

## Exercice - Session 4

L'évaluation de la teneur en Carbone Organique (CO) d'un sol est un enjeu important pour quantifier les niveaux de pollution. La méthode la plus fiable pour mesurer cette teneur en Carbone Organique repose sur une analyse physico-chimique coûteuse, qui n'est pas envisageable pour une évaluation rapide du niveau de pollution d'une surface importante. D'autres méthodes sont donc envisagées dont la spectrométrie proche infra-rouge.

Afin d'évaluer le niveau de précision de cette méthode, on procède à une expérience sur une parcelle située dans le sud tunisien, sur un chott. Sur cette zone, on définit une grille régulière de 144 points de mesure sur lesquels on prélève un échantillon de sol pour une analyse déterminant sa teneur en CO (g/kg). Chaque échantillon de sol est également soumis à spectrométrie proche infra-rouge (on dispose d'une mesure spectrale par nombre d'onde entre 400 et 2500 nm).

Les données de spectrométrie et de concentration en carbone organique sont disponibles dans les fichiers `nirstozeur.txt` et `carbontozeur.txt` respectivement.

La problématique est donc la suivante : comment prédire la concentration en carbone organique à partir de l'analyse par spectrométrie proche infra-rouge ?

Avant toute utilisation de ces données de spectrométrie proche infra-rouge, les spécialistes recommandent de leur appliquer une transformation (appelé SNV pour Standard Normal Variate) consistant à centrer et réduire chaque spectre (en lui soustrayant sa propre moyenne et en le divisant ensuite par son propre écart-type).

1. Représentez graphiquement les spectres proche-infra rouges après transformation par SNV.
2. Selon vous, quel(s) intervalle(s) de nombre d'onde apporte(nt) l'information la plus intéressante pour prédire la concentration en carbone organique ?
3. Proposez un modèle statistique permettant d'étudier la relation entre la concentration en carbone organique et l'analyse par spectrométrie proche infra-rouge d'un échantillon de sol.

Une des méthodes d'estimation possibles du modèle de la question 3, très utilisée par les chimiométriciens et appelée régression PLS, s'appuie sur l'extraction d'un nombre limité de variables latentes à partir du spectre.

4. En utilisant cette méthode, proposez une équation de prédiction de la concentration en carbone organique à partir de l'analyse par spectrométrie proche infra-rouge.
5. Donnez une interprétation des deux premières composantes PLS.
6. Évaluez la précision de cette méthode de prédiction.