

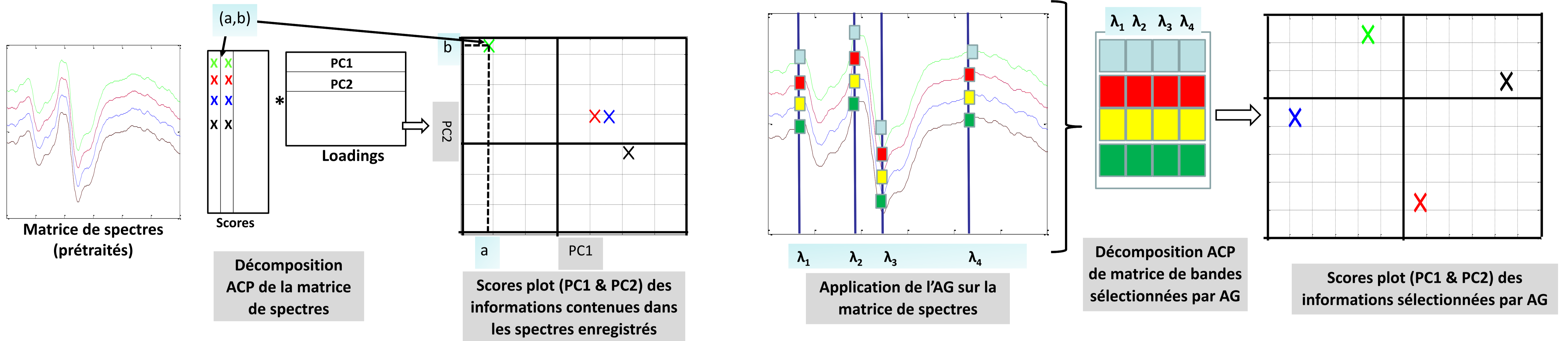
1. Introduction

Un des défis actuels dans l'analyse et la modélisation des processus de biodégradation de la biomasse végétale par spectroscopies MIR/NIR est de déterminer les modes de vibrations (donc les nombres d'ondes), intrinsèquement liées à la composition chimique et qui sont les plus discriminants par rapport à la cinétique de biodégradation.

Objectifs:

- implémenter un Algorithme Génétique (AG) qui permet la sélection automatique de nombres d'ondes les plus discriminants. Tester différentes fonctions fitness qui visent à avoir les clusters les plus compacts possibles et les plus séparés.
- Cette étude aborde également la possibilité de combiner les informations spectrales MIR et NIR avec l'objectif d'améliorer la discrimination des échantillons lors du processus de biodégradation.

2. Méthodologie



Décomposition ACP de la matrice des spectres et représentation de scores plot

- Les scores ACP permettent d'analyser la séparabilité des échantillons selon la cinétique de biodégradation.
- L'AG sélectionne des bandes spectrales optimales permettant une meilleure séparabilité : Les meilleurs résultats obtenus sont avec la fonction fitness Davies-Bouldin (DB) qui maximise toutes les distances inter-cluster et minimise la distance intra-cluster pour chaque cluster.

Sélection de bandes spectrales par AG, décomposition ACP des informations sélectionnées et représentation de scores plot.

- Pour quantifier la séparabilité par rapport à la cinétique de biodégradation, l'Indice de Dunn (DI) est calculé avec les scores obtenus :

$$DI = \frac{d_{min}}{d_{max}}$$

où d_{min} désigne la distance minimale entre deux classes différentes et d_{max} la distance maximale entre deux spectres d'une même classe.

3. Application

► Biomasse lignocellulosique :

Racines de maïs issues de deux lignées parentales distinctes (F2 et F292) et deux mutants de ces lignées (F2bm1 et F292bm3), analysées sur 5 périodes de biodégradation: $t_1=0$, $t_2=14$, $t_3=36$, $t_4=57$ et $t_5=112$ jours (G. Machinet and al. 2009).

► Combinaison des informations spectrales MIR-NIR :

Nous mettons bout à bout le segment [800 1800] cm^{-1} de MIR et le segment [4000 6000] cm^{-1} de NIR. Chaque spectre est prétraité séparément.

► Spectres :

- Les spectres ont été enregistrés par l'IRTF Nicolet 6700 Thermo en mode DRIFT dans les gammes 800-1800 cm^{-1} en MIR et 4000-6000 cm^{-1} en NIR.
- Les spectres ont été prétraités par filtrage *Savitzky-Golay* (SG) de 1^{er} ordre avec un lissage sur 17 points et un polynôme d'ordre 4, suivi d'une normalisation de type *Standard Normal Variate*.

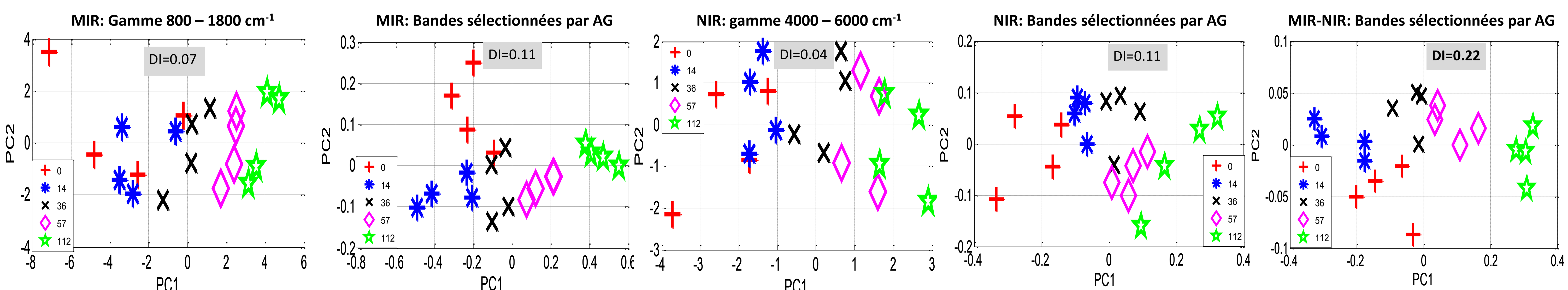
► Bandes sélectionnées par AG :

MIR	NIR	MIR-NIR : bout à bout
858; 953; 1383 et 1707 cm^{-1}	4850; 5195; 5540; 5705 cm^{-1}	858; 1385; 5541 et 5705 cm^{-1}

- 858 cm^{-1} : Vibrations du squelette aromatiques combiné avec CH déformation (balancement),
- 953 cm^{-1} : C-O-C élongation de la polysaccharides,
- 1383 cm^{-1} : Cellulose avec lignine (élongation -CH de groupement CH₃),
- 1707 cm^{-1} : Hemicellulose (élongation C=O de groupements cétones non conjugués, carbonyles et esters)
- 4850 cm^{-1} : élongation C = O de groupement CH₃, et OH + déformation O-H,
- 5195 cm^{-1} : Eau (O-H antisymétrique, vibration de valence + vibration de déformation O-H dans H₂O),
- 5540 cm^{-1} : Le groupe fonctionnel CH (première harmonique de CH₃ et -CH = CH-),
- 5705 cm^{-1} : première harmonique d élongation CH.

Les nombres d'ondes sélectionnés par l'AG correspondent aux vibrations principales des groupes fonctionnels chimiques. Ils peuvent ainsi être attribués à l'évolution chimique des échantillons étudiés au cours de la biodégradation.

► Scores plots :



L'application de la ACP sur les informations spectrales MIR et NIR sélectionnées aux nombres d'ondes identifiés par l'AG donne une meilleure discrimination des échantillons par rapport à la cinétique de biodégradation que l'application classique de l'ACP sur les informations enregistrées à tous les nombres d'ondes en MIR et NIR.

La prise en compte des informations conjointes en MIR et NIR donne la meilleure discrimination des échantillons par rapport à la cinétique de biodégradation.

4. Conclusion et perspectives

- L'algorithme génétique avec la fonction fitness Davies-Bouldin permet l'identification de nombres d'ondes dans les deux gammes spectrales MIR et NIR correspondant aux groupes fonctionnels chimiques qui peuvent ainsi être attribués à l'évolution chimique des échantillons étudiés au cours de la biodégradation. Nous avons comparé différentes fonctions fitness qui visent à avoir les clusters les plus compacts possibles et les plus séparés, la fonction Davies-Bouldin offrant le meilleur résultat.
- Les scores plot ACP montrent une meilleure discrimination selon les périodes de biodégradation que l'application classique sur toutes les gammes spectrales MIR et NIR. Ces résultats sont à vérifier sur d'autres biomasses lignocellulosiques.
- En combinant des informations des deux gammes spectrales MIR et NIR, la discrimination par rapport à la cinétique de biodégradation est améliorée. Cette approche doit être confortée sur d'autres biomasses lignocellulosiques, voire développée.