




La spectroscopie proche-infrarouge en agronomie

Martin Ecartot, INRAE / AGAP Institut





Plan du TD

- Le phénotypage et la SPIR
- Données SPIR: Comment est généré le spectre
- Applications en agronomie
- Prétraitements et Analyse en Composantes Principales
- Au boulot ! Application avec 
 - Prise en main des données
 - Pré-traitements
 - ACP
- Discussion

Le Phénotypage et la SPIR

Besoin de phénotypage pour la **sélection variétale** et la qualité des produits

- L'œil humain

- Qualité des épis/grains
- Caractérisation des feuilles (senescence, maladies)
- Architecture des plantes (Hauteur, port foliaire,...)
- Evaluation du dispositif



- L'analyse chimique

- Azote (feuille, bottillon)
- Protéines des grains
- Lipides

- Le génotypage

- Sélection génomique



Le Phénotypage et la SPIR

Le phénotypage optique ou « Haut-débit »

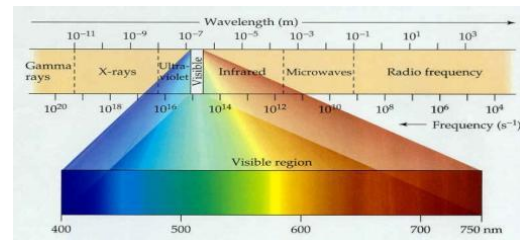
Avantages

- Mesures objectives et Répétables
- Bon marché
- Pas de produit chimique
- Mesures non destructive
- Rapide
- Facilement embarquable
- Voir dans l'Infra-rouge



Inconvénients

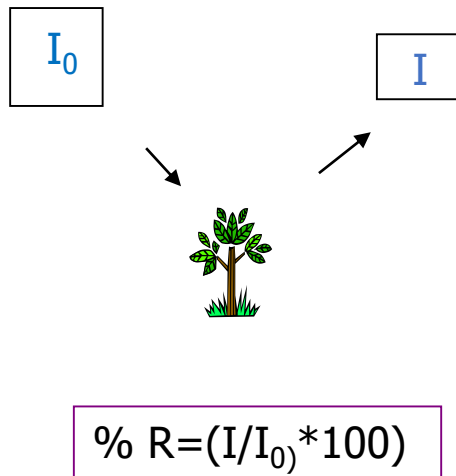
- Technicité pour mesurer
- Nécessite un étalonnage



Le Phénotypage et la SPIR

Pourquoi le Proche Infrarouge?

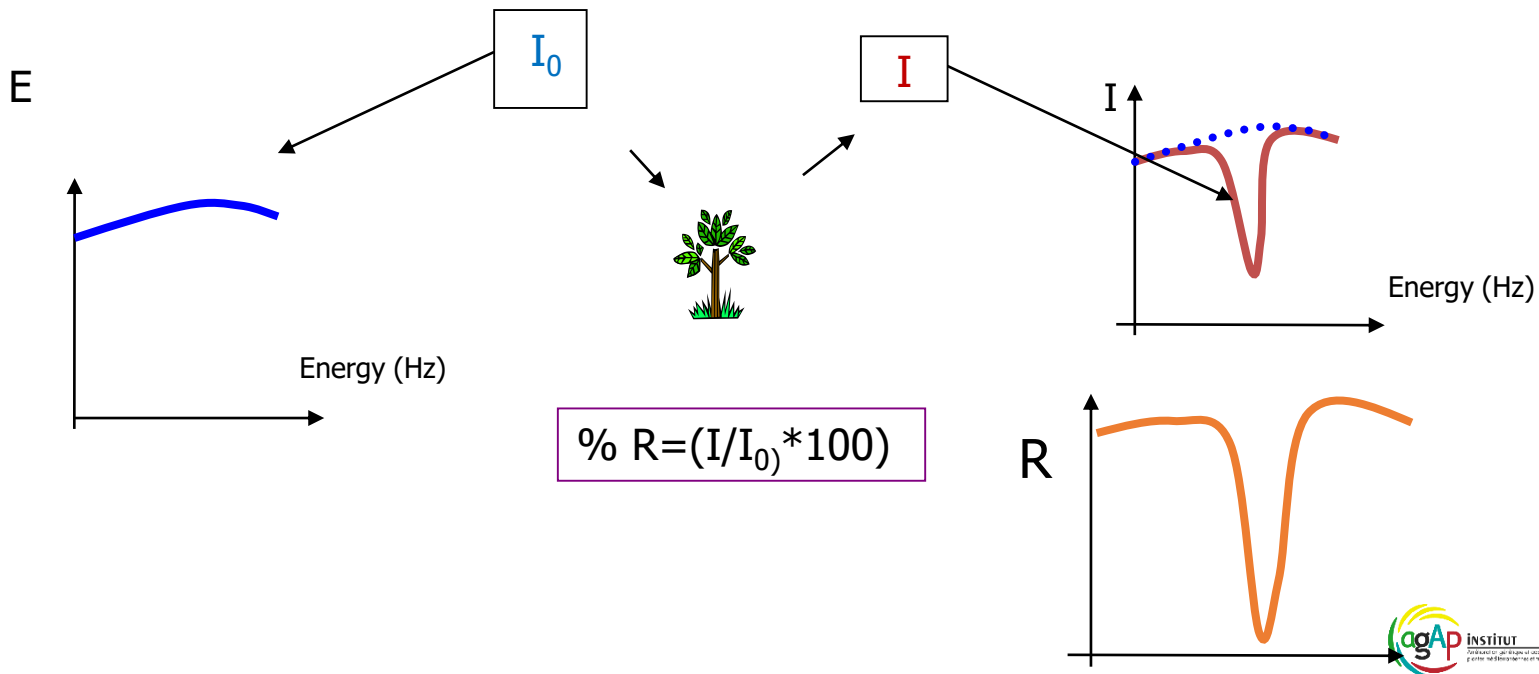
Mesure Optique



Le Phénotypage et la SPIR

Pourquoi le Proche Infrarouge?

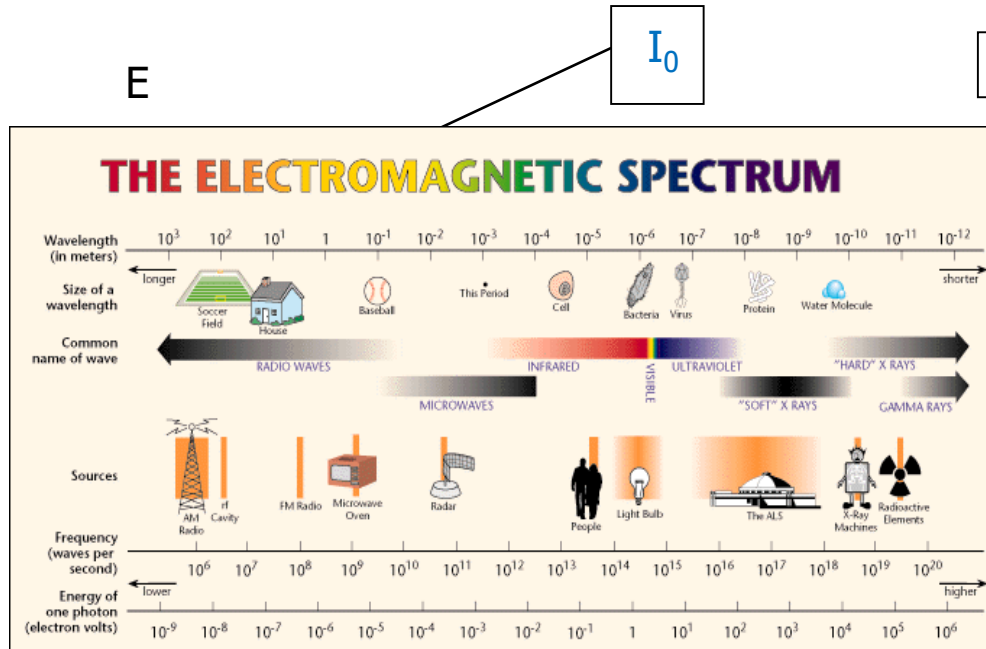
Mesure par spectroscopie



Le Phénotypage et la SPIR

Pourquoi le Proche Infrarouge?

Mesure par spectroscopie



E

I_0

I

I

Energy (Hz)

R

Le Phénotypage et la SPIR

Pourquoi le Proche Infrarouge?

Identifier et Quantifier un composé:

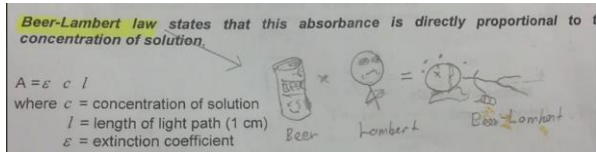
Loi de Beer-Lambert

$$A = \epsilon \cdot l \cdot c = -\log(R)$$

ϵ : coef. d'absorption

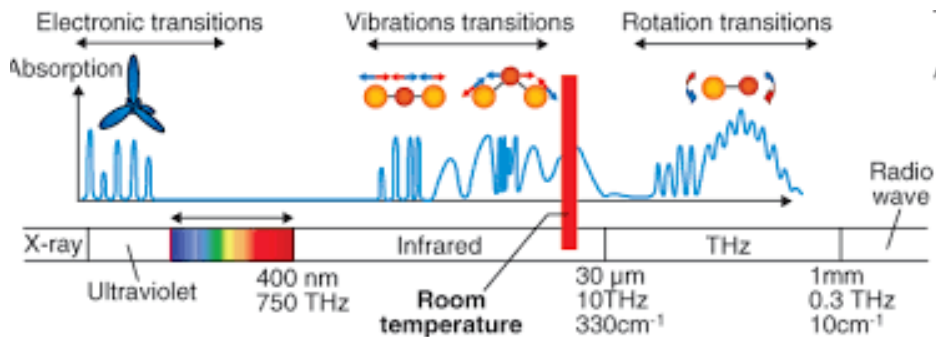
l : longueur du trajet optique

c : concentration du composé



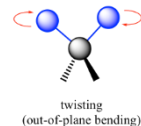
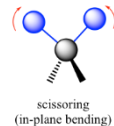
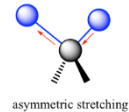
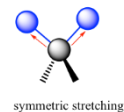
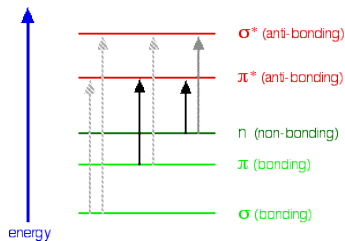
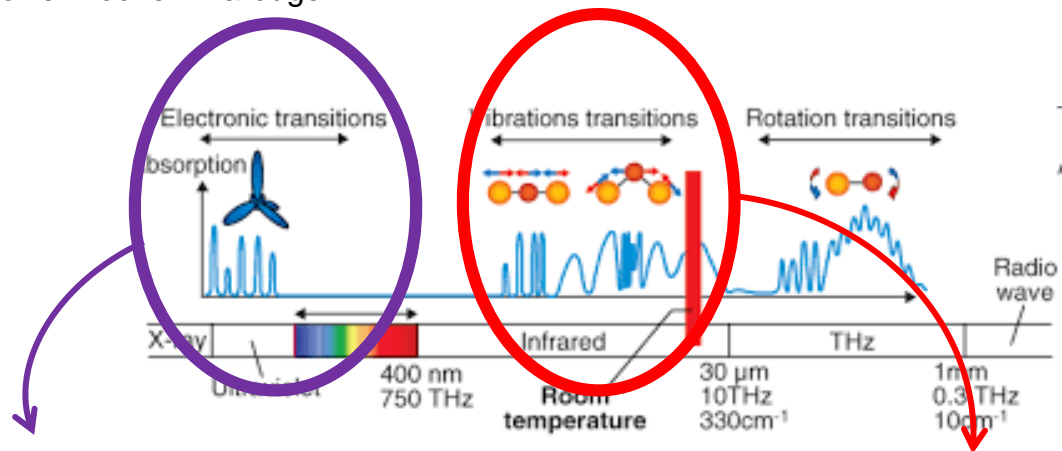
Le Phénotypage et la SPIR

Pourquoi le Proche Infrarouge?



Le Phénotypage et la SPIR

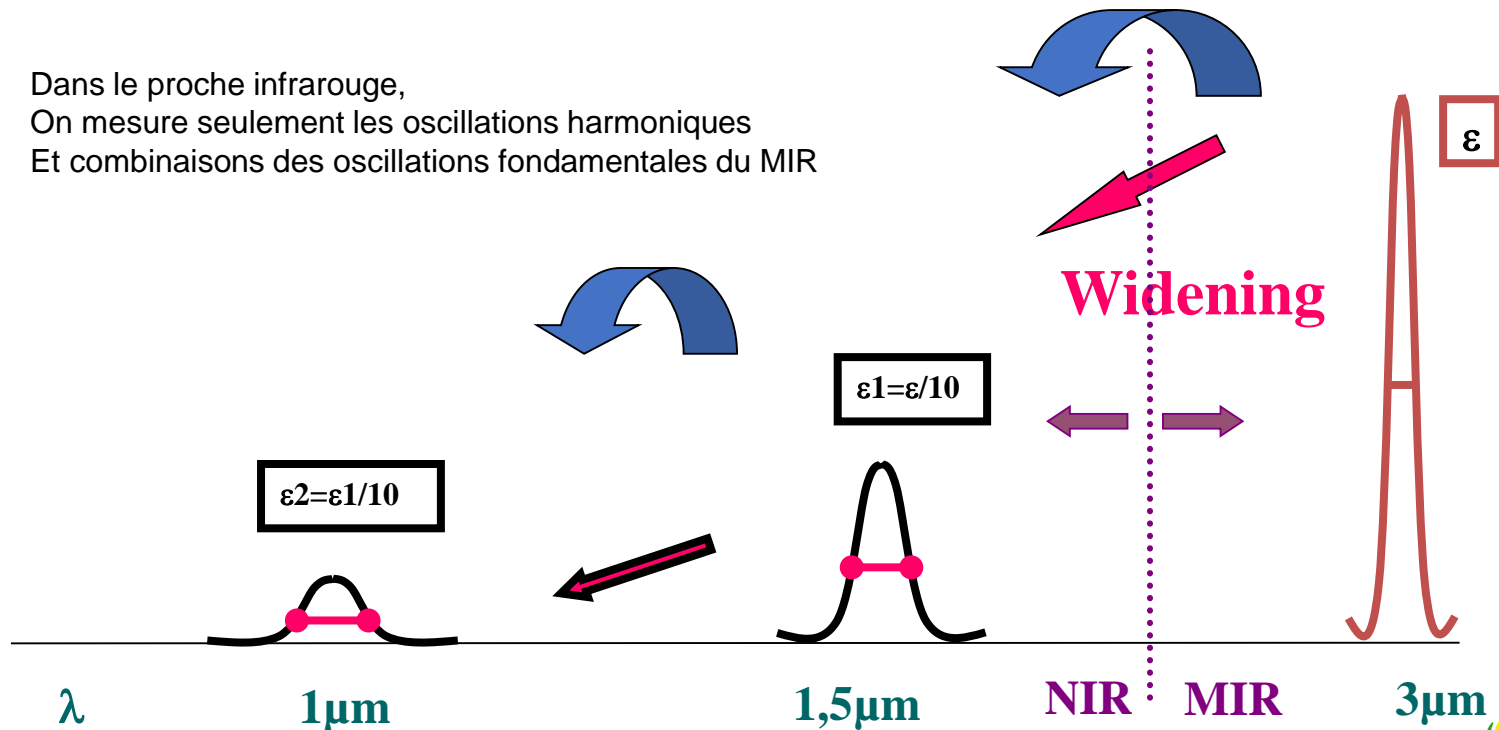
Pourquoi le Proche Infrarouge?



Le Phénotypage et la SPIR

Pourquoi le Proche Infrarouge?

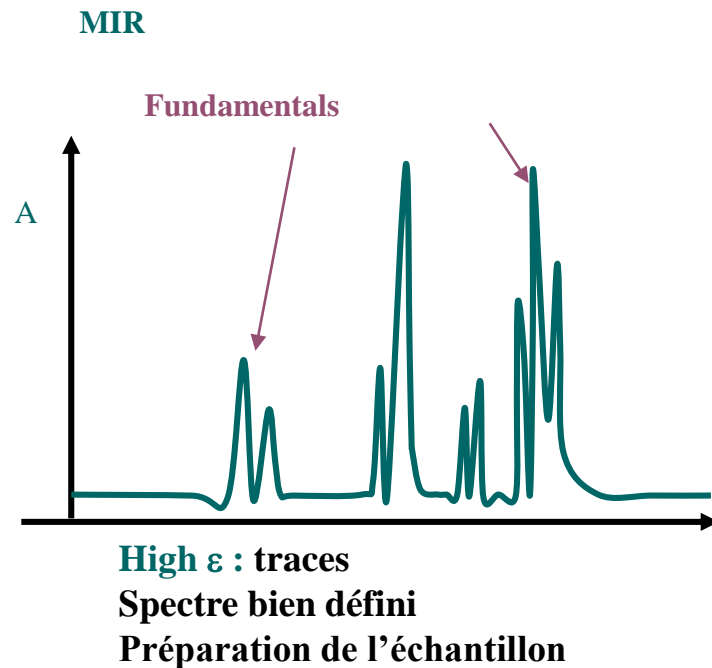
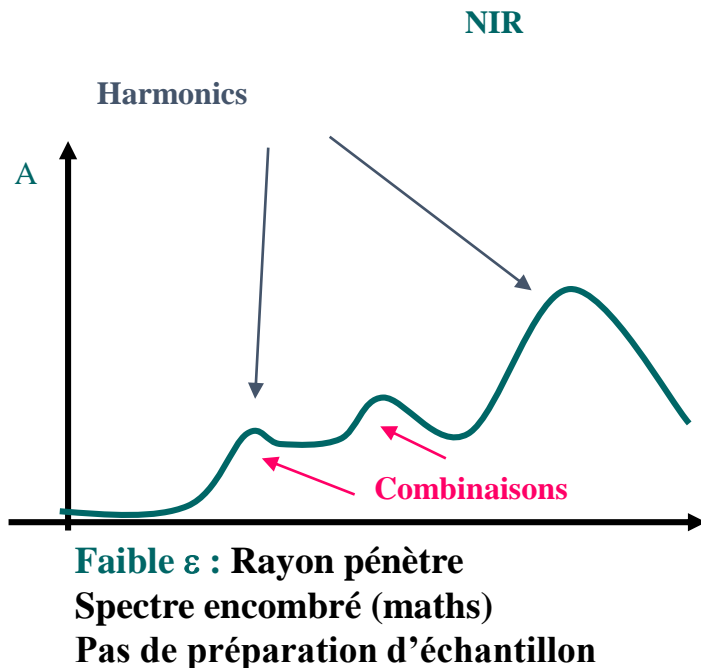
Dans le proche infrarouge,
On mesure seulement les oscillations harmoniques
Et combinaisons des oscillations fondamentales du MIR





Le Phénotypage et la SPIR

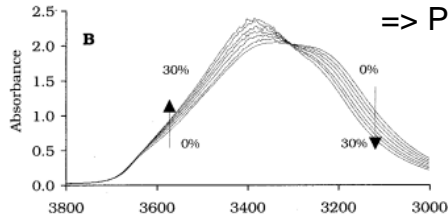
Pourquoi le Proche Infrarouge?



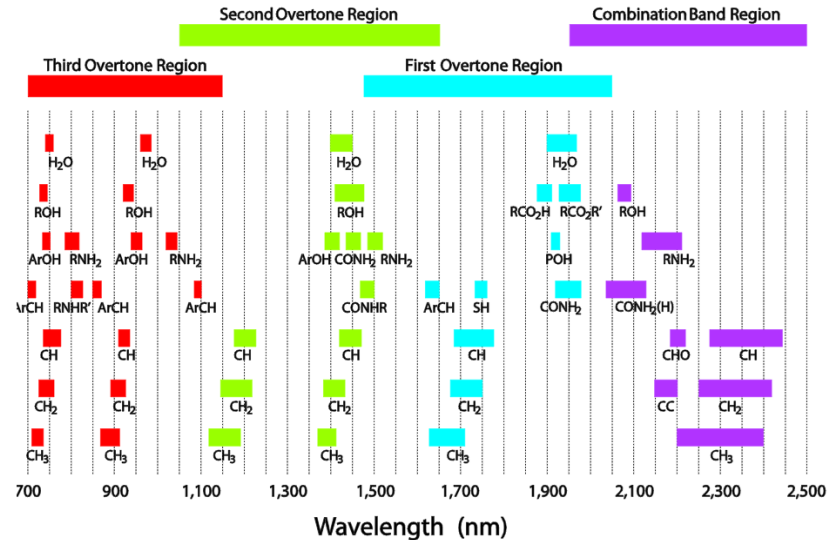
Le Phénotypage et la SPIR

Pourquoi le Proche Infrarouge?

- Avantages
 - Pénétration de la lumière + profonde que MIR (>1mm)
 - Nombreux composés sensibles dans cette gamme
- Inconvénients
 - Bandes larges -> Recouvrement
 - Faible sensibilité (difficile si < 0.1%)
 - Nombreuses perturbations
 - Température
 - Granulométrie
 - Interactions



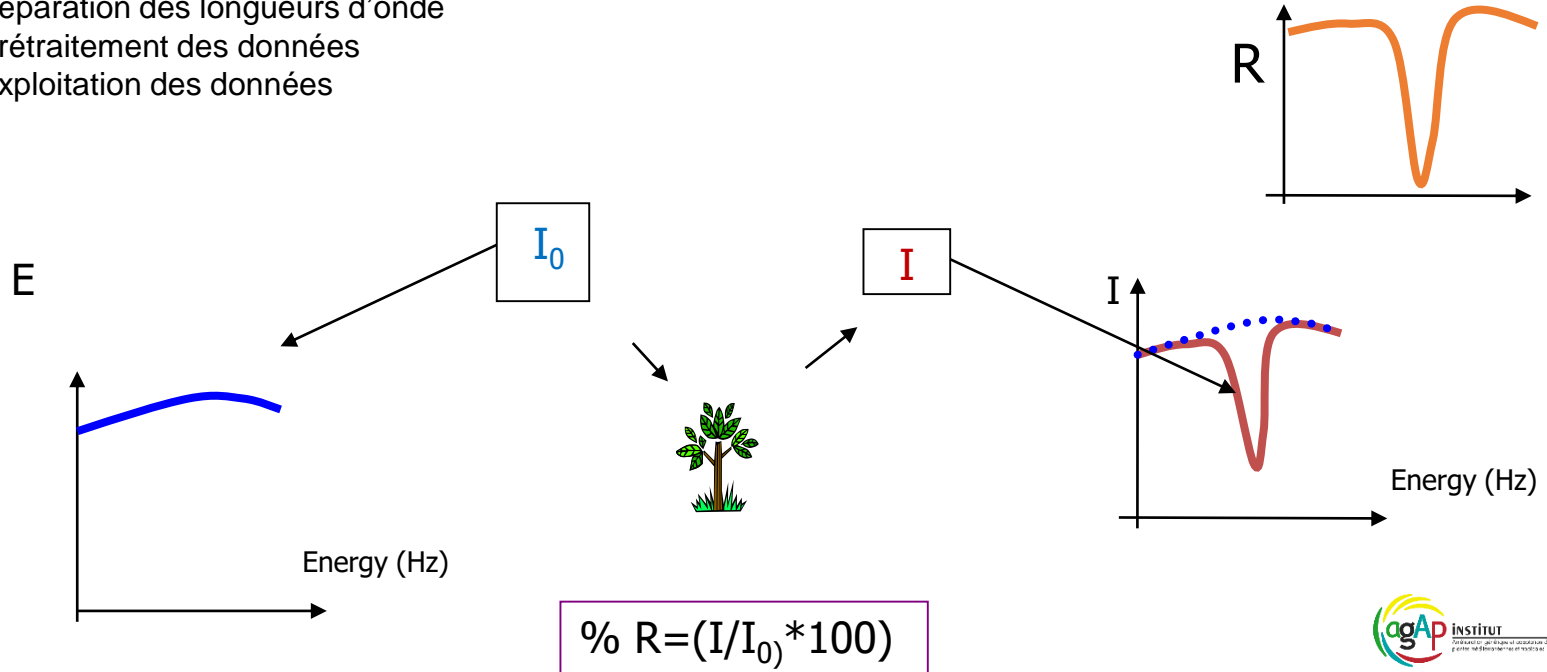
=> Pas de modèle simple
suivant Beer-Lambert



La SPIR: génération et traitement des données

Les données SPIR, de A à Z

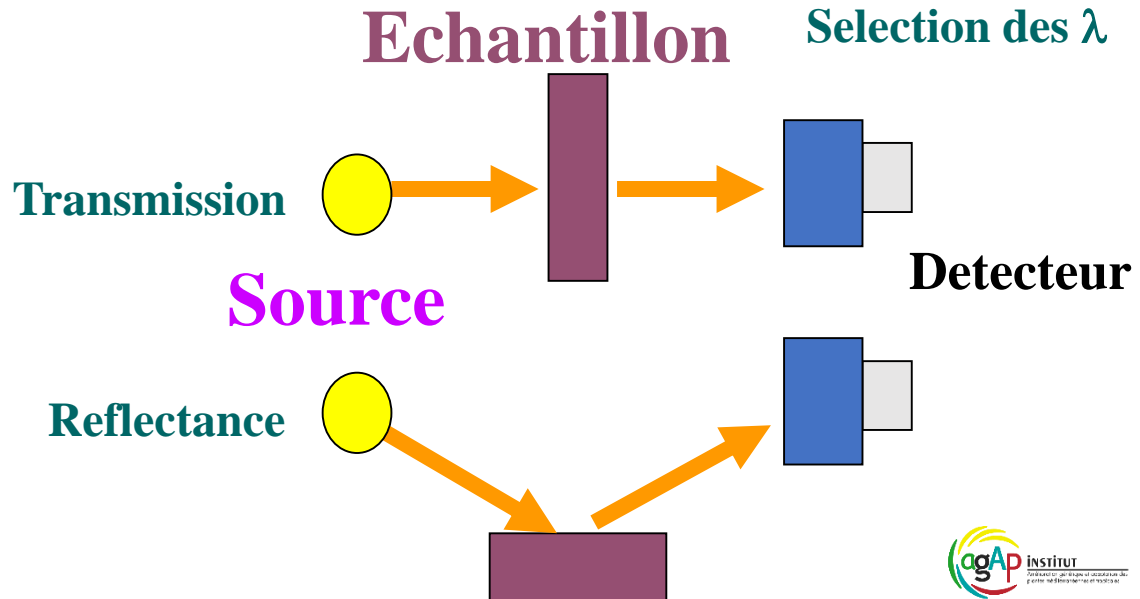
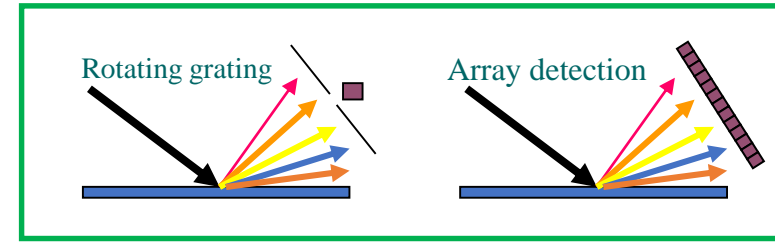
- Présentation de l'échantillon
- Séparation des longueurs d'onde
- Prétraitement des données
- Exploitation des données



La SPIR: génération et traitement des données

Les données SPIR, de A à Z

- Présentation de l'échantillon
 - Reflexion/Transmission
- Séparation des longueurs d'onde
 - Filtre
 - Réseau de diffraction
 - Transformée de Fourier
- Prétraitement des données
- Exploitation des données



La SPIR: génération et traitement des données

Les données SPIR, de A à Z

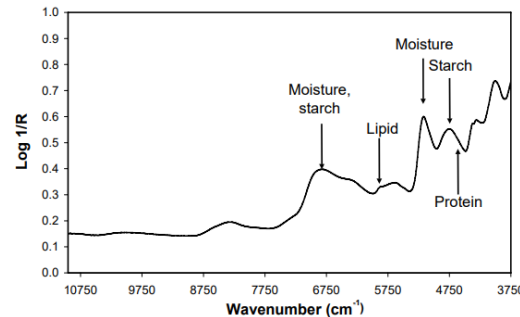
- Prétraitement des données

Qu'est-ce qu'un spectre?

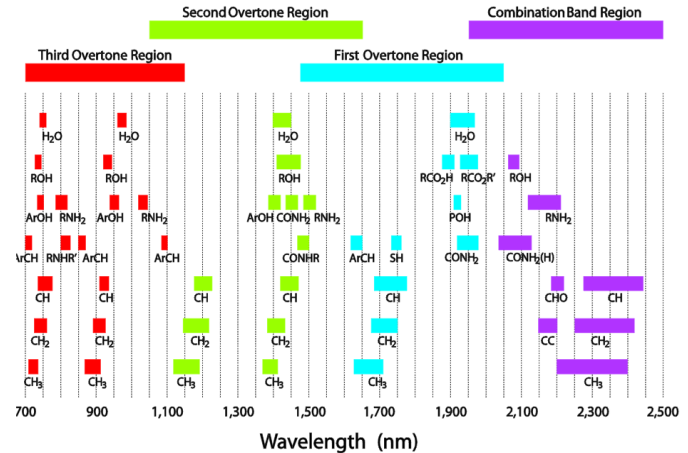
- Interaction lumière/matière
- Influence de la composition chimique
- Influence de la physique
- Perturbations: électronique, contamination,...



Grains de blé (reflectance)



Farine de blé, Bhandari et al, 2006



La SPIR: génération et traitement des données

Les données SPIR, de A à Z

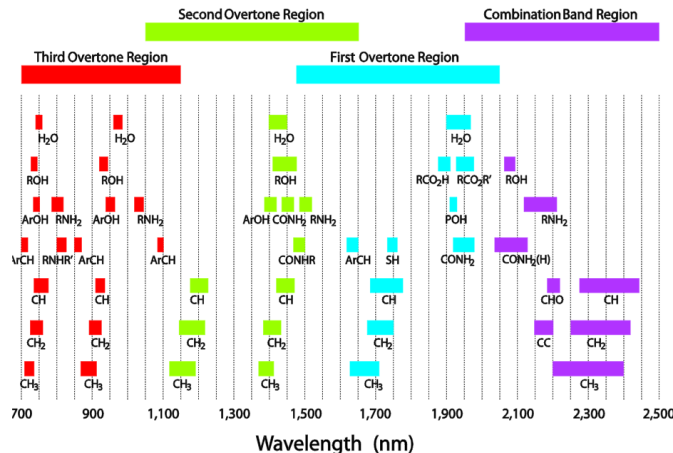
- Prétraitement des données

Qu'est-ce qu'un spectre?

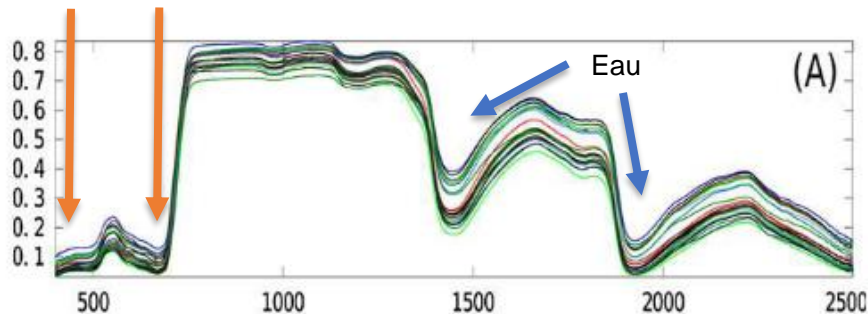
- Interaction lumière/matière
- Influence de la composition chimique
- Influence de la physique
- Perturbations: électronique, contamination,...



Spectres de feuilles



Chlorophylles



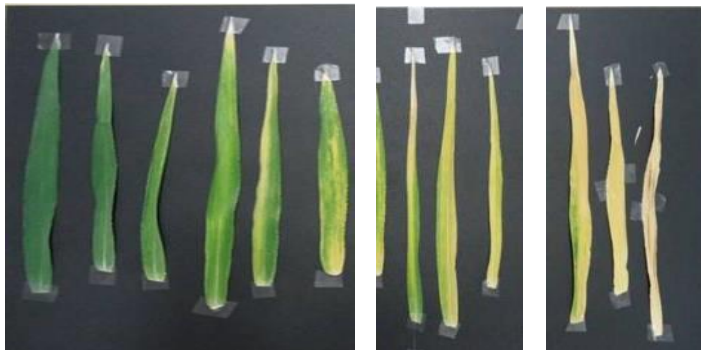
La SPIR: génération et traitement des données

Les données SPIR, de A à Z

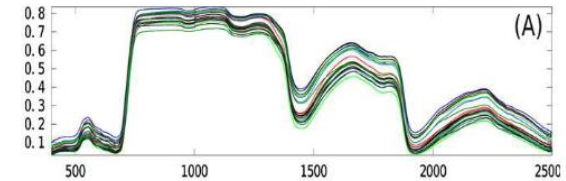
- Prétraitement des données

Qu'est-ce qu'un spectre?

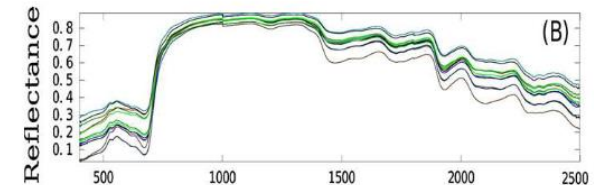
- Interaction lumière/matière
- Influence de la composition chimique
- Influence de la physique
- Perturbations: électronique, contamination,...



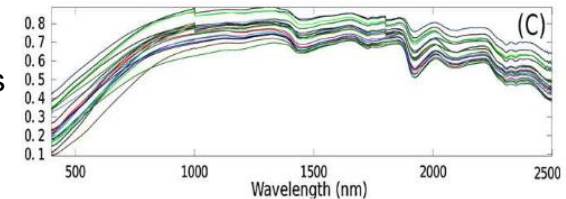
Feuilles fraîches



Feuilles sèches



Feuilles sénescentes



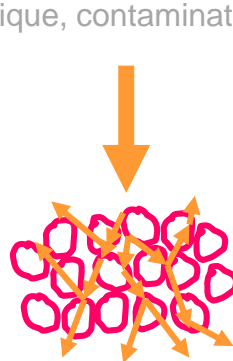
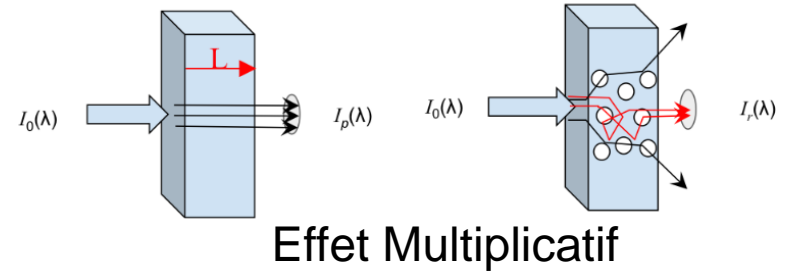
La SPIR: génération et traitement des données

Les données SPIR, de A à Z

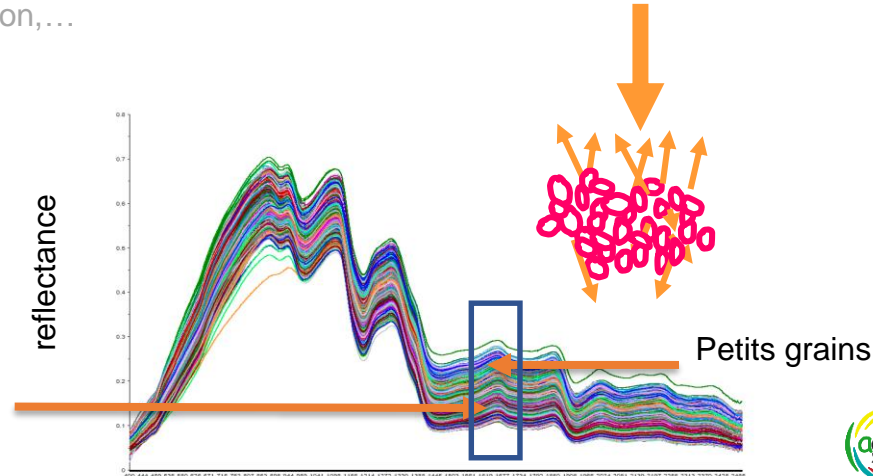
- Prétraitement des données

Qu'est-ce qu'un spectre?

- Interaction lumière/matière
- Influence de la composition chimique
- Influence de la physique
- Perturbations: électronique, contamination,...



Gros Grains



Petits grains

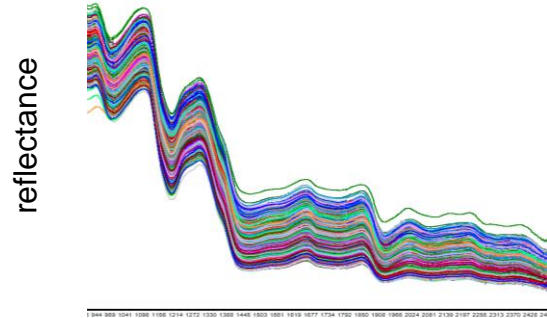
La SPIR: génération et traitement des données

Les données SPIR, de A à Z

- Prétraitement des données

Qu'est-ce qu'un spectre?

- Interaction lumière/matière
- Influence de la composition chimique
- Influence de la physique
- Perturbations: électronique, contamination,...



Dérive de ligne de base:
La diffusion de la lumière croît avec la
longueur d'onde

La SPIR: génération et traitement des données

Les données SPIR, de A à Z

- Prétraitement des données

Qu'est-ce qu'un spectre?

- Interaction lumière/matière
- Influence de la composition chimique
- Influence de la physique
- Perturbations: électronique, contamination,...

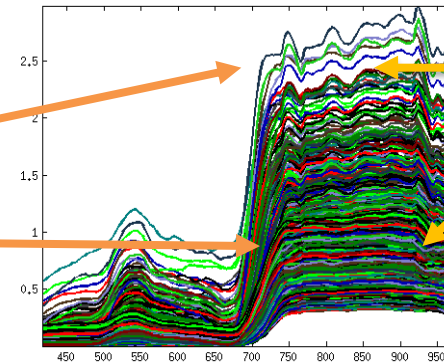
Réflexion spéculaire



(a) comportement spéculaire



(b) comportement diffus



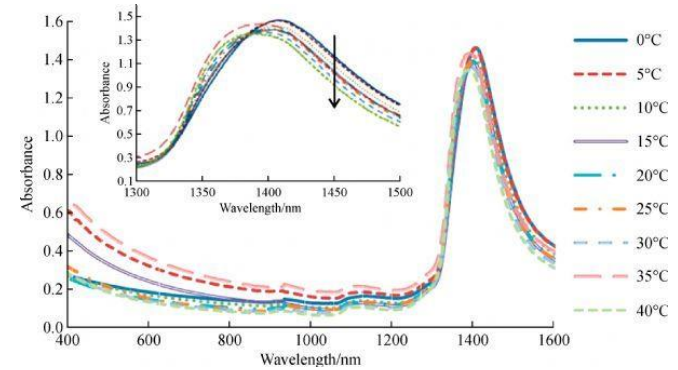
La SPIR: génération et traitement des données

Les données SPIR, de A à Z

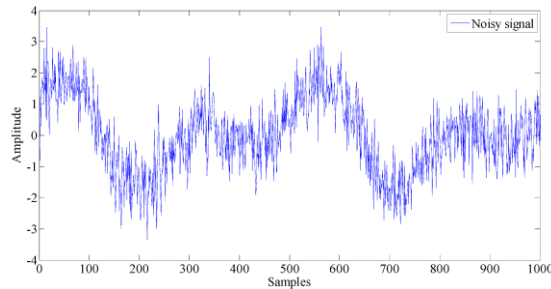
- Prétraitement des données

Qu'est-ce qu'un spectre?

- Interaction lumière/matière
- Influence de la composition chimique
- Influence de la physique
- Perturbations: électronique, contamination,...



Xu et al, 2019





La SPIR: génération et traitement des données

Les données SPIR, de A à Z

- Prétraitement des données

Qu'est-ce qu'un spectre?

- Interaction lumière/matière
- Influence de la composition chimique
- Influence de la physique
- Perturbations: électronique, contamination,...

Les traitement mathématiques permettant de prendre en compte ses différentes sources d'influences sont appelés les **prétraitements**

On en parle juste après...

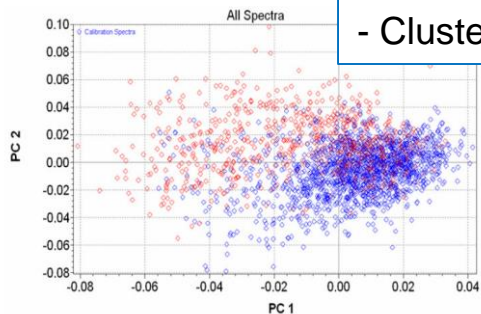
La SPIR: génération et traitement des données

Les données SPIR, de A à Z

- Présentation de l'échantillon
- Séparation des longueurs d'onde
- Prétraitement des données
- Exploitation des données

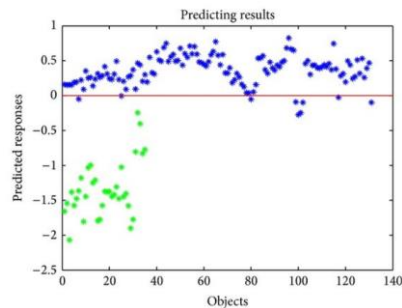
Méthodes exploratoires

- ACP
- Clustering



Méthodes prédictives

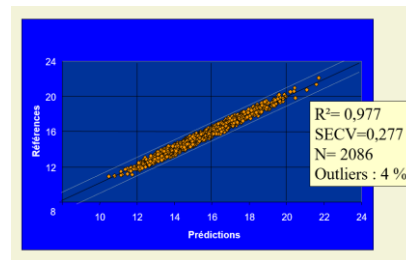
Modèles qualitatifs



Yan et al, 2014

Modèles quantitatifs

$$\begin{matrix} \text{réponses} & & \text{prédicteurs} & & \text{b-coefficients} & & \text{résidus} \\ \left[\begin{matrix} Y \end{matrix} \right] & = & \left[\begin{matrix} X \end{matrix} \right] \cdot \left[\begin{matrix} B \end{matrix} \right] + \left[\begin{matrix} R \end{matrix} \right] \end{matrix}$$





La SPIR: Applications en Agronomie

Une technique utilisée dans de nombreux domaines

Agro alimentaire / Agronomie

Pharmacie

Médecine

Sciences du sol

Géologie

Tri sélectif

Détection de fraude

Police scientifique

La SPIR: Applications en Agronomie

Sur les grains:

Teneur en protéines (CTPS)



Humidité / Matière sèche
Protéines
Sucres
Lipides

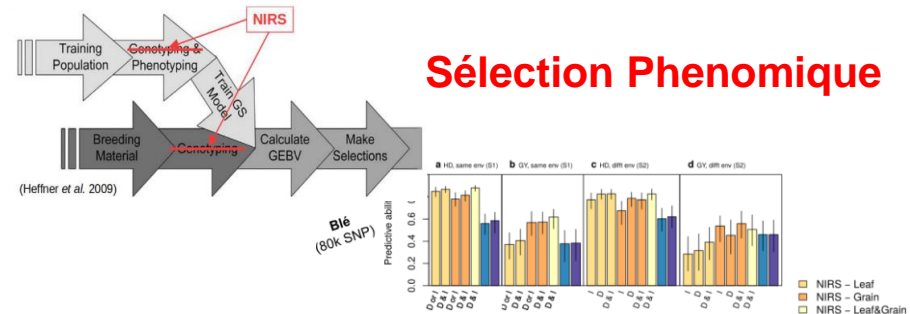


Soja OGM
Roundup Ready™



Soja conventionnel

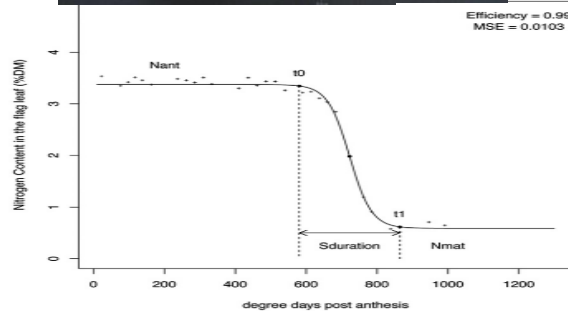
Bonne classif.
de 92,4%



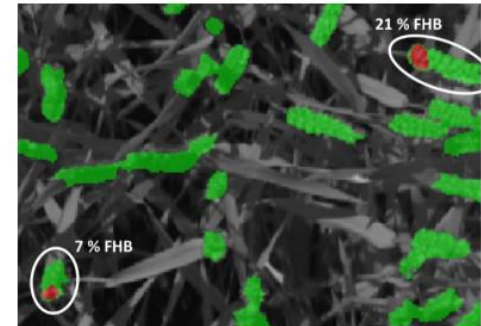
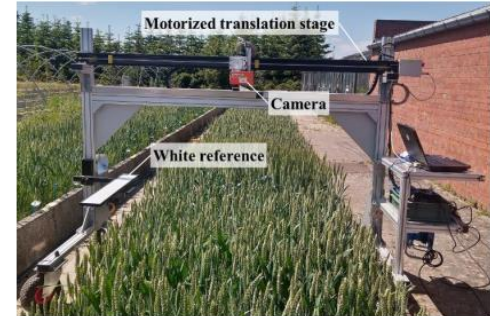
La SPIR: Applications en Agronomie

Sur les plantes:

Teneur en azote des feuilles



Estimation de la fusariose
(Thèse Damien Vincke CRA-W)



La SPIR: Applications en Agronomie

Sur les fruits:

Analyse du contenu en sucre de pommes en ligne



Echantillons

$$X_{2,1}$$

...

$$X_{3,1}$$

...

$$X_{4.1}$$

...

X_{5.1}

• • •

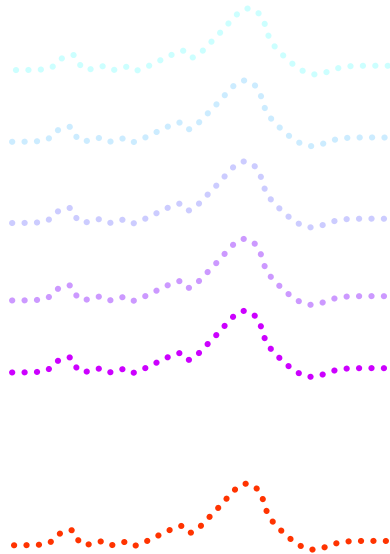
...

$$X_{n.1}$$

• • •

$$X_{n,p}$$
$$X_{n,p}$$
$$X_{n,p}$$
$$X_{n,p}$$
$$X_{n,p}$$

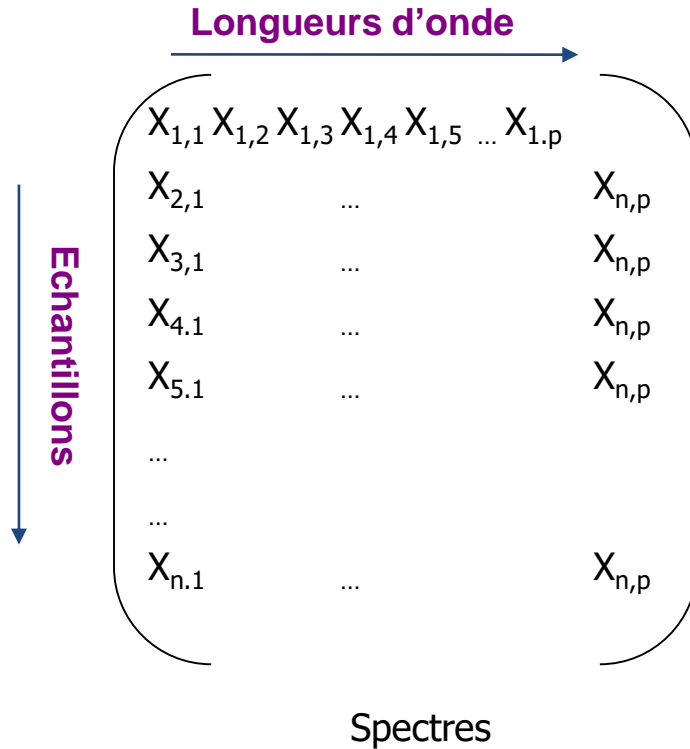
Longueurs d'onde



Spectres

La SPIR: Quelles précautions pour traiter les données

Entrez dans la matrice...



Dans la Matrice des spectres:

Lignes = Individus

Colonnes = Variables

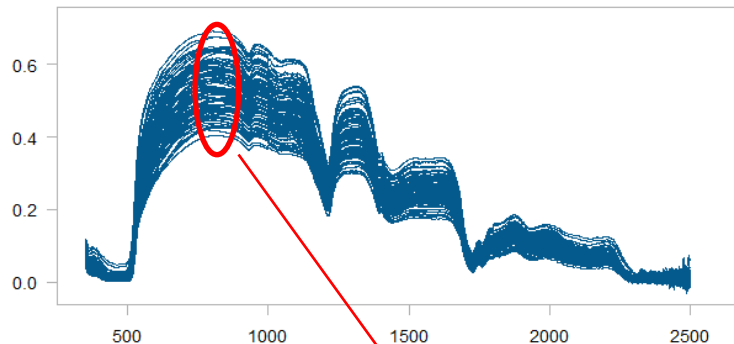
Exactement comme :

```
> head(iris)
  Sepal.Length Sepal.Width Petal.Length Petal.Width Species
1          5.1         3.5          1.4         0.2  setosa
2          4.9         3.0          1.4         0.2  setosa
3          4.7         3.2          1.3         0.2  setosa
4          4.6         3.1          1.5         0.2  setosa
5          5.0         3.6          1.4         0.2  setosa
6          5.4         3.9          1.7         0.4  setosa
```

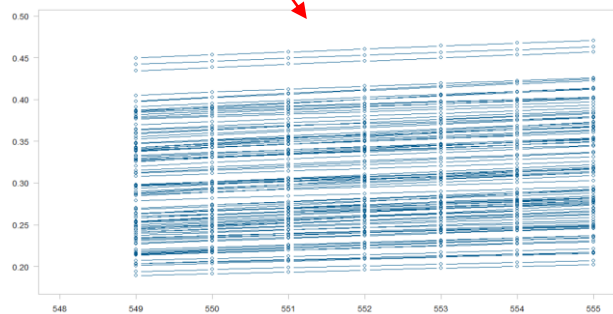
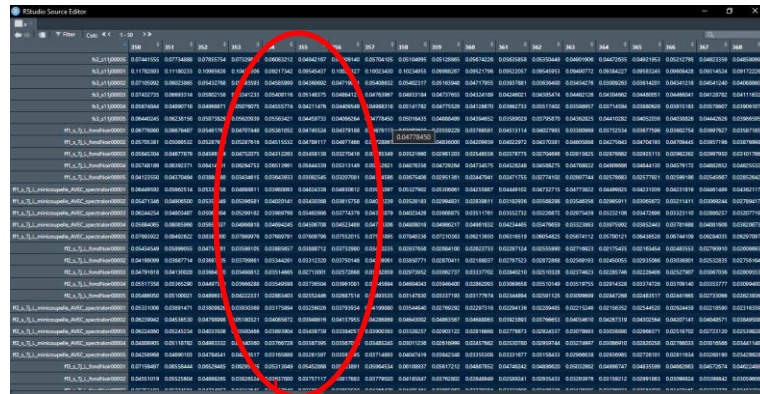



La SPIR: Comment traiter les données

Entrez dans la matrice...



=



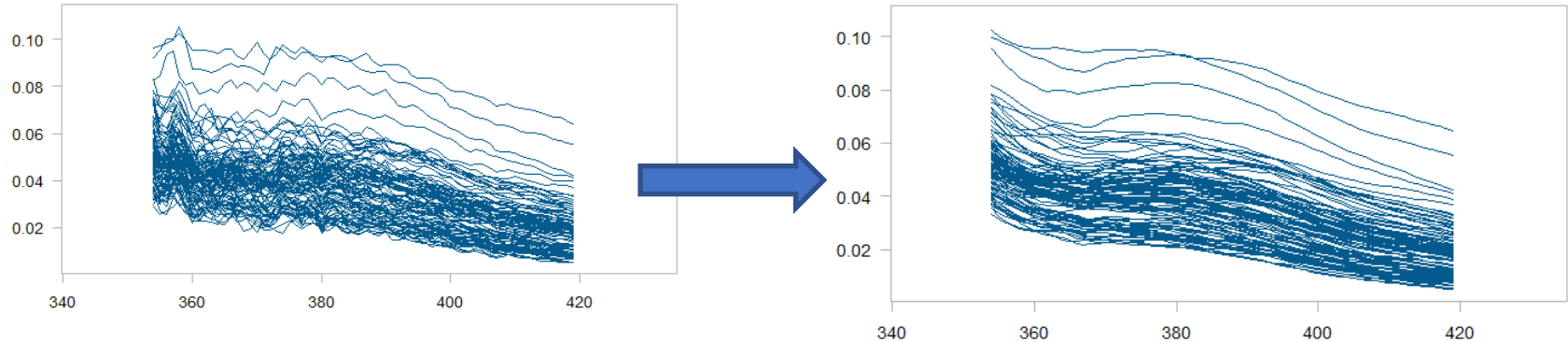
=

	549	550	551	552	553
fs2_s1j00005	0.3977160	0.4022721	0.4065713	0.4107540	0.4148661
fs3_s1j00001	0.4421590	0.4423801	0.4496612	0.4531382	0.4564772
fs3_s1j00002	0.3392124	0.3430121	0.3466758	0.3501895	0.3536320
fs3_s1j00003	0.3639590	0.3665730	0.3690074	0.3713782	0.3737252
fs3_s1j00004	0.3469997	0.3506492	0.3541951	0.3575716	0.3608734
fs3_s1j00005	0.3261270	0.3302771	0.3342712	0.3381723	0.3419623

La SPIR: Comment traiter les données

Les prétraitements pour corriger les perturbations

- Réduire le bruit « blanc »
 - Lissage. Par ex. moyenne glissante



La SPIR: Comment traiter les données

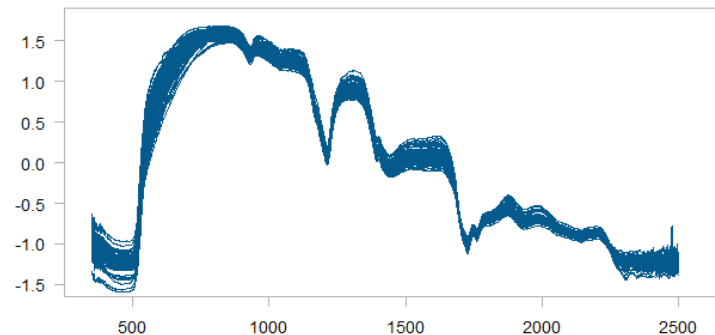
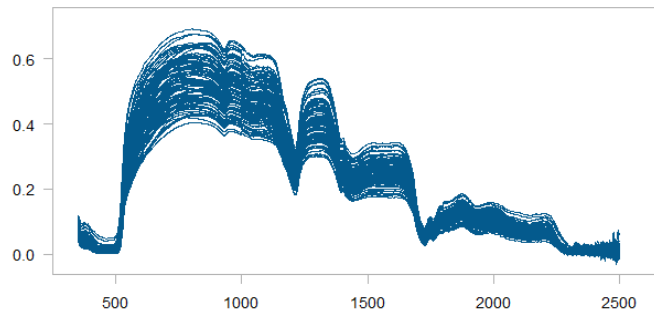
Les prétraitements pour corriger les perturbations

- Réduire un effet multiplicatif

- Normalisation SNV

Pour chaque spectre, on soustrait la moyenne, et on divise par l'écart-type

$$x_{i,j}^{SNV} = \frac{(x_{i,j} - \bar{x}_i)}{\sqrt{\frac{\sum_{j=1}^p (x_{i,j} - \bar{x}_i)^2}{p-1}}}$$

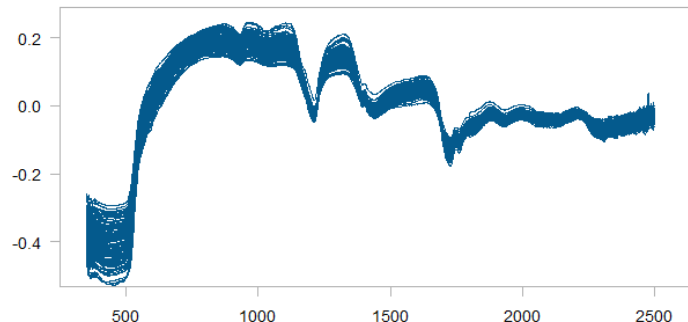
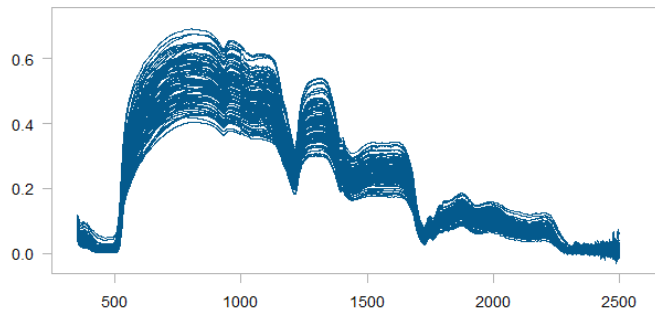


- Multi-scatter correction (MSC)
 - Logarithme

La SPIR: Comment traiter les données

Les prétraitements pour corriger les perturbations

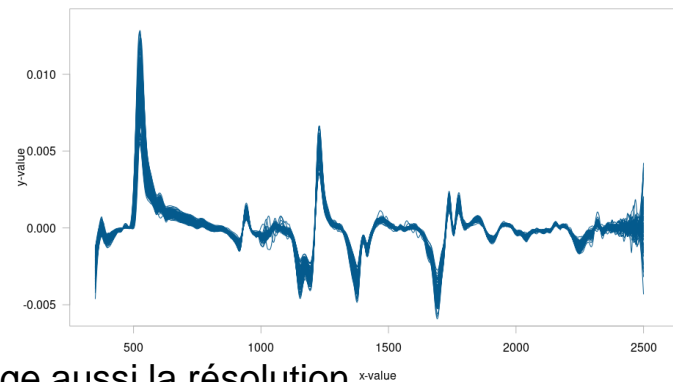
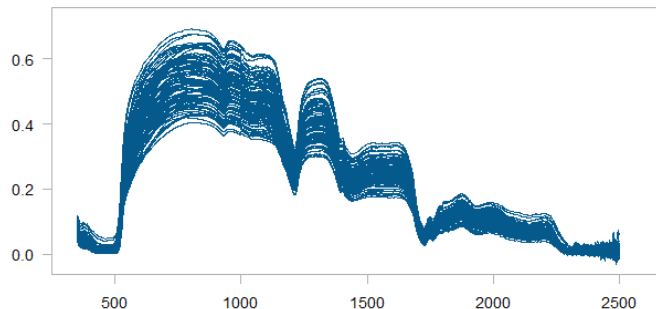
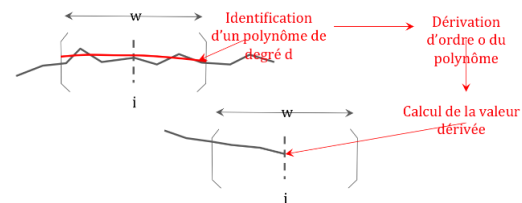
- Réduire une dérive de ligne de base
 - Algorithme « Detrend »: On soustrait un polynôme ajusté d'ordre (1, 2, ...)



La SPIR: Comment traiter les données

Les prétraitements pour corriger les perturbations

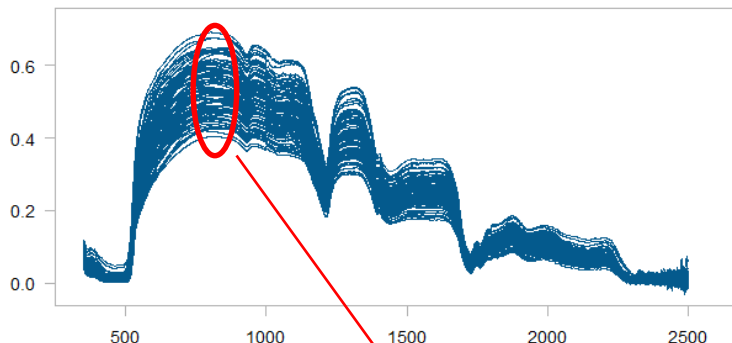
- Réduire une dérive de ligne de base
 - Dérivée (1^{ère}, 2^{ème})
- Algorithme de Savitzky-Golay



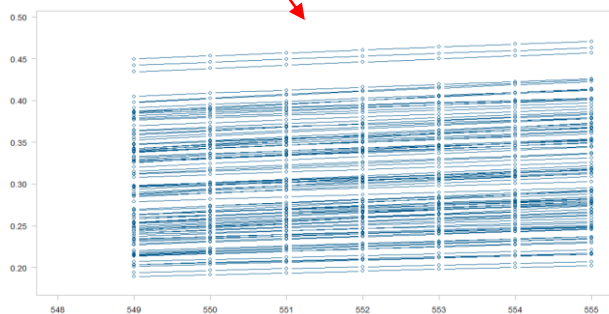
Change aussi la résolution
des pics!!

La SPIR: Comment traiter les données

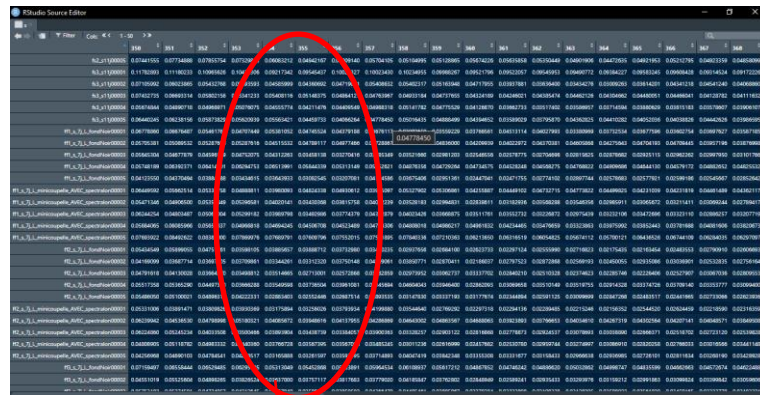
Entrez dans la matrice...



=



=

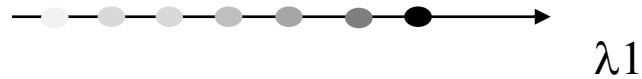


	549	550	551	552	553
fs2_s1j00005	0.3977160	0.4022721	0.4065713	0.4107540	0.4148661
fs3_s1j00001	0.4421590	0.4423801	0.4496612	0.4531382	0.4564772
fs3_s1j00002	0.3392124	0.3430121	0.3466758	0.3501895	0.3536320
fs3_s1j00003	0.3639590	0.3665730	0.3690074	0.3713782	0.3737252
fs3_s1j00004	0.3469997	0.3506492	0.3541951	0.3575716	0.3608734
fs3_s1j00005	0.3261270	0.3302771	0.3342712	0.3381723	0.3419623

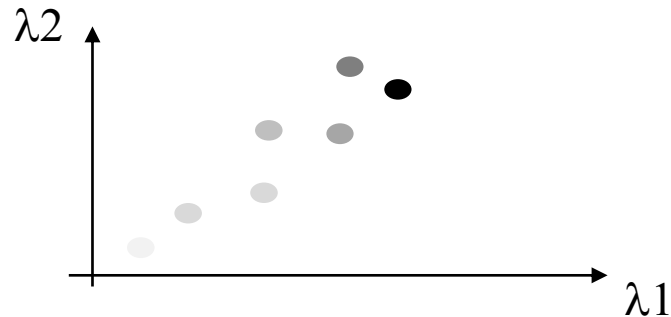
La SPIR: Comment traiter les données

Comment classer les spectres

Si le spectromètre mesure 1 longueur d'onde



Si le spectromètre mesure 2 longueurs d'onde



La SPIR: Comment traiter les données

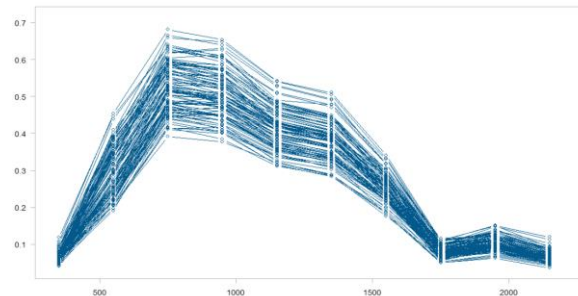
Comment classer les spectres

Si le spectromètre mesure 2000 longueurs d'onde



Quelles longueurs d'onde choisir?

- Si on choisit 1 longueur d'onde toutes les X...
 - On ne voit plus les pics
- Si on choisit une gamme réduite
 - Très forte colinéarité entre les variables



La SPIR: Comment traiter les données

Comment classer les spectres

Si le spectromètre mesure 2000 longueurs d'onde

Solution?

.....



La SPIR: Comment traiter les données

Comment classer les spectres

Si le spectromètre mesure 2000 longueurs d'onde

Quelles longueurs d'onde choisir?

Analyse en Composante principales!

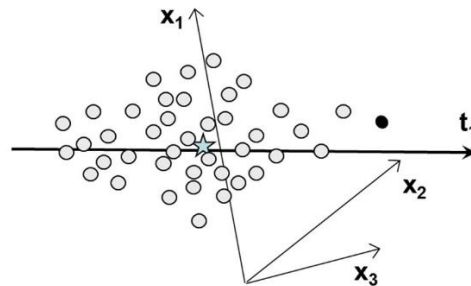
La SPIR: Comment traiter les données

Analyse en Composante principales

- Pour explorer les données (classement d'individus, outliers)
- Déterminer l'importance des longueurs d'onde

Une composante principale = une combinaison linéaire des longueurs d'onde

$$CP1 = \alpha \times \lambda_1 + \beta \times \lambda_2 + \dots + \gamma \times \lambda_p$$



La SPIR: Comment traiter les données

Modéliser la matrice X , en :

- loadings V (= poids des variables dans le modèle)
- scores C (= « coordonnées » des individus)

$$X = C V^T + E \quad (E = \text{résidu de modélisation})$$

Longueurs d'onde →

Echantillons ↓

$X_{1,1}$	$X_{1,2}$	$X_{1,3}$	$X_{1,4}$	$X_{1,5}$... $X_{1,p}$
$X_{2,1}$...			$X_{n,p}$
$X_{3,1}$...			$X_{n,p}$
$X_{4,1}$...			$X_{n,p}$
$X_{5,1}$...			$X_{n,p}$
...					
...					
$X_{n,1}$...			$X_{n,p}$

Spectres

X



CP1 = $\alpha \times \lambda_1 + \beta \times \lambda_2 + \dots + \gamma \times \lambda_p$

Composantes Principales →

Echantillons ↓

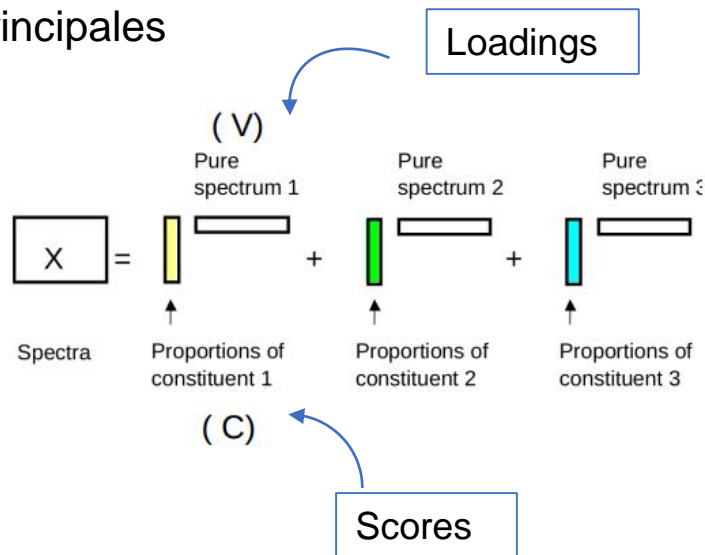
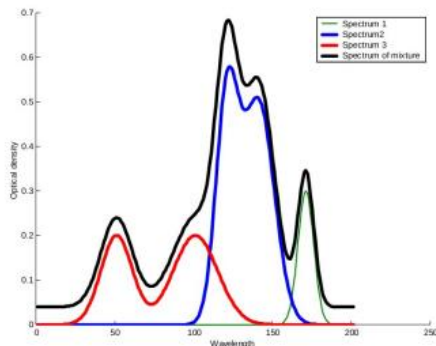
$s_{1,1}$	$s_{1,2}$	$s_{1,3}$	$s_{1,4}$	$s_{1,5}$... $s_{1,k}$
$s_{2,1}$...			$s_{n,k}$
$s_{3,1}$...			$s_{n,k}$
$s_{4,1}$...			$s_{n,k}$
$s_{5,1}$...			$s_{n,k}$
...					
...					
$s_{n,1}$...			$s_{n,k}$

Scores

C

La SPIR: Comment traiter les données

Analyse en Composante principales



$$CP1 = \alpha \times \lambda_1 + \beta \times \lambda_2 + \dots + \gamma \times \lambda_p$$

On peut évaluer les longueurs d'onde les plus importantes pour le modèle.

La SPIR: Comment traiter les données

Rstudio est ton ami!