spectroscopie proche infrarouge via le NIRS portable afin de caractériser la diversité génétique du ray-grass anglais. Cette caractérisation se fera sur des individus appartenant à des populations de ray-grass anglais représentatives de la diversité phénologique de l'espèce et installées dans deux lieux différents. Ces deux lieux sont équipés de capteurs de température et d'humidité du sol et de l'air.

Premièrement, au pollinarium de Poitiers, ce lieu bénéficie d'un microclimat urbain, d'une protection contre les vents grâce à des murs en pierre et d'une orientation plein sud, autant de facteurs pouvant affecter le développement et la précocité de l'épiaison et de l'émission du pollen. Il regroupe les plantes allergènes de la commune de Poitiers et de ses alentours, incluant le ray-grass anglais. Deuxièmement, à l'INRAE, sur la commune de Lusignan, la pépinière est exposée aux aléas météorologiques, bénéficiant d'un ensoleillement constant de par sa localisation en plein champ.

Cette étude représente un test de prédiction phénomique sur le ray-grass anglais, une première pour cette espèce. Pour cela, nous testons la prédiction de différents traits phénotypiques liés à la phénologie, tels que la date d'épiaison, la date de début et de fin d'émission de pollen et certains caractères de croissance. Ces prédictions seront réalisées à l'aide d'un NIRS portable. L'établissement des populations dans deux lieux nous permettra de rechercher quelles conditions climatiques influencent l'émission de pollen après épiaison.

Nos résultats favoriseront le développement de la sélection phénomique comme alternative au génotypage dans le cadre de la sélection végétale.



Sponsors de l'association HélioSPIR en 2025

Sponsors institutionnels



Sponsors privés

Sponsors Platinum



Sponsors Silver



Association HélioSPIR
Réseau de spectroscopie proche infrarouge
www.heliospir.net