

## Caractérisation de la concentration en carbone organique du sol à l'échelle de la France par spectrométrie proche et moyen infrarouge

CLAIROTTE Michaël<sup>1</sup>, GRINAND Clovis<sup>2</sup>, KOUAKOUA Ernest<sup>2</sup>, THÉBAULT Aurélie<sup>1,3</sup>, JOLIVET Claudy<sup>4</sup>, ARROUAYS Dominique<sup>4</sup>, BERNOUX Martial<sup>2</sup>, BARTHÈS, Bernard<sup>2</sup>

<sup>1</sup> : UMR Eco&Sols, INRA, 2 place Viala, 34060 Montpellier, France ; <michael.clairotte@supagro.inra.fr>

<sup>2</sup> : UMR Eco&Sols, IRD, 2 place Viala, 34060 Montpellier, France ; <c.grinand@etcterra.org>, <ernest.kouakoua@ird.fr>, <martial.bernoux@ird.fr>, <bernard.barthes@ird.fr>

<sup>3</sup> adresse actuelle : ITK, Cap Alpha, avenue de l'Europe, 34830 Clapiers, France ; <thebault.aurelie@gmail.com>

<sup>4</sup> : Infosol, INRA, BP 20619, 45166 Olivet, France ; <claudy.jolivet@orleans.inra.fr>, <dominique.arrouays@orleans.inra.fr>

### Résumé

La spectrométrie infrarouge quantitative permet la caractérisation rapide et peu coûteuse de nombreuses propriétés des sols, en particulier leur concentration en carbone organique ( $C_{org}$ ). Cette technique est basée sur des étalonnages, qui expriment les propriétés d'un échantillon (par exemple  $C_{org}$ ) en fonction de son spectre. Ces étalonnages sont construits à partir de la caractérisation conventionnelle et spectrale d'échantillons étalons, et permettent de prédire les propriétés considérées sur de nouveaux échantillons d'après leur seul spectre. Jusqu'à présent, la construction d'étalonnages de propriétés des sols valides sur de vastes territoires reste rare. Ce travail présente l'intérêt de la spectrométrie proche et moyen infrarouge (PIR et MIR, respectivement) pour caractériser  $C_{org}$  à l'échelle de la France.

Une grosse librairie spectrale a été constituée en utilisant une collection d'échantillons fournie par le Réseau de mesures de la qualité des sols (RMQS, INRA-Orléans). Cette collection nationale comprend environ 3800 échantillons prélevés à 0-30 et 30-50 cm de profondeur sur la base d'une grille carrée de 16 km de côté sur tout le territoire métropolitain (552 000 km<sup>2</sup>). Les spectres de réflectance ont été acquis au laboratoire sur sol séché à l'air et broyé à 0.2 mm avec deux appareils : (i) un spectrophotomètre proche infrarouge NIRSystems 5000 (Foss, Silver Spring, MD, USA) couvrant le domaine 1100-2500 nm à un pas de 2 nm ; et (ii) un spectrophotomètre moyen infrarouge à transformée de Fourier Nicolet 6700 (Thermo Fisher Scientific, Madison, WI, USA) couvrant le domaine 4000-400 cm<sup>-1</sup> à un pas de 3.86 cm<sup>-1</sup>. Selon l'analyse conventionnelle de  $C_{org}$ , réalisée au laboratoire d'analyse des sols de l'INRA-Arras par combustion sèche après éventuelle décarbonatation (ISO 10694:1995),  $C_{org}$  sur la population d'échantillons considérée varie de 0.16 à 177 g kg<sup>-1</sup>, avec une moyenne de 19.1 g kg<sup>-1</sup> et une médiane de 13.6 g kg<sup>-1</sup>. Toutes les analyses spectrales ont été réalisées avec le logiciel R et les packages *prospectr*, *pls* et *resemble*.

La population totale d'échantillons a été divisée en trois groupes : (i) une sous-population d'étalonnage sur laquelle ont été construits les modèles de prédiction ; cette sous-population incluait environ 3000 échantillons, sélectionnés avec l'algorithme de Kennard-Stone pour être les plus représentatifs spectralement de la population totale ; (ii) une sous-population de validation sur laquelle ont été testés les modèles ; cette sous-population comptait 380 échantillons sélectionnés avec le même algorithme pour être les mieux représentés par le reste de la population totale ; et (iii) une sous-population de réglage, regroupant les 380 échantillons restant, sur laquelle ont été déterminés certains paramètres des modèles, lorsque nécessaire. Le nombre d'échantillons d'étalonnage effectivement utilisés pour construire

les modèles a été varié de 10% à 100% de la sous-population d'étalonnage, afin d'identifier l'intensité d'étalonnage optimale (permettant d'obtenir un modèle précis avec un minimum d'étalons). Les étalonnages ont été réalisés par régression des moindres carrés partiels (PLS en anglais, pour partial least squares), soit globale, soit locale. Dans l'étalonnage global, un seul modèle est construit pour tous les échantillons à prédire (il s'agit du type de régression multivariée usuel actuellement en spectrométrie infrarouge quantitative). Dans l'étalonnage local, chaque échantillon est prédit à partir d'un modèle propre utilisant les étalons qui lui sont voisins spectralement, dont le mode de sélection et le nombre ont été définis sur la sous-population de réglage. La qualité des prédictions a été estimée sur la sous-population de validation à l'aide des critères suivants : erreur standard de prédiction corrigée du biais (SEPC, qu'on souhaite minimale) ; coefficient de détermination entre prédictions et mesures ( $R^2_{\text{val}}$ , qu'on souhaite le plus proche de 1) ; et coefficient RPD (ratio entre écart type de validation et SEPC, qu'on souhaite maximal).

Avec étalonnage global, le MIR donne des prédictions plus précises que le PIR sur les échantillons considérés : SEPC = 2.6 vs. 4.4 g kg<sup>-1</sup>,  $R^2_{\text{val}}$  = 0.88 vs. 0.87, et RPD = 2.7 vs. 2.3, respectivement ; on considère en général que les prédictions des propriétés des sols sont précises quand RPD  $\geq$  2, ce qui est ici le cas dans les deux domaines spectraux (Fig. 1). Les prédictions sont plus précises encore avec étalonnage local, et restent meilleures avec MIR qu'avec PIR : SEPC = 2.0 vs. 3.5 g kg<sup>-1</sup>,  $R^2_{\text{val}}$  = 0.92 vs. 0.89, RPD = 3.5 vs. 2.9. Ces résultats ont été obtenus en utilisant 100% des étalons ; mais l'intensité optimale d'étalonnage a été estimée à 30% de la sous-population d'étalonnage (utiliser plus d'étalons améliore faiblement les prédictions) ; dans ces conditions, avec étalonnage local, les critères de performance du modèle MIR sont les suivants : SEPC = 2.6 g kg<sup>-1</sup>,  $R^2_{\text{val}}$  = 0.88, RPD = 2.6. Sur une telle campagne d'échantillonnage, on peut donc n'analyser conventionnellement que 30% des échantillons pour prédire correctement les autres. Mais  $C_{\text{org}}$  ayant été déterminé sur tous les échantillons du RMQS, il est possible actuellement d'utiliser un modèle construit avec l'ensemble des 3800 échantillons pour prédire  $C_{\text{org}}$  sur n'importe quel sol de France avec une erreur standard de l'ordre de 2 g kg<sup>-1</sup> (si l'échantillon est bien représenté dans la population totale) ; cela rend la spectrométrie infrarouge de laboratoire compétitive avec les approches conventionnelles, surtout la spectrométrie MIR (au moins pour les sols des régions tempérées). Certaines voies sont susceptibles de permettre d'améliorer encore la précision des prédictions, par exemple le recours à certaines méthodes de régression multivariée non linéaires, ou l'augmentation de la population d'étalonnage (grille de prélèvement plus serrée).

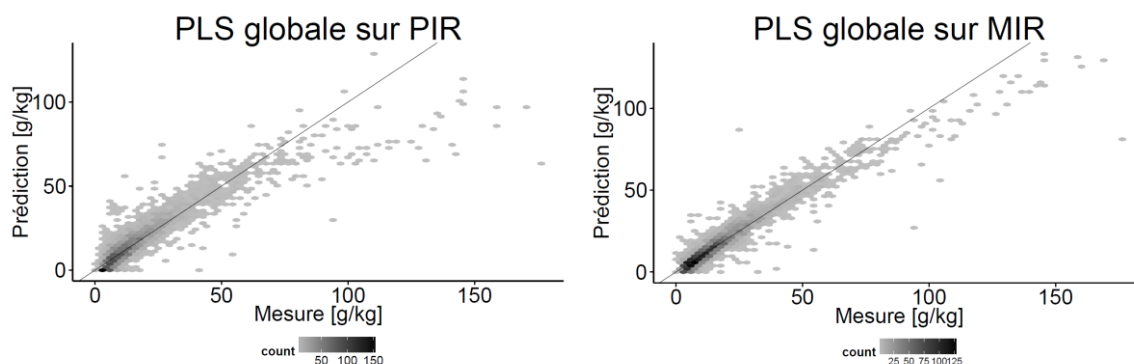


Fig. 1. Comparaisons sur la sous-population de validation entre mesures (combustion sèche) et prédictions de  $C_{\text{org}}$  par PLS globale en PIR (à gauche) et MIR (à droite) ; la droite représente la première bissectrice (prédiction = mesure).