



Les 21ème Rencontres HelioSPIR
29/06/2021 - VISIO

Analyse de données spectrales massives avec parSketch-PLSDA pour la description de géotypes de tournesols

Maxime RYCKEWAERT, Maxime METZ, Daphné HERAN, Ryad BENDOULA



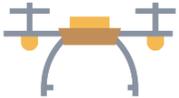
➤ Introduction

La quantité de données en agriculture

Les vecteurs



Satellites



Drone



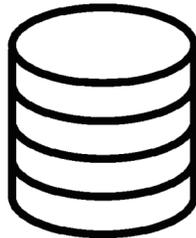
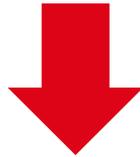
Moissonneuse



Tracteur



Brouette



Les mesures



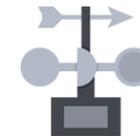
Lidar



Imagerie



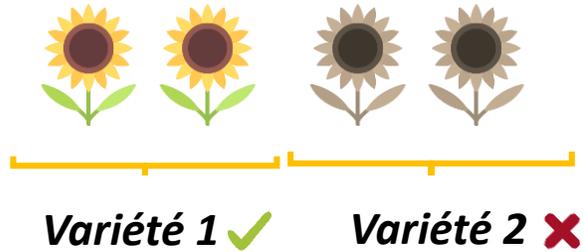
IoT



Weather station

➤ Introduction

La quantité de données en sélection variétale (*Phénotypage*)



Besoin d'instruments précis



Micro-parcelles/grandes surfaces



Différencier des variétés selon des **critères spécifiques**



Rendement

Adaptation aux conditions pédoclimatique



Tolérances aux stress

Biotique



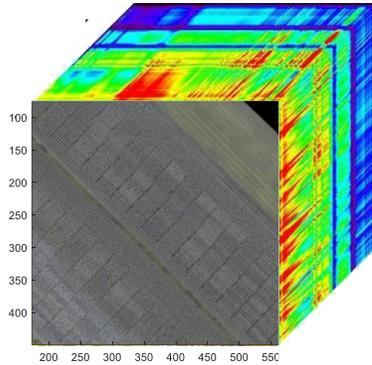
Abiotique

Meilleures qualités nutritionnelles

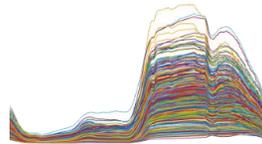
Besoin d'une information riche
-> spectrale par exemple

➤ Introduction

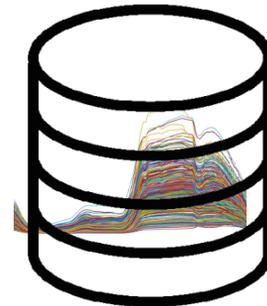
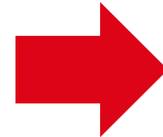
La quantité de données en sélection variétale (Phénotypage)



L'imagerie hyperspectrale
Résolutions spectrale et spatiale élevées



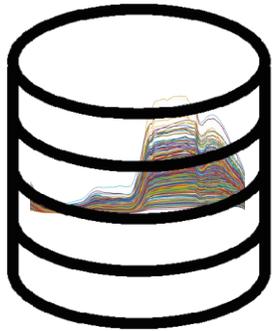
Lorsqu'elle est utilisée en routine de phénotypage



➤ Introduction

La quantité de données... pose problème

De simples opérations deviennent problématiques



- Détection de valeurs aberrantes
- Prétraitements
- Méthodes linéaires challengées

- Time consuming
- Ressources informatiques
- Structure complexe
- Non-linéarité

[Szymańska, 2018]

[Dardenne et al., 2000]

[Bertran et al., 1999]

Une nouvelle méthode parSketch-PLSDA

[Metz et al., 2020]

Méthode locale pour un grand volume de données



INRAE

parSketch-PLSDA pour la discrimination de variétés

29/06/2021 / HelioSPIR 2021

➤ La méthode parSketch-PLSDA

parSketch-PLS-DA [Metz et al., 2020]

parSketch

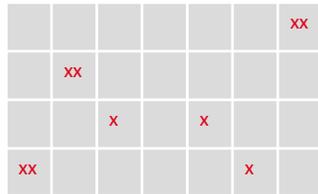
1. Réduction de dimensions
Lemme de Johnson Lindenstrauss (1986)

$$\mathbf{T} = \mathbf{XP}$$

[-1, 1, -1 ...]

INDEX

2. Création d'une grille



3. Renvoi des voisins

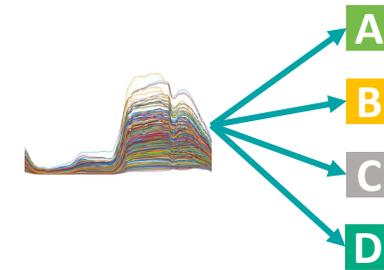
Nombre de voisins à définir

Paramétrisation: 3 paramètres à trouver

PLS-DA

Partial Least Square –
Discriminant Analysis (PLS-DA)

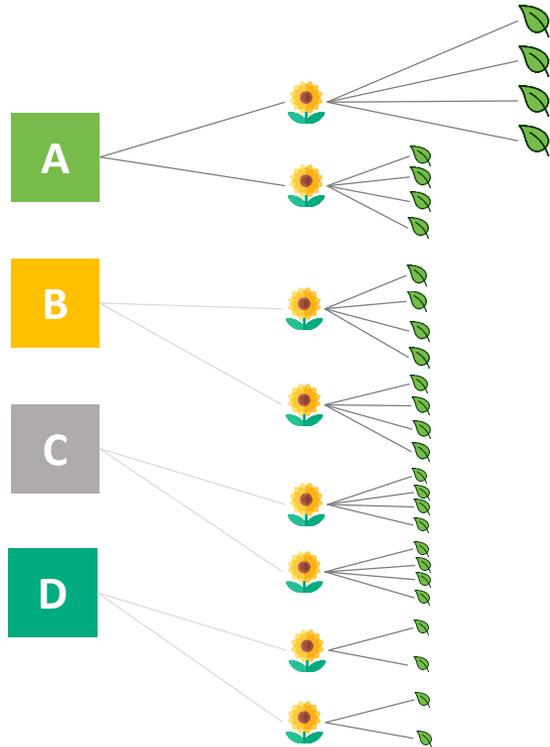
[Barker et Rayens, 2003]



➤ Cas d'étude:

Discrimination de 4 variétés de tournesols

Plan d'expérience



Acquisition d'images hyperspectrales



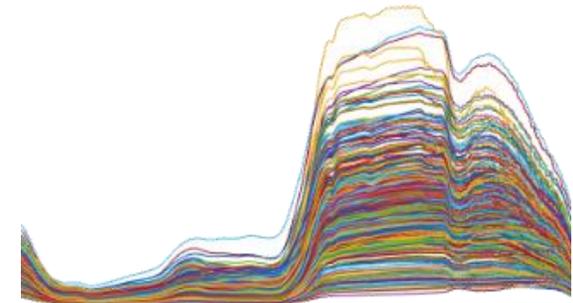
Camera: NEO Hypspec, VNIR-1600
Spectral Range: [400:1000 nm]



Total: 28 images

➤ Cas d'étude:

Constitution de la base de données



28 images



1 300 000 spectres

14 images pour la calibration : 650 000 spectres

14 images pour le test (réduit): 14 000 spectres sélectionnés aléatoirement (dans les 14 images)

> Cas d'étude:

Méthodes utilisées

parSketch-PLS-DA

[Metz et al., 2020]

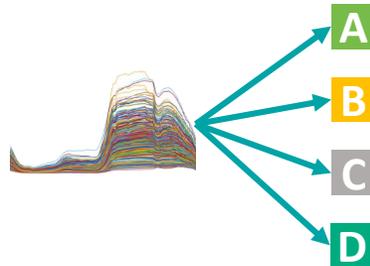
INDEX



PLS-DA

1. Réduction de dimensions
2. Création d'une grille
3. Renvoi des voisins

paramétrisation

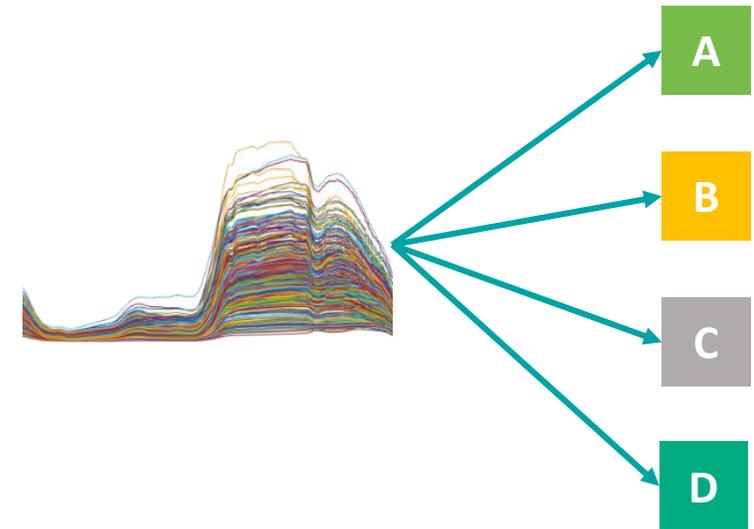


Sélection

PLS-DA

[Barker et Rayens, 2003]

Utilisée comme **méthode de référence**



Cross-validation (kfold: 5 blocks x 3)



INRAE

➤ Cas d'étude:

Critères d'évaluation

A partir d'une matrice de confusion

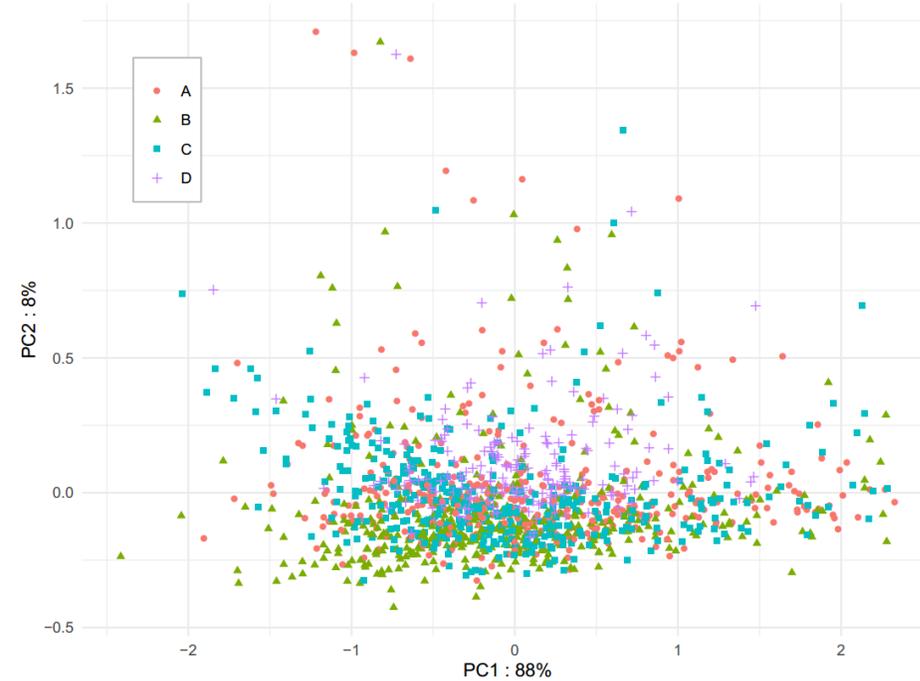
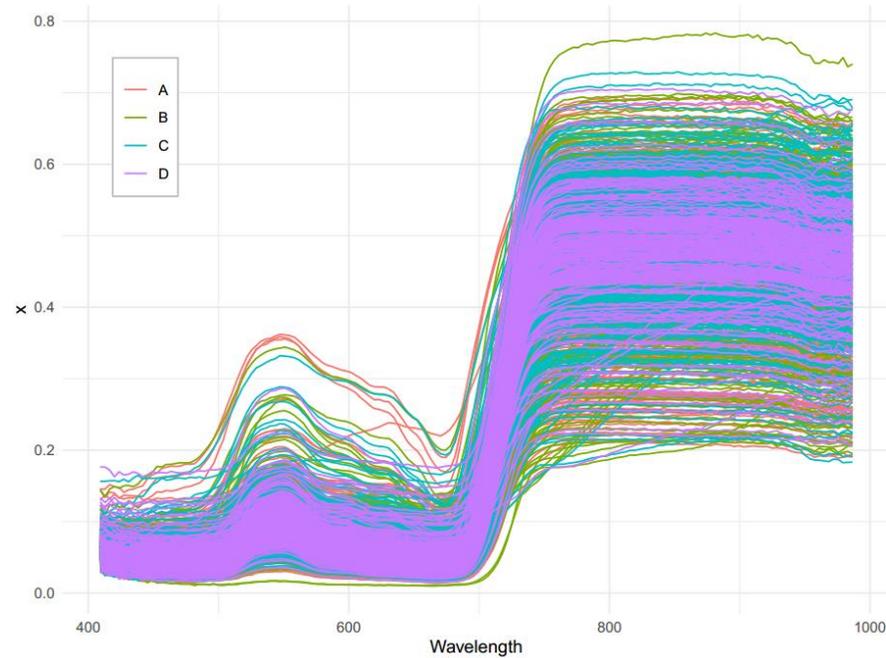
	A	B	C	D	Recall (%)
A	<i>tp</i>	<i>fn</i>	<i>fn</i>	<i>fn</i>	
B	<i>fp</i>				
C	<i>fp</i>				
D	<i>fp</i>				
Precision(%)					

$$\text{Precision} = \frac{tp}{tp + fp}$$

$$\text{Recall} = \frac{tp}{tp + fn}$$

➤ Résultats

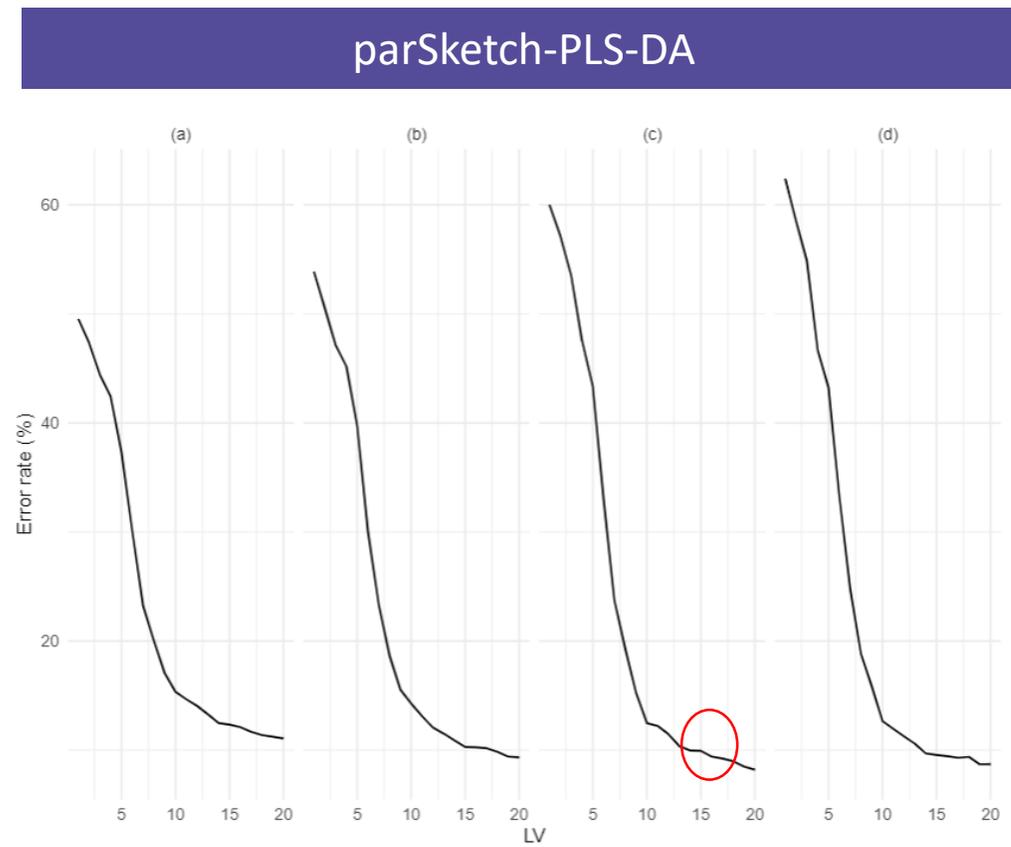
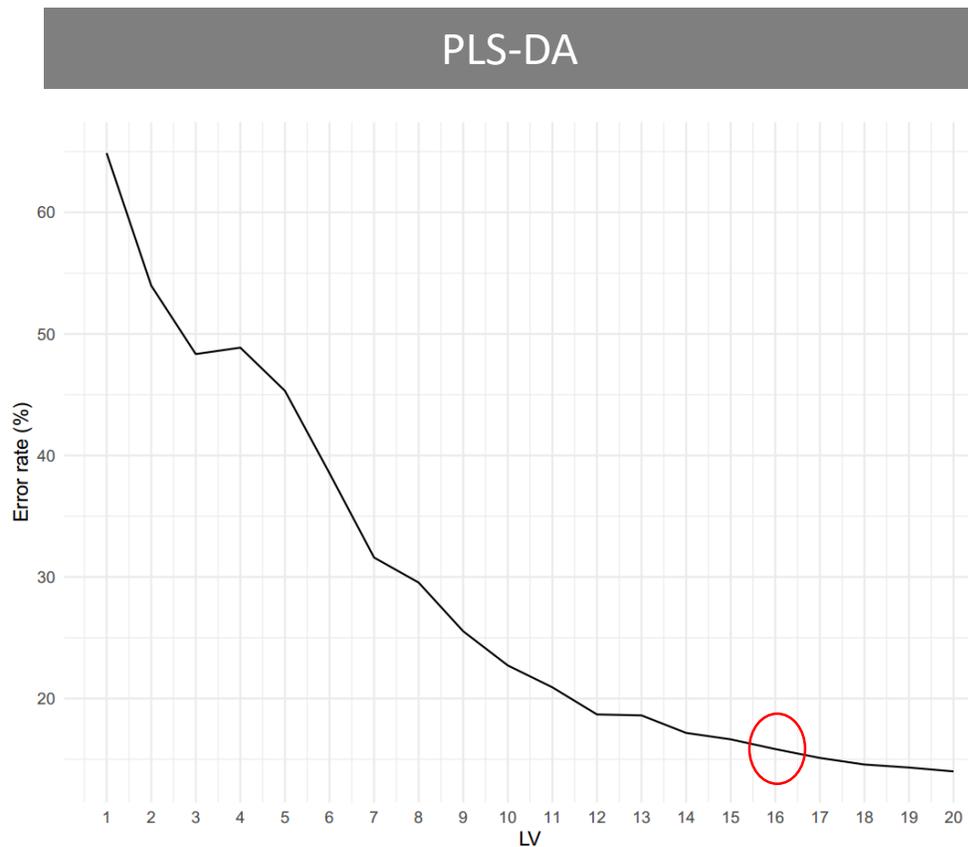
Visualisation des spectres



Pas de distinctions évidentes des génotypes

➤ Résultats

Choix du nombre de composantes



➤ Résultats

PLS-DA

	A	B	C	D	Recall (%)
A	3120	121	339	250	81
B	238	2812	619	154	74
C	127	241	3463	75	89
D	283	173	270	1213	63
Precision(%)	83	84	74	72	

parSketch-PLS-DA

	A	B	C	D	Recall (%)
A	3547	114	115	54	93
B	40	3256	473	54	85
C	58	211	3590	47	92
D	51	112	260	1516	78
Precision(%)	96	88	81	91	

➤ Conclusion et perspectives

Conclusion

parSketch-PLS-DA réduit de 10% l'erreur de classification par rapport à la PLS-DA
Présence de non-linéarité dans le jeu de données
Adaptée pour traiter une grande quantité de données à la différence de la PLS-DA

Perspectives

A tester sur d'autres jeux de données contenant des non-linéarités
Pour d'autres applications en sélection variétale : Maladies/Stress ...
-> Anticiper le futur goulot d'étranglement de la quantité de données en Agriculture

Massive spectral data analysis for plant breeding using parSketch-PLSDA method:
discrimination of sunflower genotypes (Submitted)

*Maxime Ryczewaert, Maxime Metz, Daphné Héran, Pierre George, Bruno Grèzes-Besset, Reza Akbarinia,
Jean-Michel Roger, Ryad Bendoula*



INRAE

➤ Thank You

Is there any questions?
Maybe at the « **coffee** » break

➤ Les méthodes

parSketch-PLSDA

