

COMPARAISON DES PERFORMANCES DES MODÈLES D'ANALYSE DISCRIMINANTE PLS ET SVM

- APPLICATION SUR DES SPECTRES PROCHE INFRAROUGE DE RACINES ET DE PAILLES DE FROMENT D'HIVER -

Contexte: L'imagerie hyperspectrale proche infrarouge est de plus en plus utilisée pour analyser des produits agricoles. Cette méthode qui combine la spectroscopie proche infrarouge et l'imagerie permet d'enregistrer, pour chaque pixel de l'image, un spectre comprenant la réflectance à différentes longueurs d'ondes. Cette analyse permet d'appréhender l'hétérogénéité et la variabilité spectrale des échantillons [1]. Des algorithmes adéquats sont nécessaires pour analyser les larges volumes de données générés. Les performances des **modèles d'analyses discriminantes PLS** (*Partial Least Squares*) et **SVM** (*Support Vector Machines*) ont été comparées dans une étude qui avait pour objectif de discriminer des spectres de racines et de pailles de froment en vue de les quantifier [2].

Méthode

Matériel

Des échantillons de sol ont été prélevés sous une culture de froment et lavés à l'eau pour en extraire les racines et les résidus de culture enfouis.



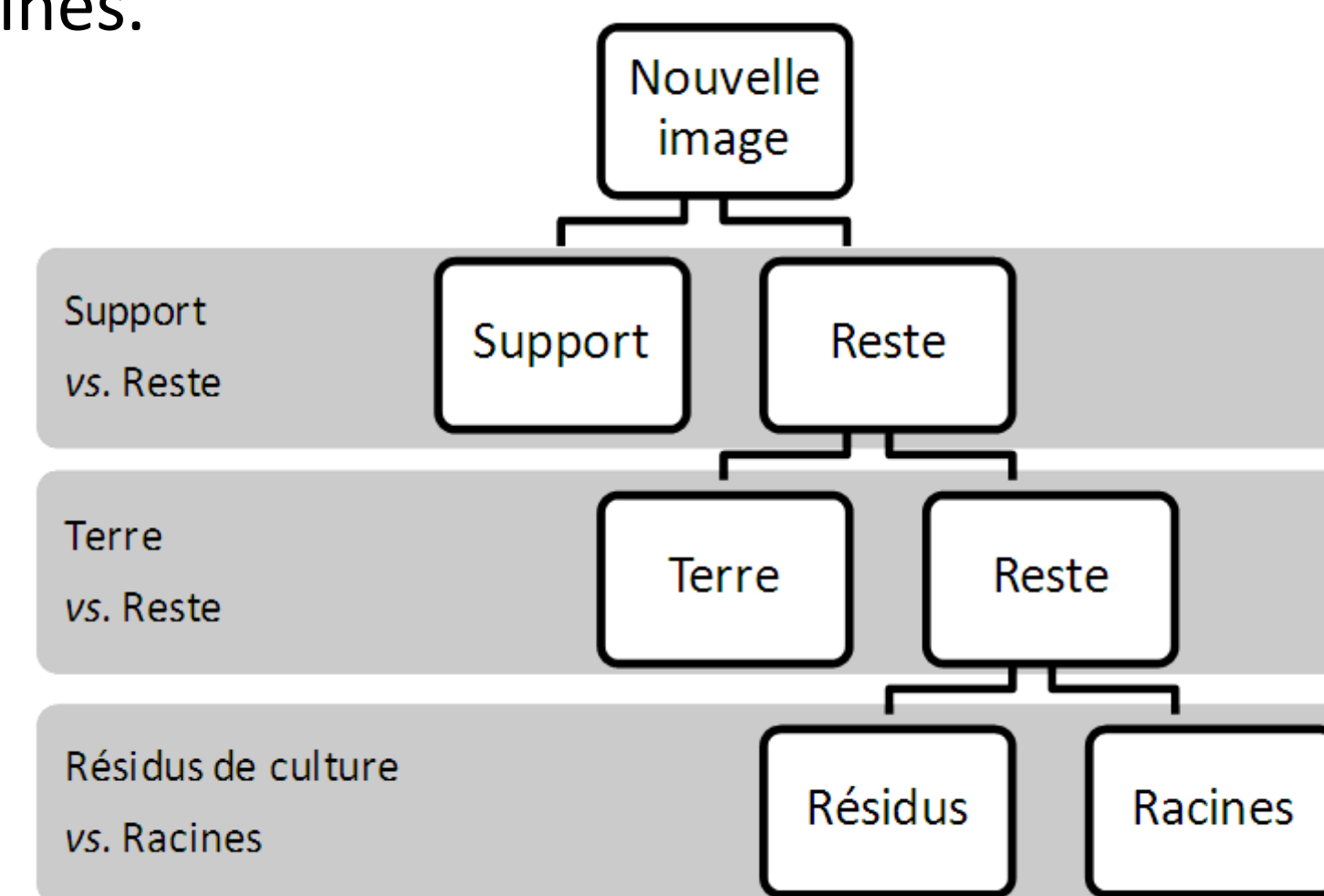
Acquisition des images

Les images sont acquises dans le proche infrarouge à l'aide d'une **camera hyperspectrale** de type "line-scan" positionnée au dessus d'une bande transporteuse sur laquelle sont posés les échantillons [3]. Un spectre complet de 209 longueurs d'ondes (1100-2498 nm) est enregistré pour chaque pixel de l'image.



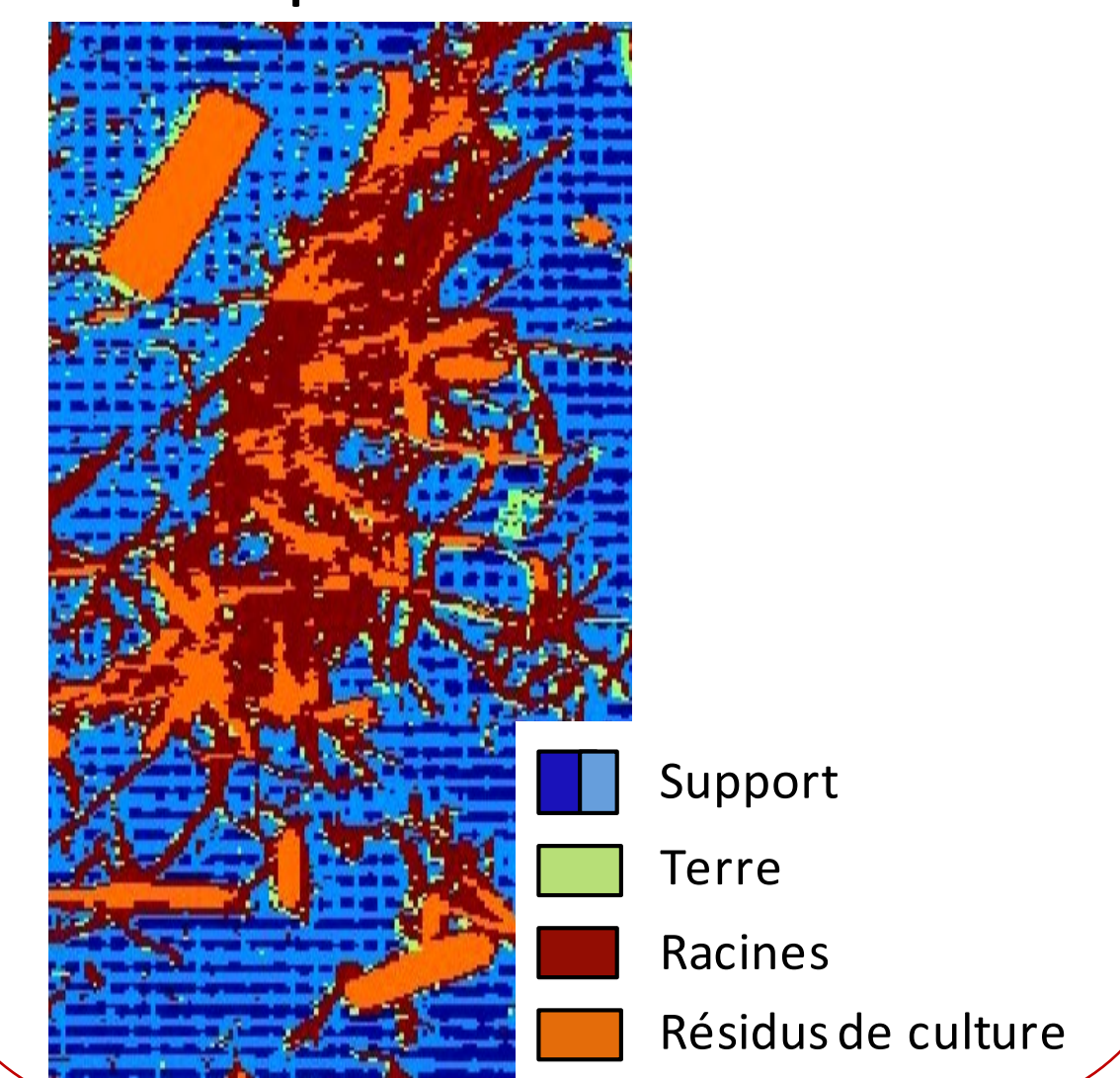
Discrimination des spectres

Les algorithmes d'analyse discriminante (PLS ou SVM) sont utilisés pour prédire la nature de chaque pixel de l'image sur base des spectres enregistrés. Un **arbre de classification dichotomique** composé de 3 algorithmes permet la séparation des spectres de l'image en 4 classes: le support, les résidus de terre, les résidus de culture et les racines.

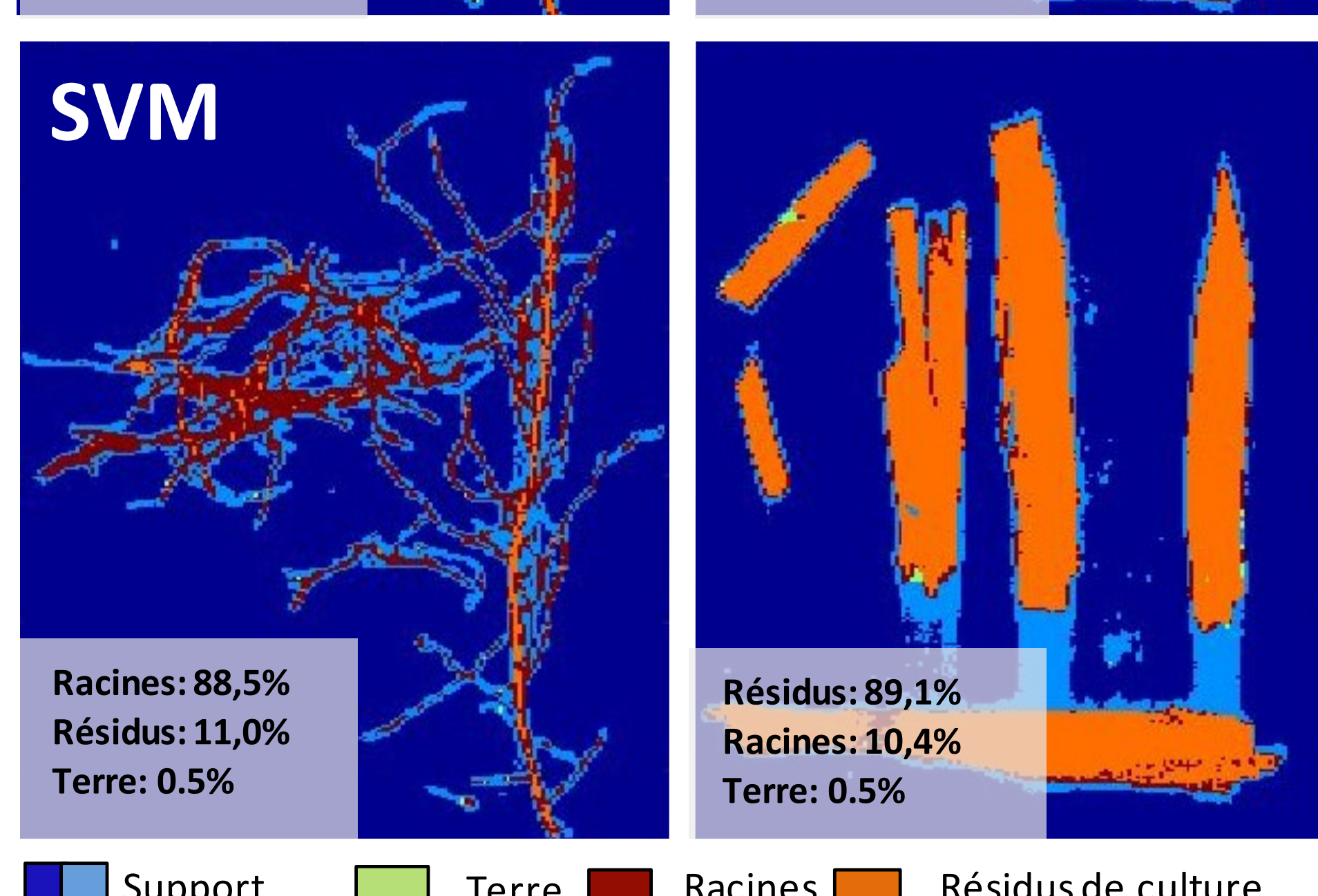
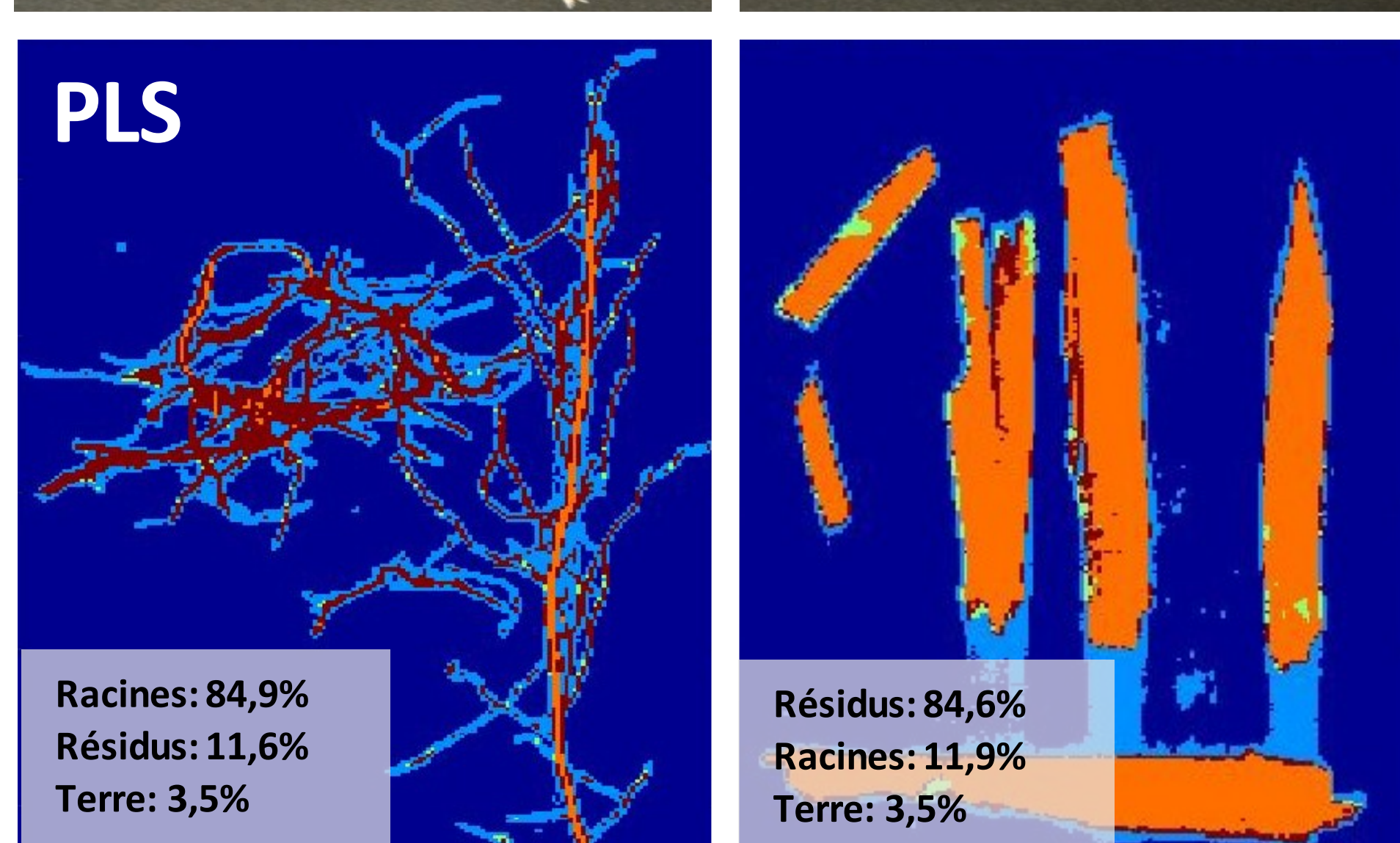
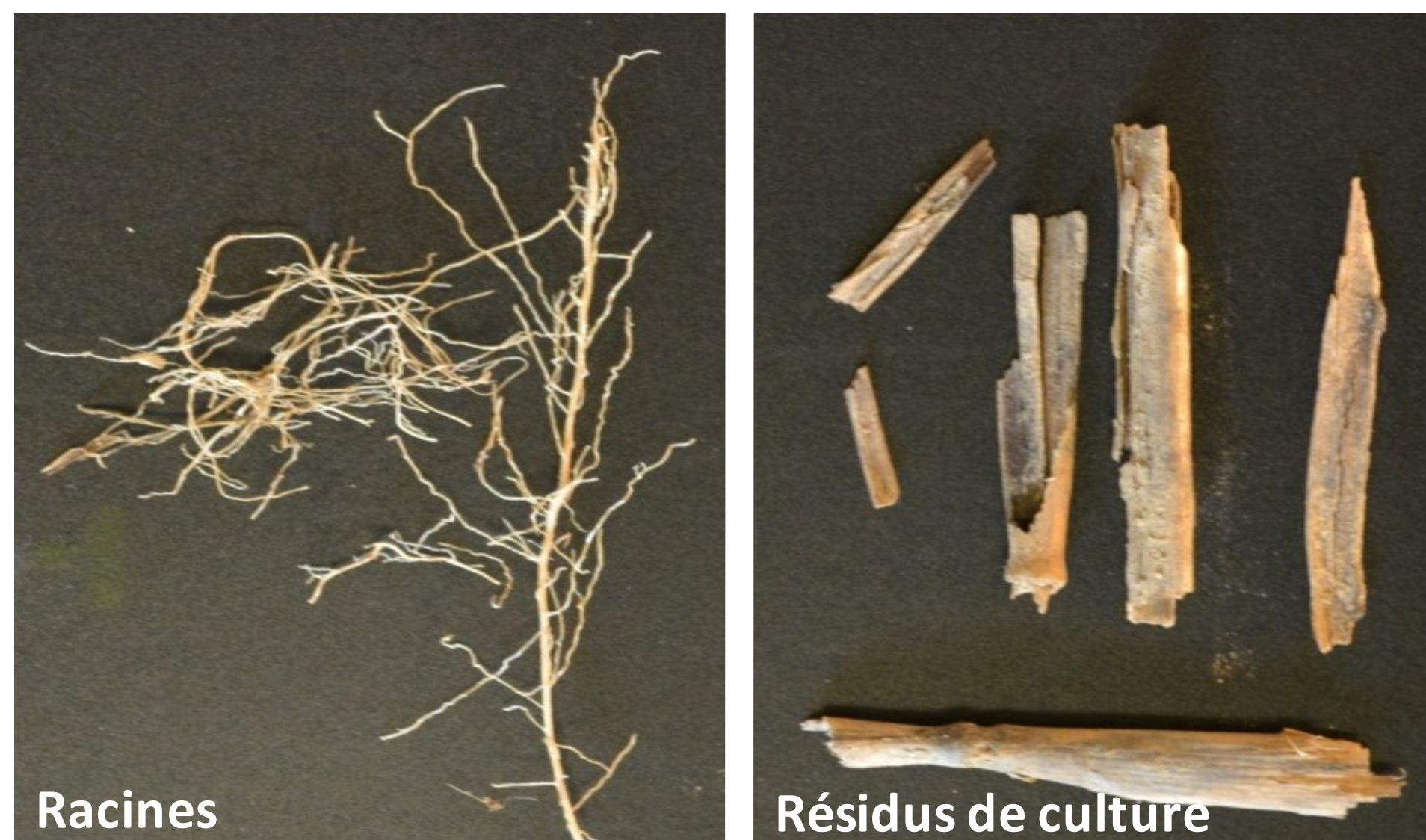


Prédiction et quantification

Une couleur est attribuée à chaque classe spectrale afin de pouvoir reconstruire l'image et visualiser le résultat des prédictions. La quantification est ensuite basée sur le nombre de pixels prédits pour chaque classe.



Résultats et conclusions



Validation sur spectres proche infrarouge:

Quelle que soit la classe spectral étudiée, les modèles SVM donnent toujours des résultats meilleurs ou égaux que les modèles PLS. Ceci s'observe aussi bien lors de la calibration des modèles que lors de leur validation avec un set de spectres indépendant. La moins bonne discrimination spectrale apparaît avec les spectres de racines et de résidus de culture qui ont, parmi les 4 classes spectrales à discriminer, les compositions chimiques les plus similaires. La différence la plus élevée entre les 2 types de modèles d'analyse discriminante a été atteinte dans ce cas particulier lors de la calibration: 9.1% des spectres de résidus de culture furent mieux prédits avec les modèles SVM.

	Modèle	Calibration		Validation	
		Sensibilité	Spécificité	Sensibilité	Spécificité
Support vs. Reste	PLS	96,1	97,5	92,1	94,2
	SVM	97,9	99,6	99,6	98,0
Terre vs. Reste	PLS	99,3	98,2	98,6	96,2
	SVM	99,2	99,7	99,0	99,4
Résidus de culture vs. Racines	PLS	88,0	86,5	85,0	82,2
	SVM	97,1	92,6	93,4	84,9

Sensibilité: proportion de spectres classés comme étant positifs dans la classe positive du modèle

Spécificité: proportion de spectres classés comme étant négatifs dans la classe négative du modèle

Validation sur images prises en proche infrarouge:

Pour estimer la capacité des modèles à prédire la véritable nature des pixels sur les images hyperspectrales, ceux-ci ont été appliqués sur des images prises sur des éléments appartenant aux différentes classes spectrales (terre, résidus de culture, racines) posés séparément sur le tapis de mesure. La proportion de pixels correctement prédits fut calculée en divisant le nombre de pixels prédits comme appartenant à une classe spectrale par le nombre total de pixels comptabilisés pour les 3 classes spectrales. Les résultats ont montré de meilleures prédictions avec les modèles SVM: un plus faible nombre de pixels est prédit comme étant du sol et les pixels en bordure des résidus de culture sont mieux prédits (cfr figure).

Conclusions:

Les modèles SVM ont donc démontré leurs plus grandes performances dans cette étude et ils ont permis de mettre au point une méthode de discrimination rapide et fiable des racines et des résidus de culture utilisable pour l'étude des systèmes racinaires.

Références:

- [1] Fernández Pierna J. A., Baeten V., Dubois J., Burger J., Lewis E. N., Dardenne P., 2009. NIR Imaging – Theory and applications, in: *Comprehensive Chemometrics*, Oxford, Elsevier (S. Brown, R. Tauler and B. Walczak eds.), Vol 4, pp. 173-196.
- [2] Eylenbosch D., Fernandez Pierna J. A., Baeten V., Bodson B., 2014. Detection of wheat root and straw in soil by use of NIR hyperspectral imaging spectroscopy and Partial Least Square discriminant analysis, in: *Proceedings of the ESA XIIIth Congress*, Debrecen, Hungary, pp. 237-238.
- [3] Vermeulen P., Fernández Pierna J. A., van Egmond H. P., Dardenne P., Baeten V., 2012. Online detection and quantification of ergot bodies in cereals using near infrared hyperspectral imaging. *Food Addit Contam A*, 29(2), pp. 232-240.

Logiciels: PLS Toolbox 7.8 (Eigenvector Research, Inc., Wenatchee, WA, USA) fonctionnant sur Matlab (The MathWorks, Inc., Natick, MA, USA).