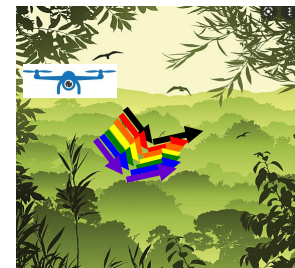


# Sélection de bandes hyperspectrales pour la prédiction non destructive des traits bio-chimiques des feuilles végétales

Catherine Baskiotis, Josselin Aval, Marwa El Bouz, Ayman Alfalou  
Light Scattering Learning Group, L@bisen, ISEN site de Brest, Yncréa-Ouest

[catherine.baskiotis@isen-ouest.yncrea.fr](mailto:catherine.baskiotis@isen-ouest.yncrea.fr)



## Suivi des paramètres physico-chimiques des feuilles :

- Chlorophylle a
- Chlorophylle b
- Carotène
- Teneur en eau
- Masse surfacique
- ...



**pour :**

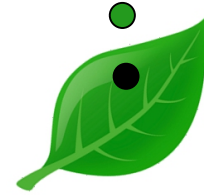
- Agriculture intelligente
- Détection précoce des maladies des végétaux
- Surveillance/Protection environnementale

## Deux grands types de méthodes :

- Méthodes physico-chimiques **destructives**  
*(chronophages, utilisation massive impossible)*

### Chlorophylle

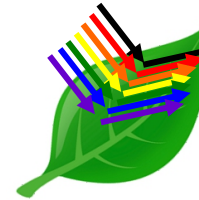
- Broyage,
- centrifugation,
- dissolution dans l'acétone



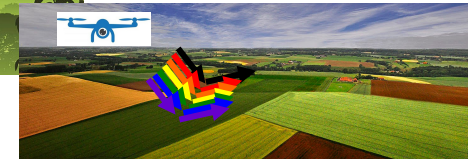
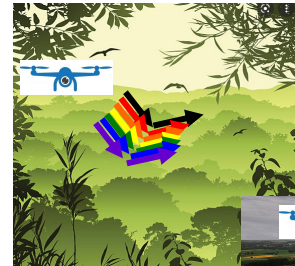
### Masse Foliaire surfacique




- Séchage dans un four,
- Pesée avant et après

- Méthodes d'imagerie hyper-spectrale **non destructives**  
*(instantanées, utilisation massive possible)*



Réflectance



VISIBLE	PROCHE INFRAROUGE	INFRAROUGE COURT
		
0.4-0.75 ( $\mu\text{m}$ )	0.75-1 ( $\mu\text{m}$ )	1-2.5 ( $\mu\text{m}$ )
Pas : 1nm	1nm	1nm
Résolution : 1-2nm	1-2nm	4-5nm

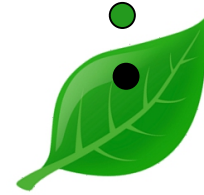
~2000 bandes hyperspectrales

# Contexte

## Deux grands types de méthodes :

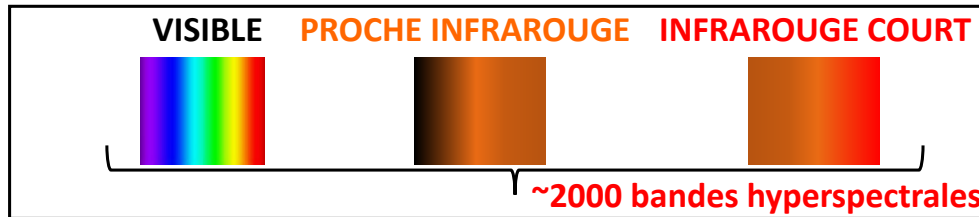
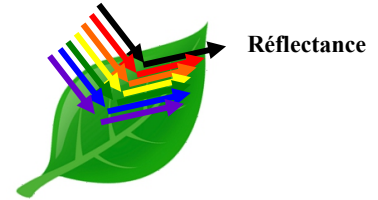
- Méthodes physico-chimiques **destructives**  
*(chronophages, utilisation massive impossible)*

Chlorophylle  
[- Broyage,  
- centrifugation,  
- dissolution dans l'acétone



Masse Foliaire surfacique  
[- Séchage dans un four,  
- Pesée avant et après

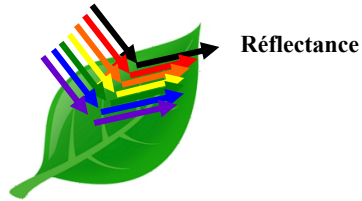
- Méthodes d'imagerie hyper-spectrale **non destructives**  
*(instantanées, utilisation massive possible)*



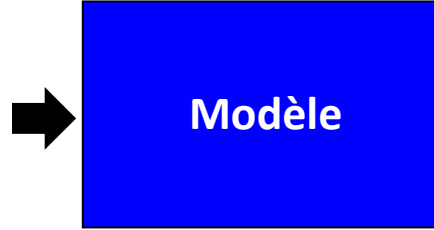
**Problème ouvert :**

**plus petit nombre de bandes hyperspectrales nécessaire à une prédiction précise des traits bio-chimiques ?**

Entrées :  
Réflectance



Réflectance



Sortie:

- Paramètre bio-chimique  
(Chl. a, Chl. b, Carotène, Masse Foliaire Surfactive, ...)

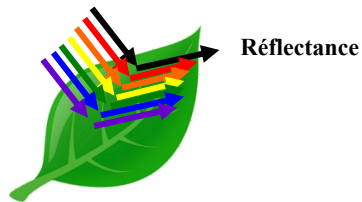
## ■ Modèle physique PROSPECT [1]

- Meilleure précision pour la prédiction des paramètres bio-chimiques
  - Chl. a, Chl. b, Carotène : erreur\* de 10-12%
  - Masse Foliaire Surfactive : erreur\* de 18%
- Plus de 100 bandes hyperspectrales

[1] J.-B. Féret et al., "Estimating leaf mass per area and equivalent water thickness based on leaf optical properties: Potential and limitations of physical modeling and machine learning", *Remote Sensing of Environment*, 231, pp. 110959, 2019.

\*Racine de l'erreur quadratique Moyenne Relative =  $\frac{\text{Racine de l'erreur quadratique moyenne}}{\text{Moyenne du parameter bio-chimique considéré}}$

Entrées :  
Réflectance



Réflectance

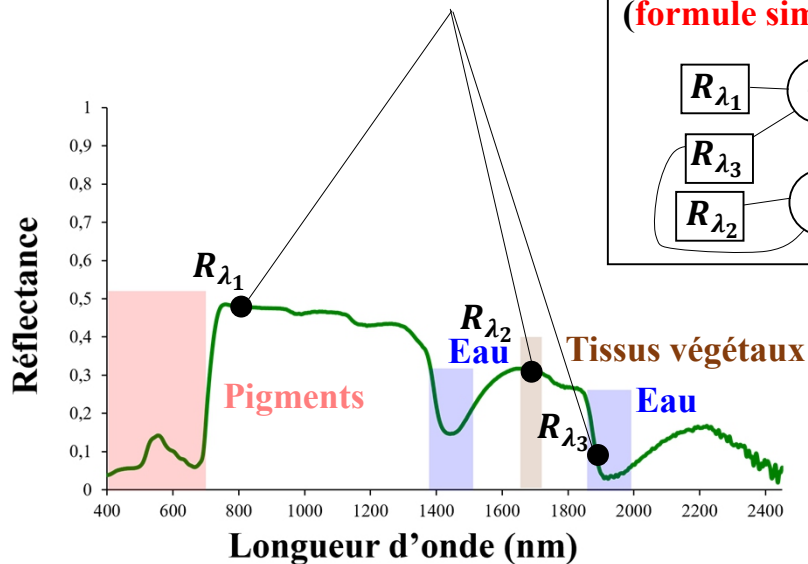
Modèle

Sortie:

- Paramètre bio-chimique  
(Chl. a, Chl. b, Carotène, Masse Foliaire Surfacique, ...)

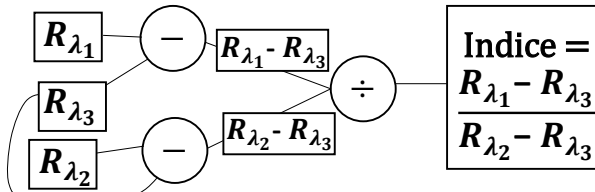
## ■ Méthodes des indices

### 1) Sélection de bandes (2 à 6)



### 2) Choix d'un indice

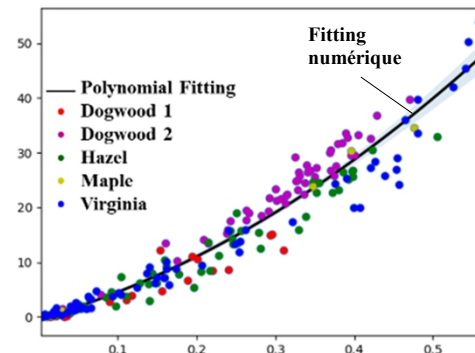
(formule simple des réflectances de bandes)



$$\text{Indice} = \frac{R_{\lambda_1} - R_{\lambda_3}}{R_{\lambda_2} - R_{\lambda_3}}$$

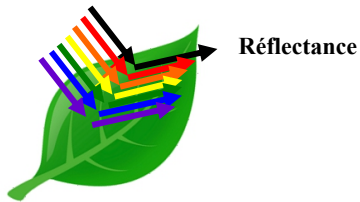
### 3) Fitting numérique

Paramètre biochimique (Chl. a,  $\mu\text{g}/\text{cm}^2$ )



$$\text{Indice} = \frac{R_{\lambda_1} - R_{\lambda_3}}{R_{\lambda_2} - R_{\lambda_3}}$$

Entrées :  
Réflectance



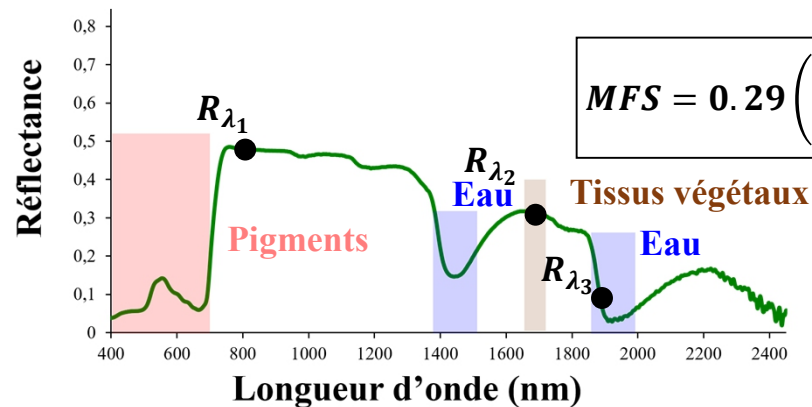
**Modèle**

Sortie:

- Paramètre bio-chimique  
(Chl. a, Chl. b, Carotène, Masse Foliaire Surfactive, ...)

## ■ Méthodes des indices

- 2 à 6 bandes hyperspectrales
- Précision presque identique à celle obtenue avec le modèle physique PROSPECT pour la Chl. a\*, Chl. b\* et la Carotène\*
- Précision très inférieure à celle obtenue avec le modèle physique pour la **Masse Foliaire Surfactive (MFS)**, **erreur de 28%**
- **Méthode de recherche des indices de force brute limitée par les puissances actuelles de calculs [2,3]**



$$MFS = 0.29 \left( \frac{R_{1710nm} - R_{1340nm}}{R_{1710nm} + R_{1340nm}} \right)^2 - 0.014 \left( \frac{R_{1710nm} - R_{1340nm}}{R_{1710nm} + R_{1340nm}} \right) - 1.7 \cdot 10^{-4}$$

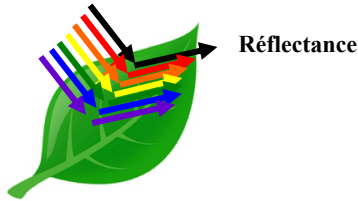
$$Chl. a = -1.8 \left( \frac{R_{750nm}}{R_{710nm}} \right)^2 + 35.1 \left( \frac{R_{750nm}}{R_{710nm}} \right) - 33.2$$

\* Erreur ~10-12% si dicotylédons

[2] Y. Chen et al., *Remote Sensing*, **13**, pp. 3761, 2021.

[3] G. le Maire et al., *Remote Sensing of Environment*, **112**, pp. 3846-3864, 2008.

Entrées :  
Réflectance



Réflectance

Modèle

Sortie:

- Paramètre bio-chimique  
(Chl. a, Chl. b, Carotène, Masse Foliaire Surfaique, ...)

## ■ Méthodes des indices

- 2 à 6 bandes hyperspectrales
- Précision très inférieure à celle obtenue avec le modèle physique pour la **Masse Foliaire Surfaique (MFS)**, erreur de 28%
- Méthode de recherche des indices de force brute limitée par la puissance actuelle de calculs [2,3]

$$MFS = 0.29 \left( \frac{R_{1710nm} - R_{1340nm}}{R_{1710nm} + R_{1340nm}} \right)^2 - 0.014 \left( \frac{R_{1710nm} - R_{1340nm}}{R_{1710nm} + R_{1340nm}} \right) - 1.7 \cdot 10^{-4}$$

## ■ Régression linéaire par moindres carrés partiels (PLSR)

- ~15 bandes hyperspectrales [1]
  - Précision médiocre pour la Masse Foliaire Surfaique (MFS), erreur de 38%
- [la PLSR se restreint au cas d'une relation linéaire entre les réflectances et la MFS, à la différence de la méthode des indices]

[2] Y. Chen et al., *Remote Sensing*, **13**, pp. 3761, 2021.

[3] G. le Maire et al., *Remote Sensing of Environment*, **112**, pp. 3846-3864, 2008.



## Littérature :

- Modèle physique PROSPECT utilisant plus de 100 bandes hyperspectrales [1]

*Meilleure précision actuelle pour la Masse Foliaire Surfaccique, erreur ~18%*

- Méthodes des indices utilisant 2 à 6 bandes hyperspectrales [2,3]

*Faible précision pour la Masse Foliaire Surfaccique, erreur ~28%*

- Régression linéaire par moindres carrés partiels utilisant ~15 de bandes hyperspectrales [2]

*Très faible précision pour la Masse Foliaire Surfaccique, erreur ~38%*

## But de l'étude :

**Augmenter la précision de prédiction de la Masse Foliaire Surfaccique  
avec un faible nombre de bandes hyperspectrales (<15)**

[1] J.-B. Féret et al., "Estimating leaf mass per area and equivalent water thickness based on leaf optical properties: Potential and limitations of physical modeling and machine learning", *Remote Sensing of Environment*, **231**, pp. 110959, 2019.

[2] Y. Chen et al., "Optimized Estimation of Leaf Mass per Area with a 3D Matrix of Vegetation Indices", *Remote Sensing*, **13**, pp. 3761, 2021.

[3] G. le Maire et al., "Calibration and validation of hyperspectral indices for the estimation of broadleaved forest leaf chlorophyll content, leaf mass per area, leaf area index and leaf canopy biomass", *Remote Sensing of Environment*, **112**, pp. 3846-3864, 2008.

**1. Introduction**

**→ 2. Matériau expérimental**

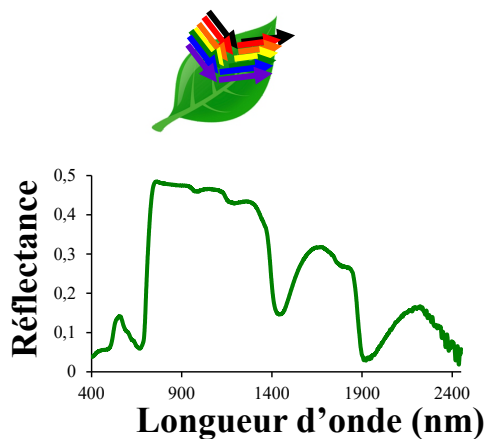
**3. Sélection des bandes par Réseaux de Neurones**

**4. Résultats**

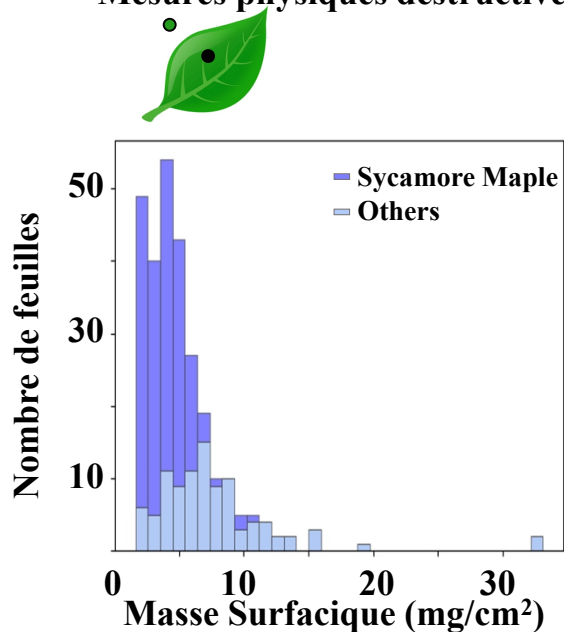
**5. Conclusion**

- Données testées : base de données publique *Angers*,  
274 échantillons, 2050 bandes hyperspectrales
- Paramètre bio-chimique testé : Masse Foliaire Surfaique

Mesures hyper-spectrales



Mesures physiques destructives



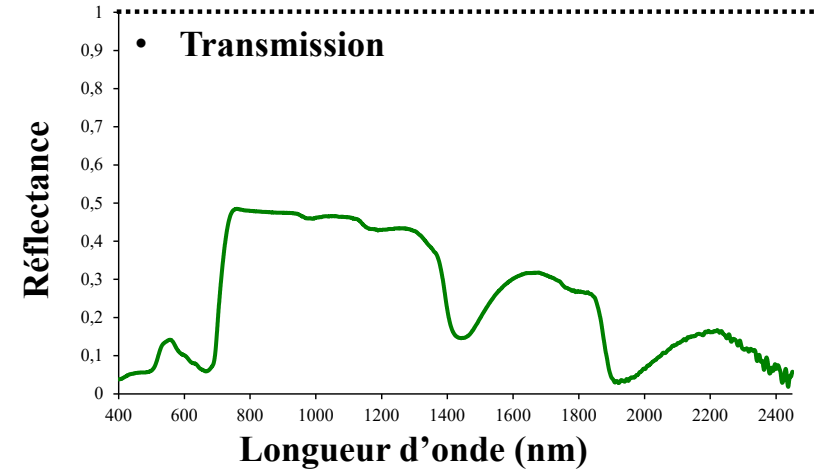
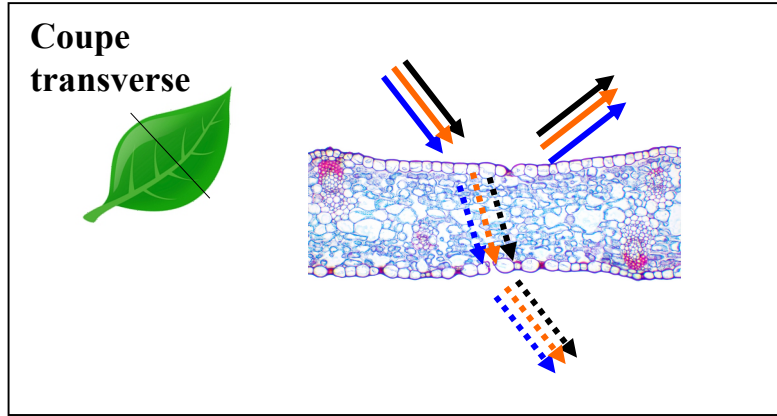
Référence (vérité terrain) :

- mesures physiques destructives de la  
masse surfaique

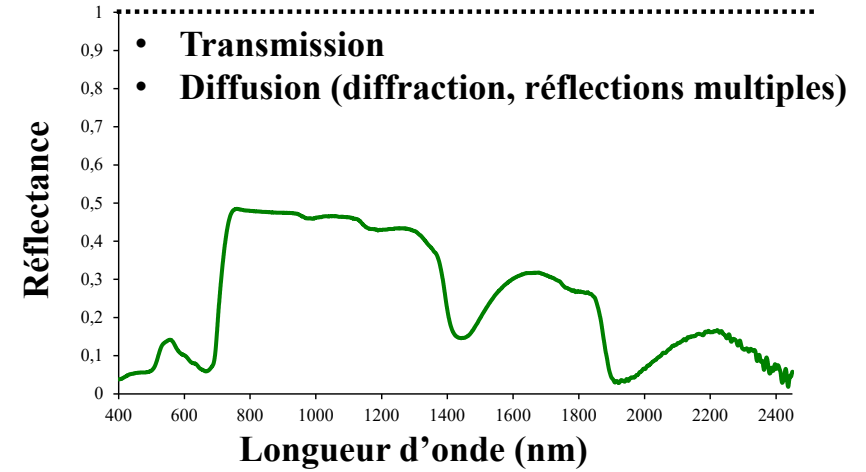
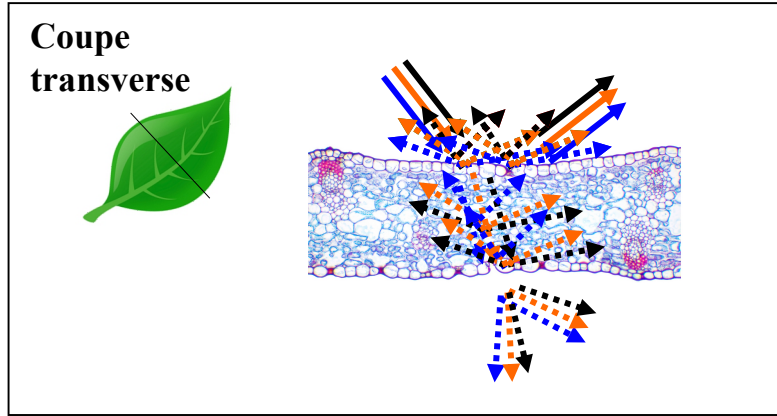
Modèles :

- prédiction de la masse surfaique à  
partir des mesures hyper-spectrales

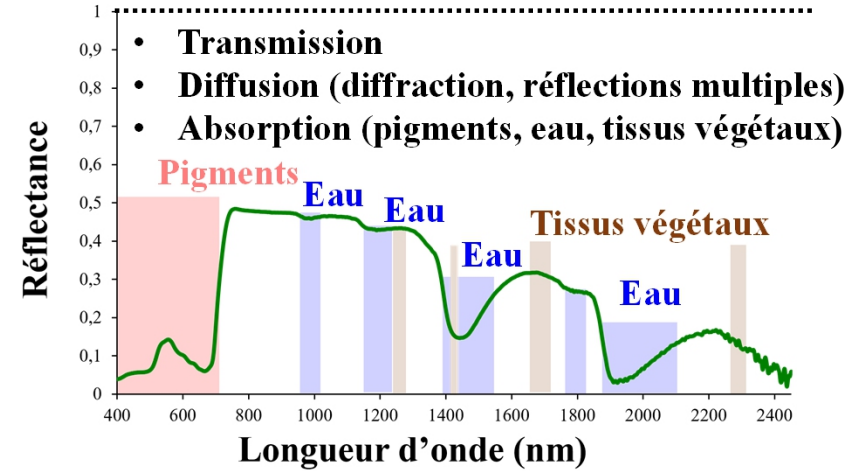
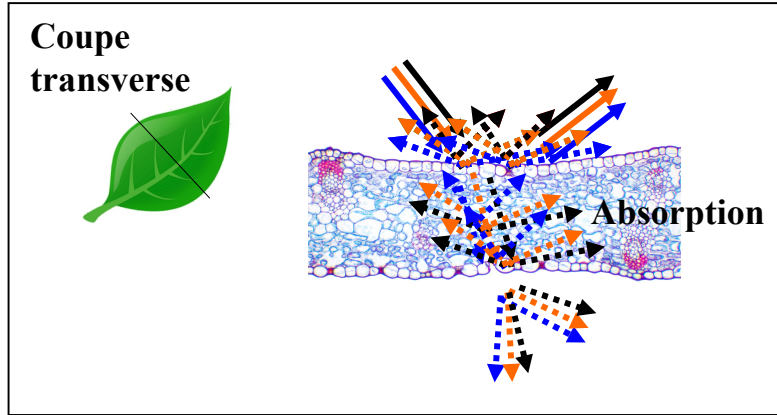
# Matériau Expérimental



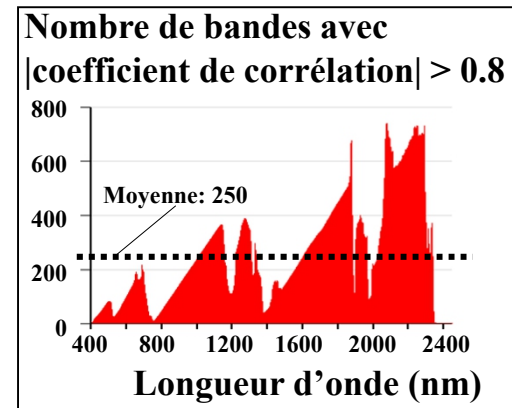
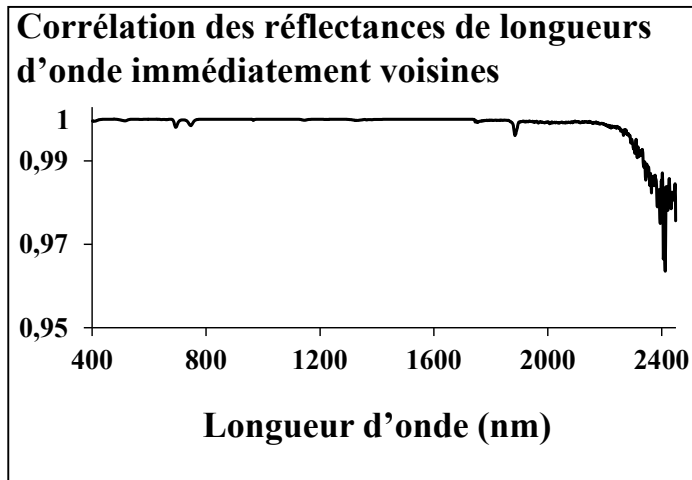
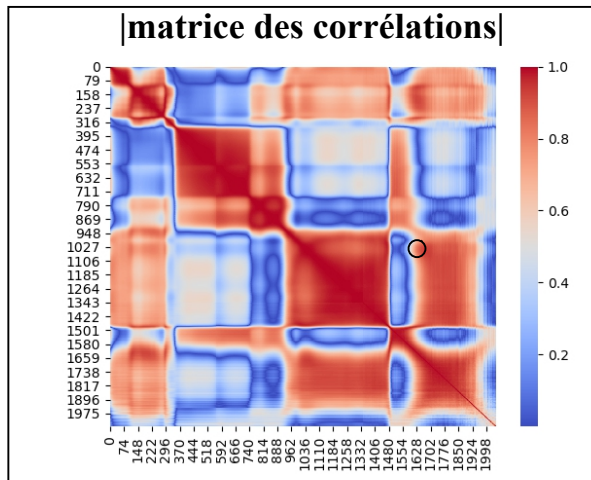
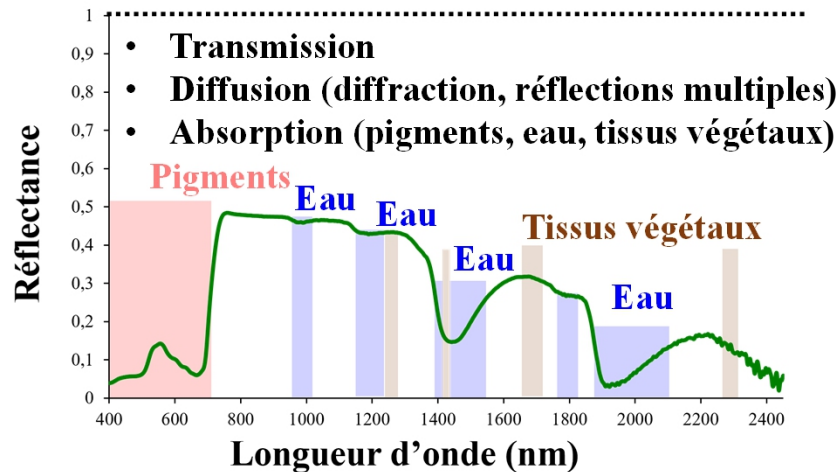
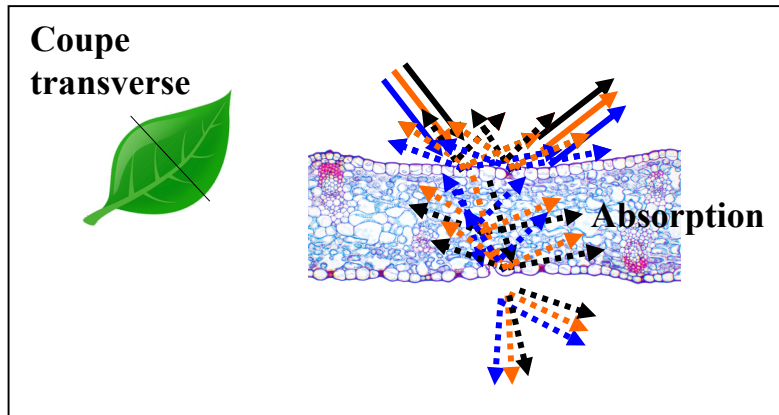
# Matériau Expérimental



# Matériau Expérimental



# Matériau Expérimental



**1. Introduction**

**2. Matériau expérimental**

**→ 3. Sélection des bandes par Réseaux de Neurones**

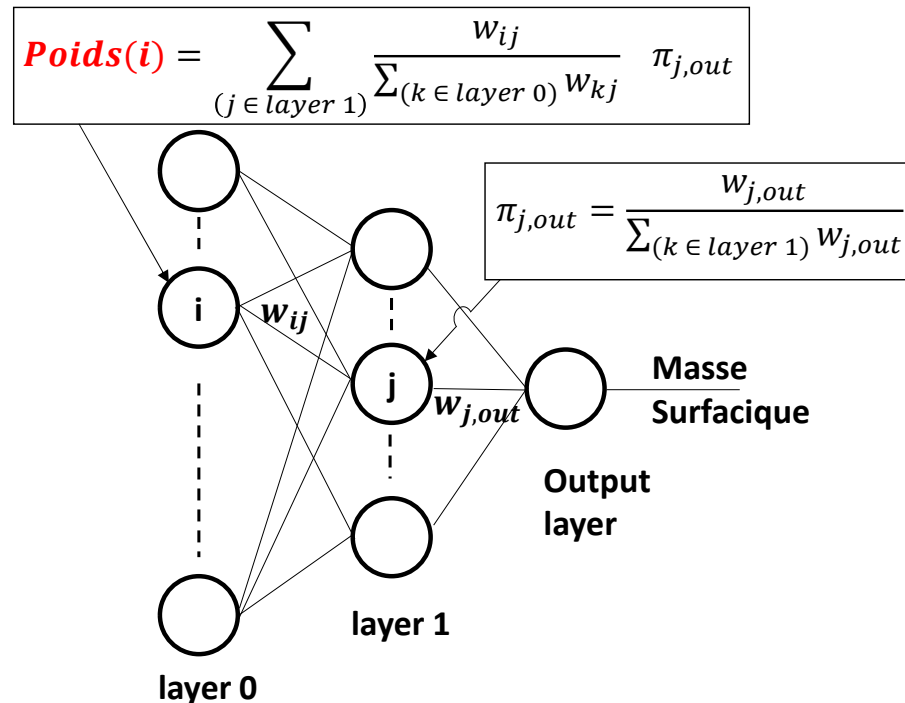
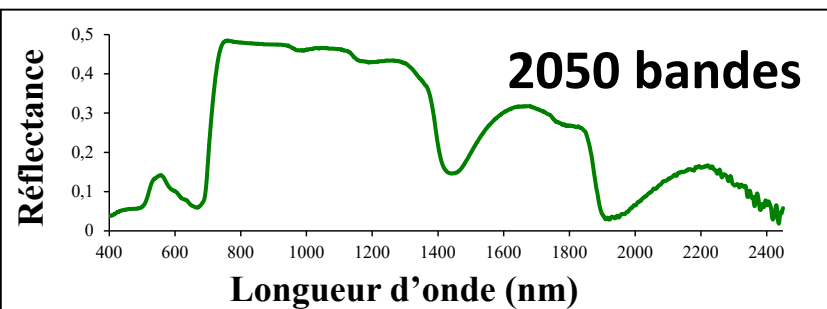
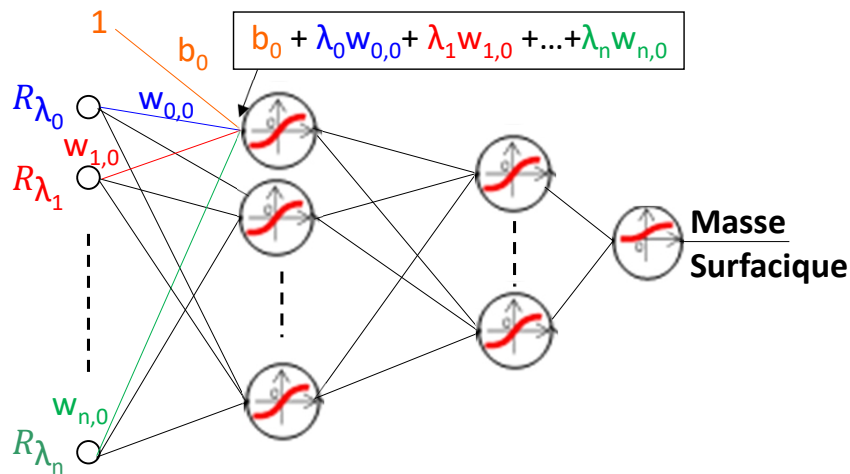
**4. Résultats**

**5. Conclusion**



# Sélection de longueurs d'onde par réseaux de neurones : algorithme

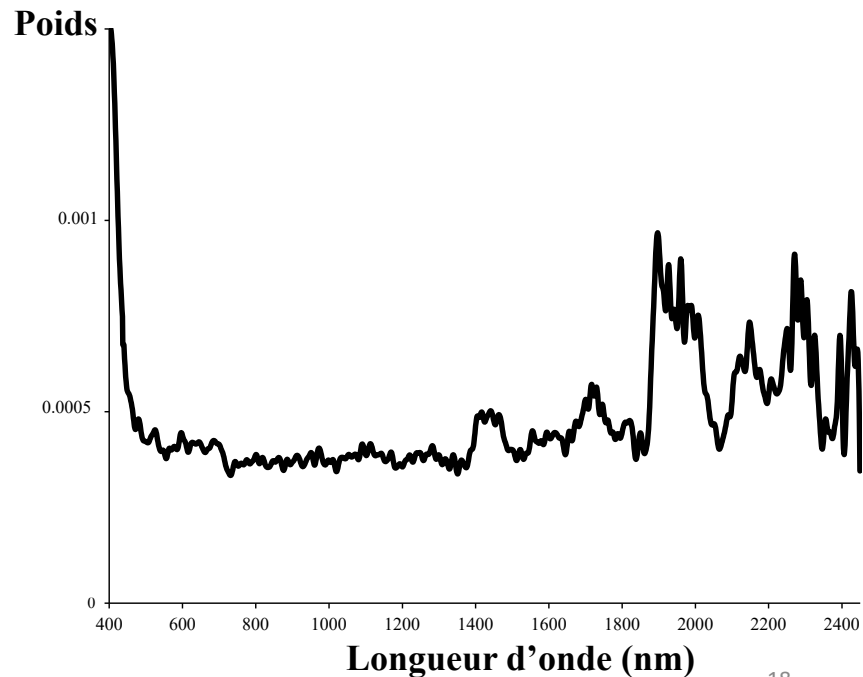
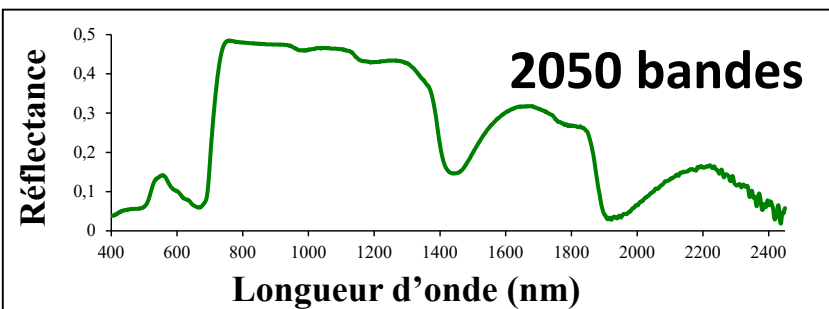
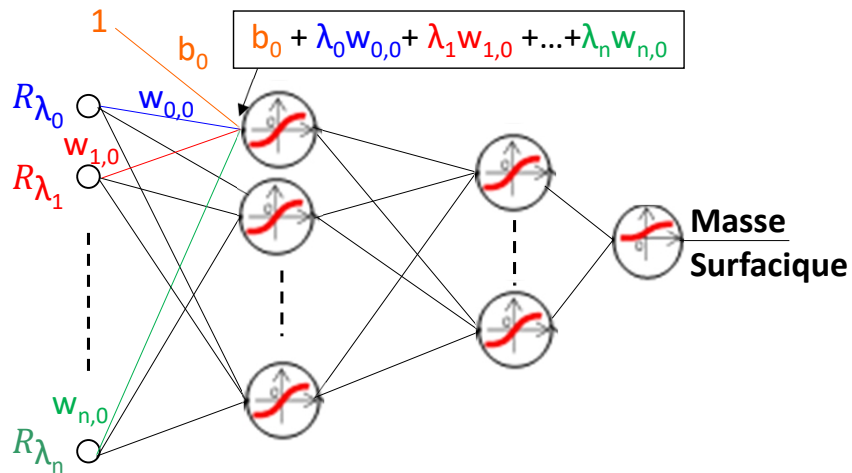
## Etape 1 : calculs des poids [Yacoub et Bennani, IJNS, 2000]



[Yacoub et Bennani, IJNS, 2000] M. Yacoub, Y. Bennani, "Features selection and architecture optimization in connectionist systems", *International Journal of Neural Systems*, **10**, pp. 379-95, 2000.

# Sélection de longueurs d'onde par réseaux de neurones : algorithme

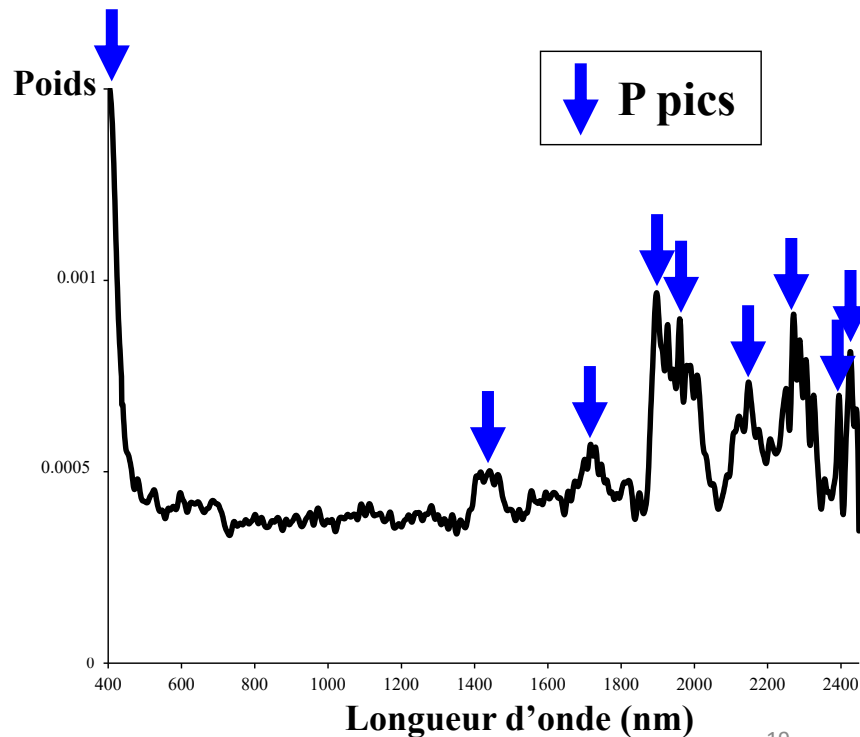
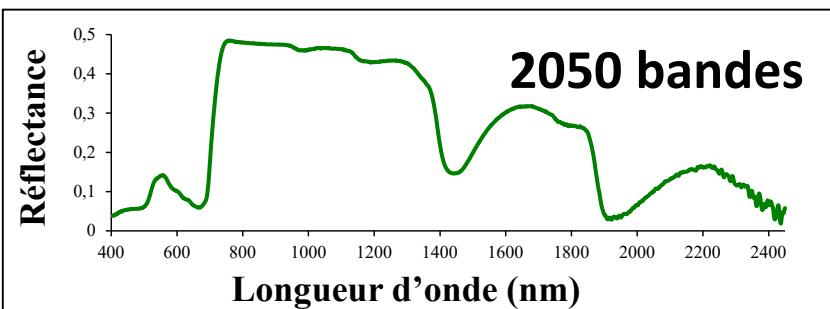
## Etape 1 : calculs des poids [Yacoub et Bennani, IJNS, 2000]



# Sélection de longueurs d'onde par réseaux de neurones : algorithme

Etape 1 : calculs des poids [Yacoub et Bennani, IJNS, 2000]

Etape 2 : sélection de P pics



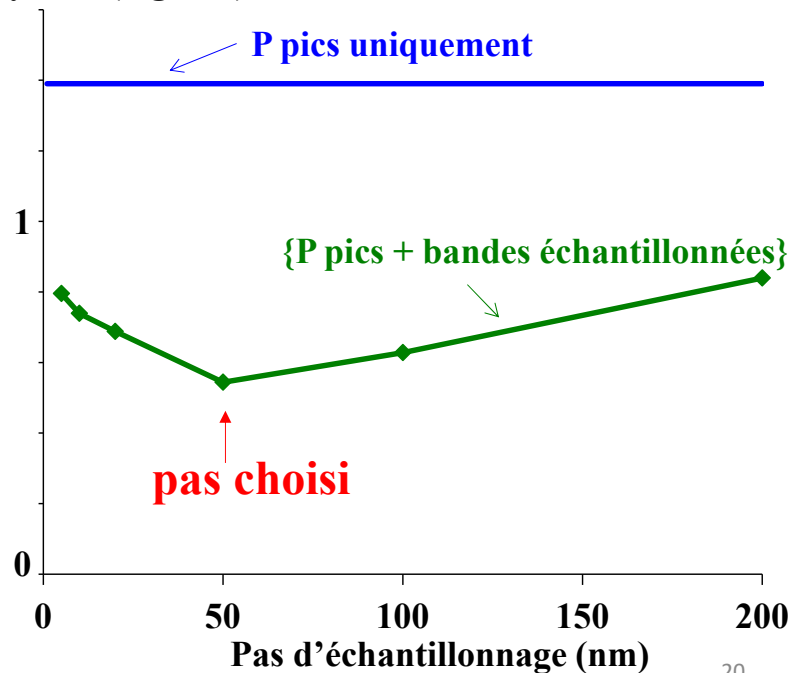
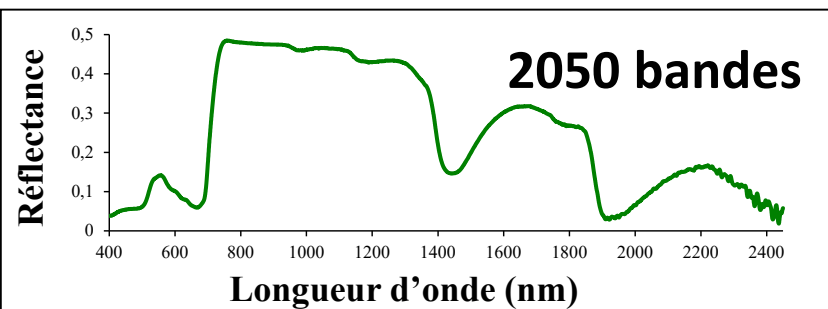
# Sélection de longueurs d'onde par réseaux de neurones : algorithme

Etape 1 : calculs des poids [Yacoub et Bennani, IJNS, 2000]

Etape 2 : sélection de P pics

Etape 3 : choix du pas d'échantillonnage

Racine de l'erreur quadratique  
moyenne (mg/cm<sup>2</sup>)



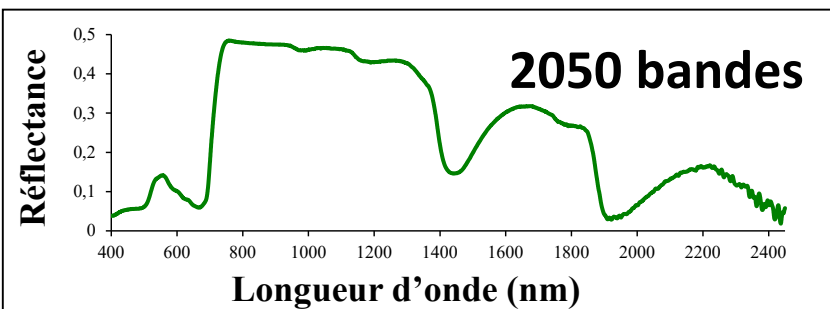
# Sélection de longueurs d'onde par réseaux de neurones : algorithme

Etape 1 : calculs des poids [Yacoub et Bennani, IJNS, 2000]

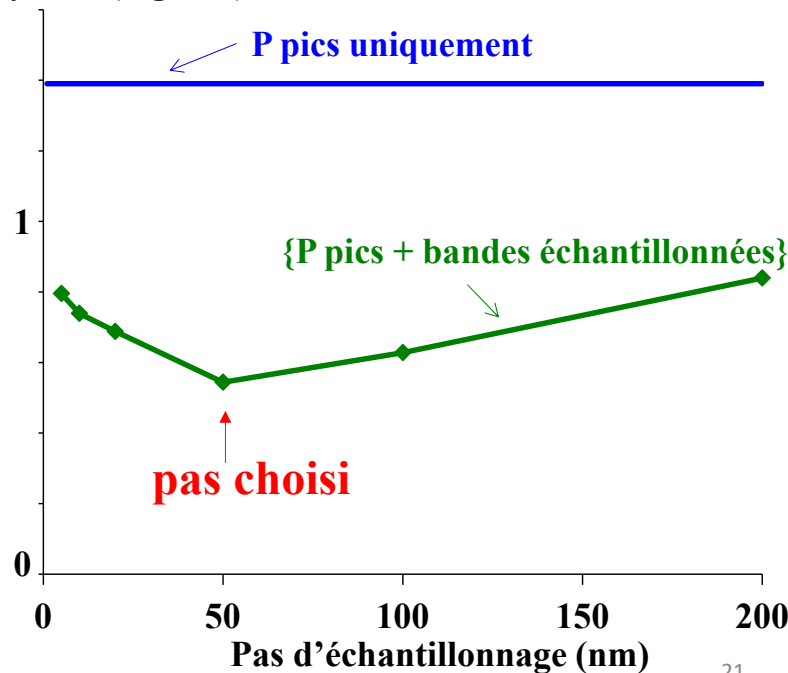
Etape 2 : sélection de P pics

Etape 3 : choix du pas d'échantillonnage

➔ **~50 bandes**



Racine de l'erreur quadratique moyenne (mg/cm<sup>2</sup>)



# Sélection de longueurs d'onde par réseaux de neurones : algorithme

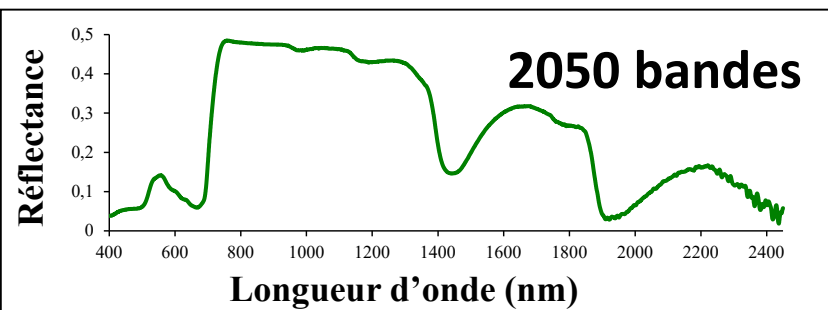
Etape 1 : calculs des poids [Yacoub et Bennani, IJNS, 2000]

Etape 2 : sélection de P pics

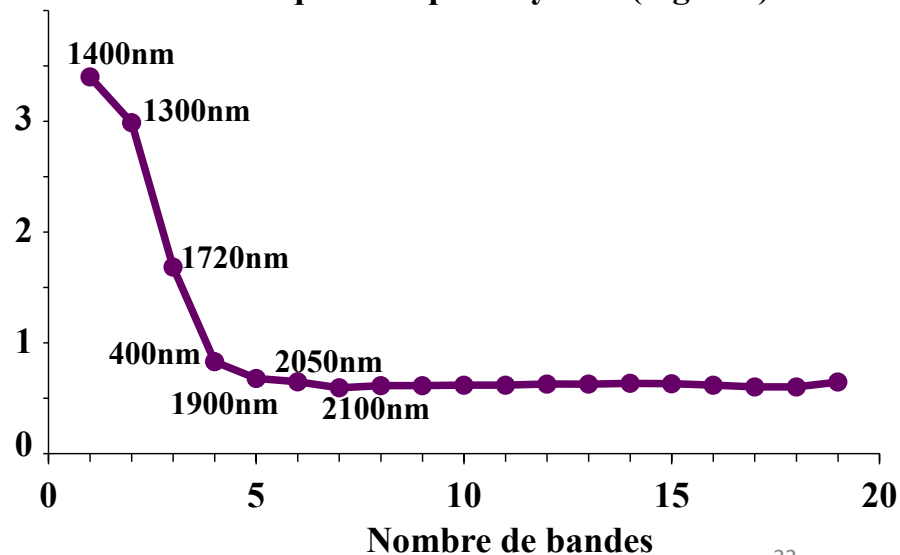
Etape 3 : choix du pas d'échantillonnage

➔ ~50 bandes

Etape 4 : boucle de suppression 1 par 1 des bandes à plus faible poids



Racine de l'erreur quadratique moyenne (mg/cm<sup>2</sup>)



# Sélection de longueurs d'onde par réseaux de neurones : algorithme

Etape 1 : calculs des poids [Yacoub et Bennani, IJNS, 2000]

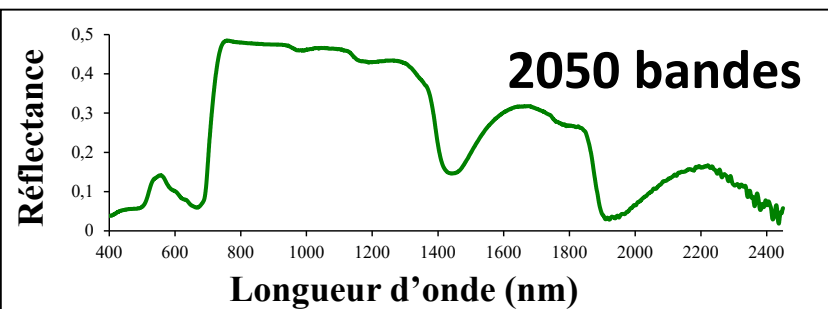
Etape 2 : sélection de P pics

Etape 3 : choix du pas d'échantillonnage

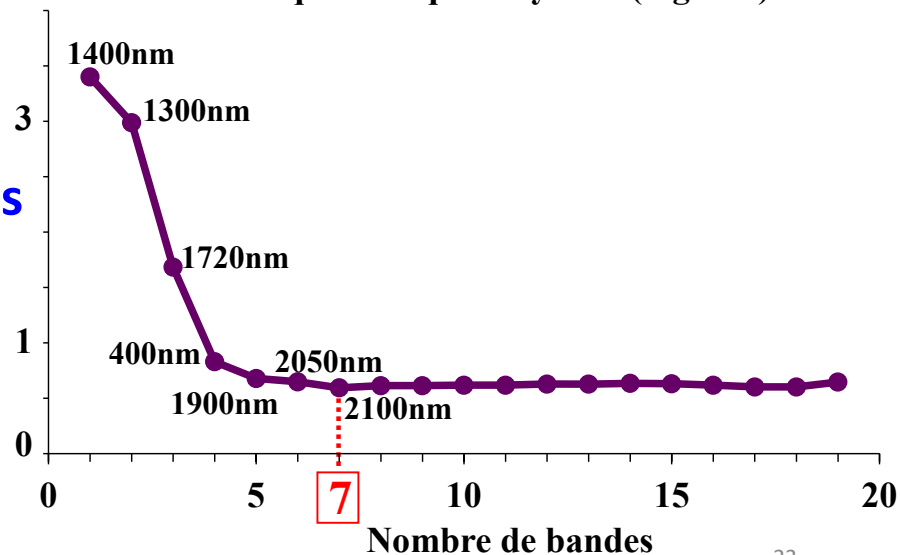
➔ ~50 bandes

Etape 4 : boucle de suppression 1 par 1  
des bandes à plus faible poids

Etape 5 : choix du nombre final de bandes



Racine de l'erreur quadratique moyenne (mg/cm<sup>2</sup>)



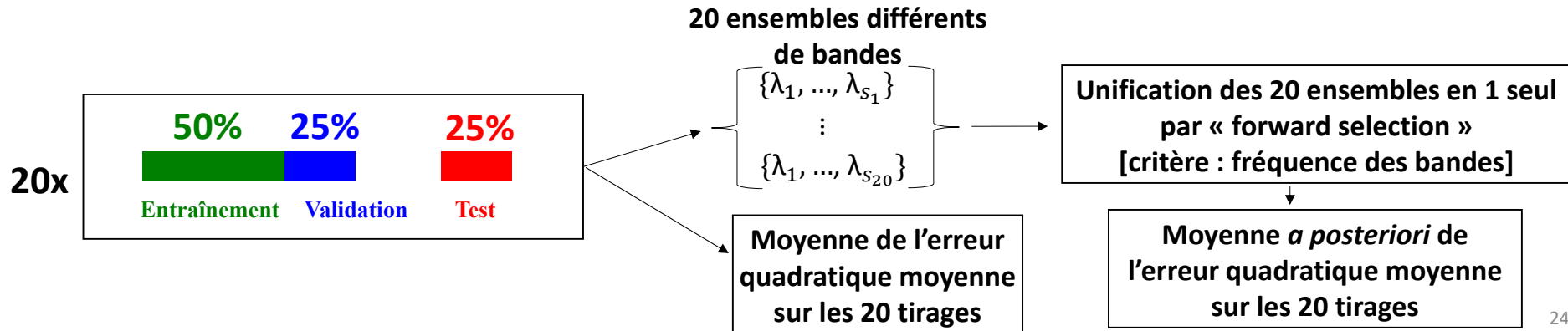
# Sélection de longueurs d'onde par réseaux de neurones : test

- Protocole de test :

- 20 tirages aléatoires en ensemble d'entraînement, de validation et de test (50%, 25%, 25%)
- Seuls les ensembles d'entraînement et de validation sont utilisés pour sélectionner les longueurs d'onde (et pas l'ensemble de test)
- 20 ensembles différents de bandes spectrales finalement sélectionnées (1 par tirage)

- Résultats

- Moyennage de l'erreur quadratique moyenne sur les 20 tirages
- Détermination d'un ensemble unique de bandes hyperspectrales (« forward selection » des bandes les plus fréquentes)
- Calcul de la moyenne *a posteriori* de l'erreur quadratique moyenne





**1. Introduction**

**2. Matériau expérimental**

**3. Sélection des bandes par Réseaux de Neurones**



**4. Résultats**

**5. Conclusion**

## Test sur Angers des méthodes de sélection de la littérature

### ■ Méthode des indices

- 17 indices de la littérature testés dont certains sont des combinaisons de 2 ou 3 indices [2]
- Meilleur indice : combinaison de 2 indices  $R_{2300}$  et  $ND = [R_{1368} - R_{1722}] / [R_{1368} + R_{1722}]$
- **Erreur en prédiction : 22%**

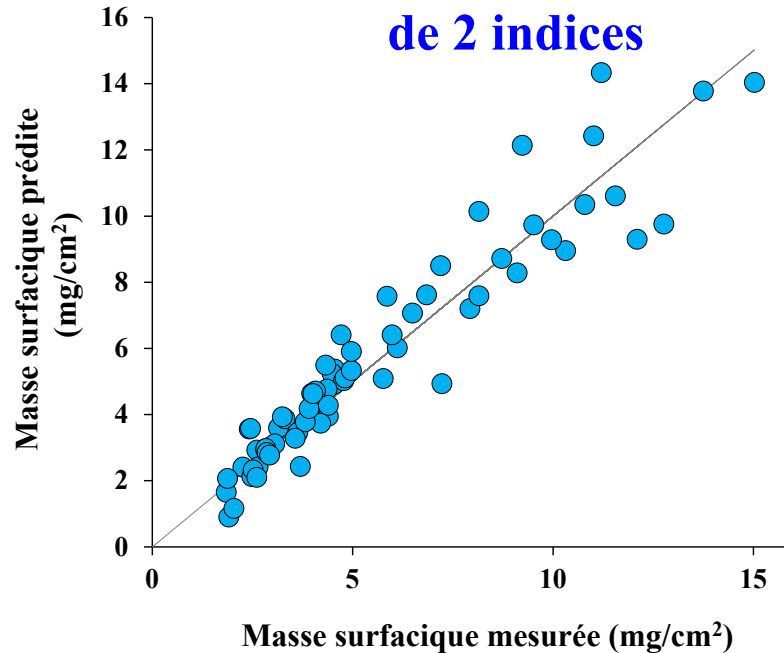
### ■ Régression linéaire par moindres carrés partiels

- Sélection de bandes par le paramètre Variable Importance in Projection (VIP)
- Entre 10 à 20 bandes selon les tirages
- **Erreur en prédiction : 28%**

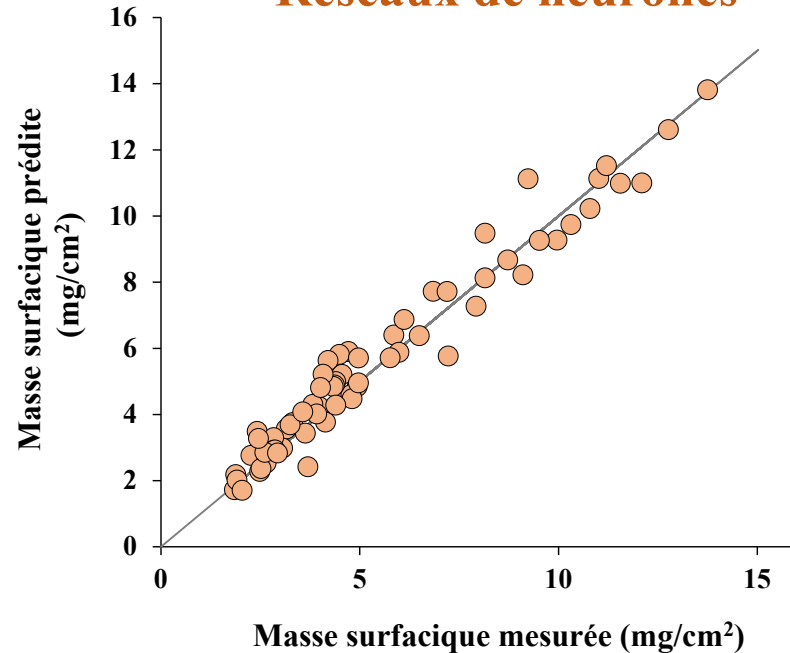
[2] Y. Chen et al., "Optimized Estimation of Leaf Mass per Area with a 3D Matrix of Vegetation Indices", *Remote Sensing*, **13**, pp. 3761, 2021.

## Prédiction de la masse surfacique : comparaison visuelle

**Meilleure combinaison  
de 2 indices**



**Réseaux de neurones**



## Prédiction de la masse surfacique : comparaison quantitative

Racine de l'erreur quadratique moyenne (RMSE) relative\* **obtenue en moyennant sur les 20 tirages :**

Modèle	Calibration (Entraînement et Validation)	Test	Nb. de bandes
Combinaison de 2 indices	18.6%	<u>22%</u>	3
Réseaux de neurones	14.8%	<u>17.5%</u>	<u>3-9</u>
Réseaux de neurones (a posteriori)	15.2%	18.1%	4

### Observations:

Avec les réseaux de neurones :

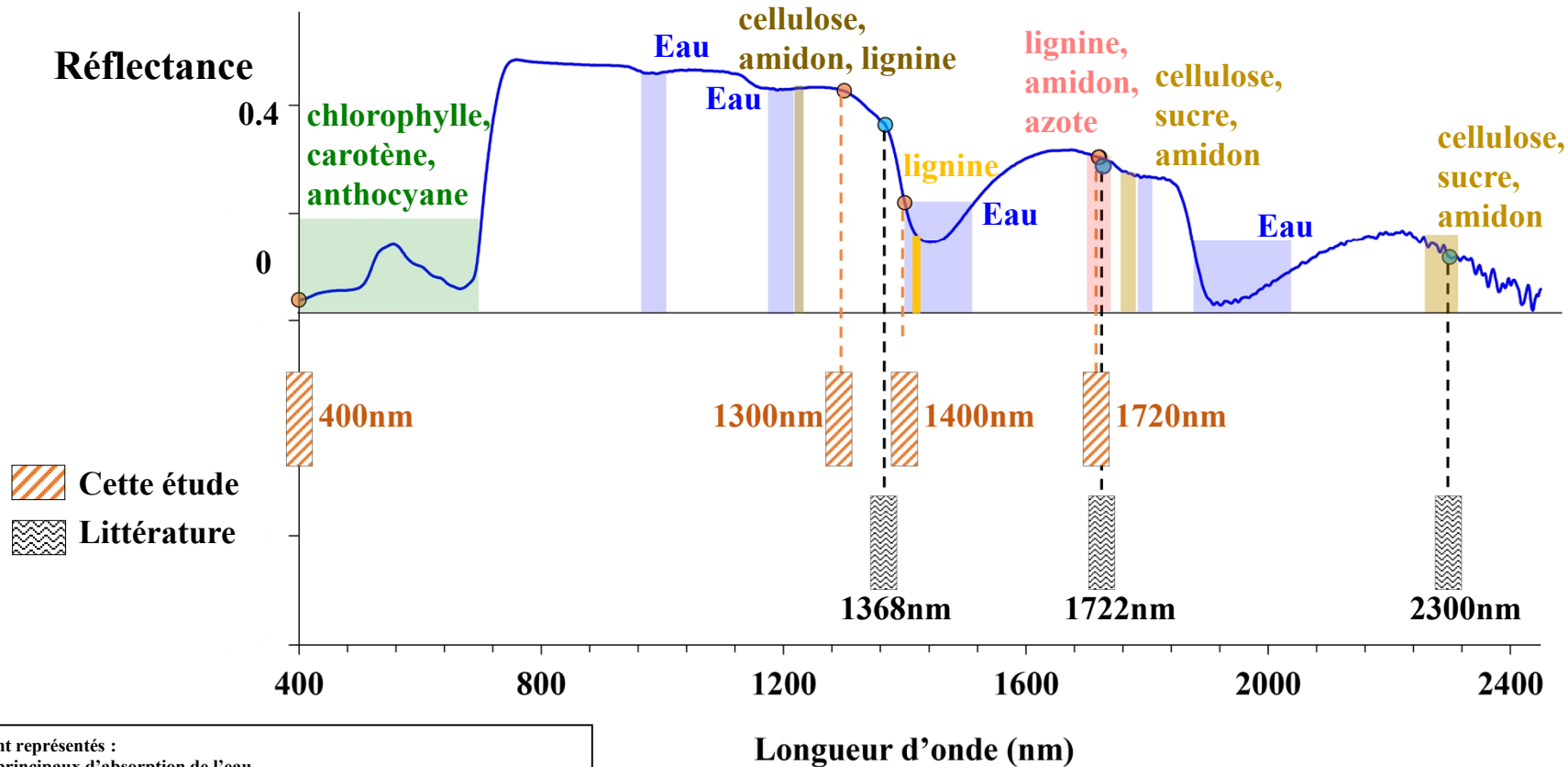
- amélioration de 4.5 points de l'erreur quadratique moyenne relative
- 3 à 9 bandes sélectionnées selon les tirages
- 4 bandes finalement retenues a posteriori

[C. Baskiotis, J. Aval, M. El Bouz, A. Al Falou, IGARSS 2022 (submitted)]

\* Racine de l'erreur quadratique Moyenne Relative =  $\frac{\text{Racine de l'erreur quadratique moyenne}}{\text{moyenne des masses surfaciques}}$

# Résultats

## Prédiction de la masse surfacique : bandes sélectionnées



\*Seuls sont représentés :  
- les pics principaux d'absorption de l'eau  
- les bandes d'absorption immédiatement voisines à des longueurs d'onde d'intérêt  
\*\*Bandes extraites de [S. Jacquemoud and S. Ustin, *Leaf Optical Properties*, Cambridge University Press, Cambridge, 2019].

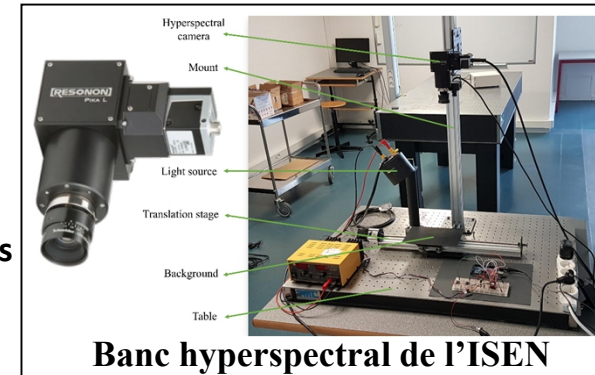
## Conclusions

Testés sur la base de données publiques *Angers*, les réseaux de neurones :

- permettent **un gain de 4.5 points dans la prédiction de la Masse Foliaire Surfaccie (MFS) [erreur relative : 17.5%]**
- tout en ne nécessitant qu'un **faible nombre de longueurs d'onde** (3 à 9)
- **4 bandes** finalement sélectionnées pour la MFS [400nm, 1300nm, 1400nm, 1720nm]

## Perspectives

- Amélioration supplémentaire de la précision de prédiction de la Masse Foliaire Surfaccie (via une nouvelle base de données expérimentale)
- Amélioration de la prédiction des autres paramètres physico-chimiques (Chlorophylle, Carotène, Anthocyane, ...)
- Proposition de formules analytiques simples remplaçant les réseaux de neurones
- Application à la détection précise du stress hydrique des feuilles de tomates



**Merci pour votre attention**

# Annexe

## Partial Least-Square Regression : principe

1) Analyse des corrélations linéaires entre :

- Réflectances ( $R_{\lambda_1}, R_{\lambda_2}, R_{\lambda_3}, \dots, R_{\lambda_n}$ )
- Trait bio-chimique Y



2) Fabrication d'un nouvel espace d'entrée ( $X_1, X_2 \dots X_n$ )

$X_i$  = Combinaison linéaire de  $R_{\lambda_1}, R_{\lambda_2}, R_{\lambda_3}, \dots, R_{\lambda_n}$  qui :

- maximisent\* la covariance entre  $X_i$  et le trait bio-chimique
- sont orthogonales entre elles

3) - Sélection d'un nombre réduit de composantes ( $X_1, X_2 \dots X_r$ ) par cross-validation

- Régression linéaire Y : fonction linéaire de  $X_1, X_2 \dots X_r$

4) Calcul de l'importance de chaque bande  
(Variable Importance in Projection, VIP)

5) Sélection du plus faible nombre de bandes hyperspectrales  
possible tout en gardant une bonne précision

$\{R_{\lambda_1}, R_{\lambda_j}, R_{\lambda_k}, R_{\lambda_l}\}$

