



Gestion des
ressources en eau
dans les milieux
ruraux tunisiens



Résumé du cours en Imagerie Vis-PIR pour les Sciences du Sol

Cécile Gomez

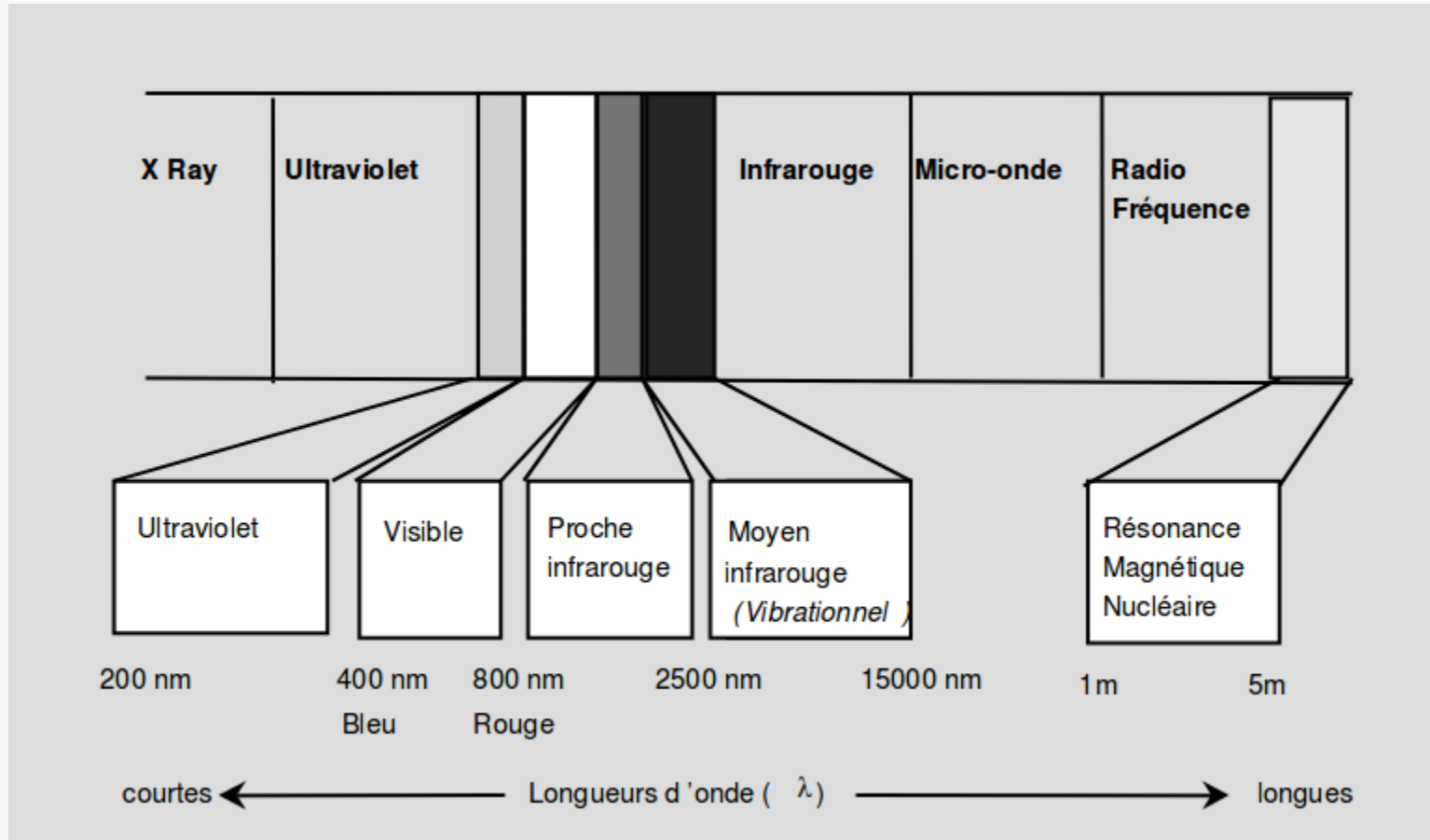
IRD

Laboratoire d'étude des Interactions Sol Agrosystème Hydrosystème (LISAH) Montpellier, France



La spectroscopie

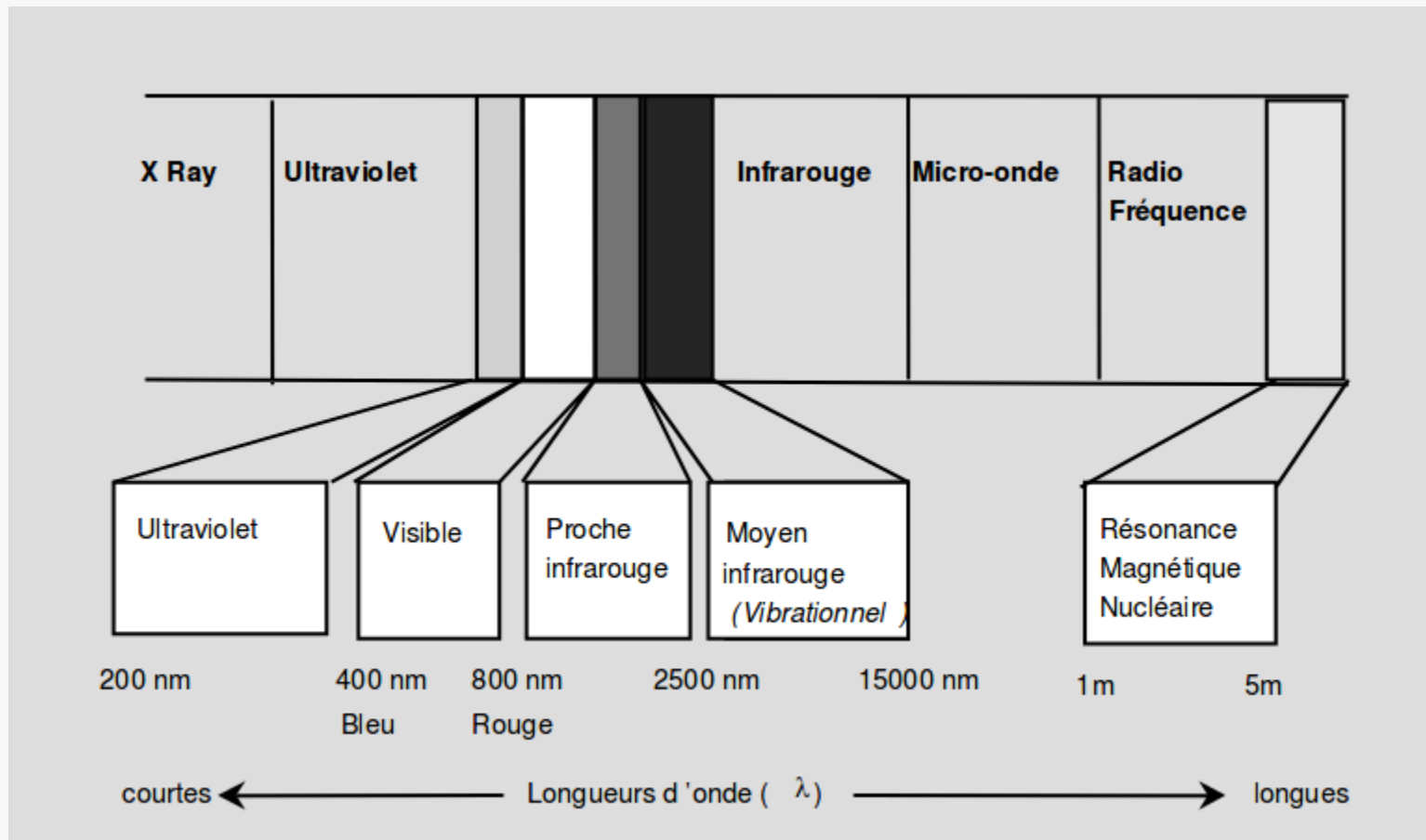
Visible Proche Infra Rouge (400 – 2500 nm)



Mesure de l'absorption de rayonnement électromagnétique par un échantillon

La spectroscopie

Visible Proche Infra Rouge (400 – 2500 nm)



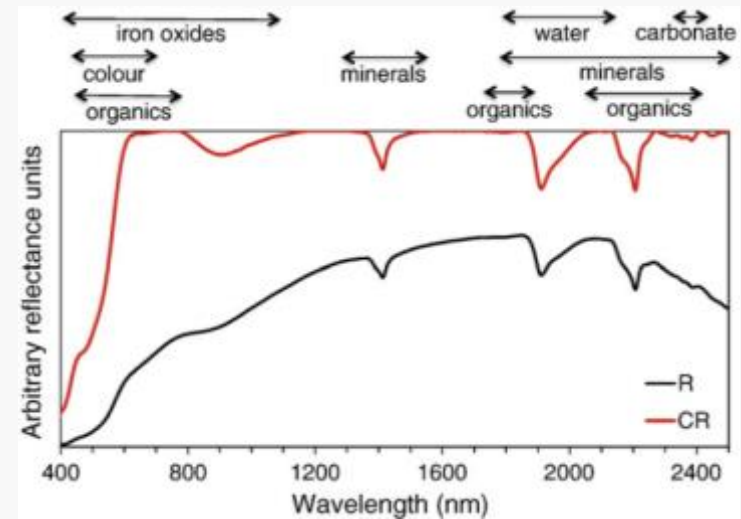
Il y a absorption de ce rayonnement électromagnétique quand la fréquence de la radiation qui vient frapper une liaison moléculaire est égale à la fréquence de vibration de cette liaison.

Exemple : Il y a absorption du rayonnement électromagnétique à 1900 nm quand la fréquence de la radiation vient frapper une liaison H-OH de la molécule d'eau.

La spectroscopie

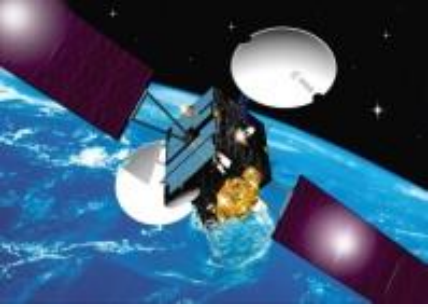
Visible Proche Infra Rouge (400 – 2500 nm)

- Enveloppe
- Bande d'absorption



En théorie :

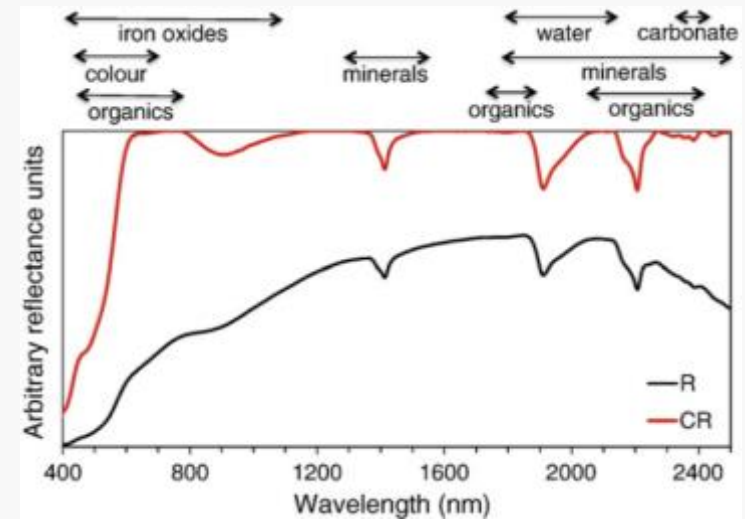
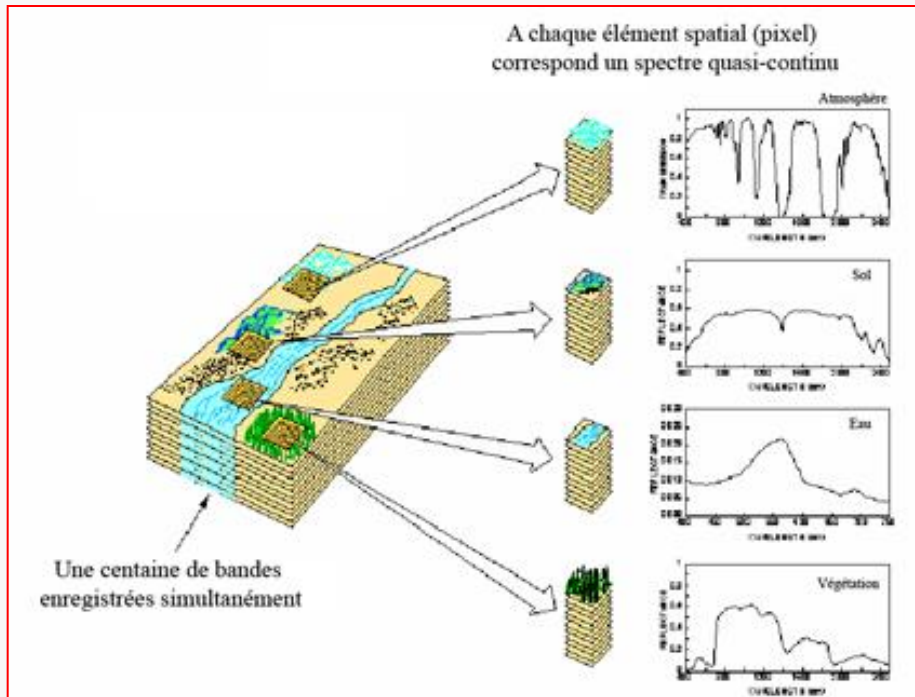
- chaque bande d'absorption est relié à un composé
- la profondeur du pic est relié à sa concentration (proportion)



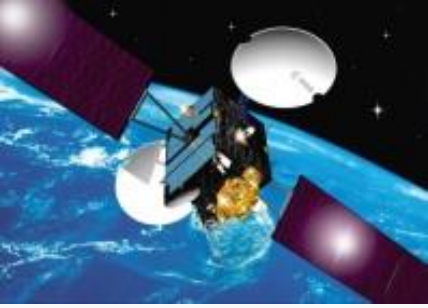
L'imagerie

Visible Proche Infra Rouge

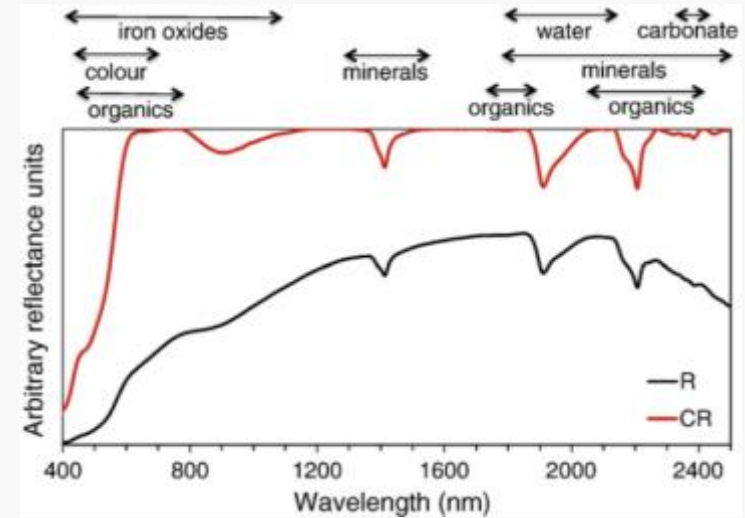
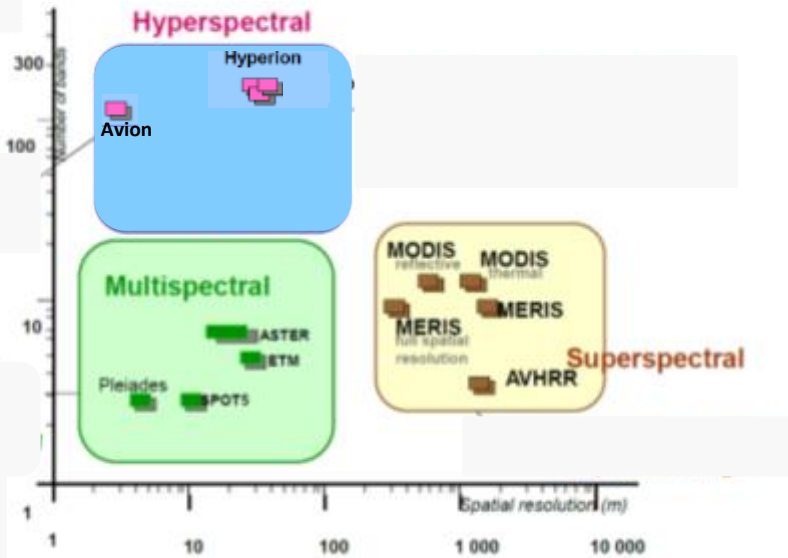
400 – 2500 nm



- Vision synoptique
- Permet d'observer des terrains difficiles d'accès
- Capteur embarqué sur différentes plateformes (drone, avion, satellite)

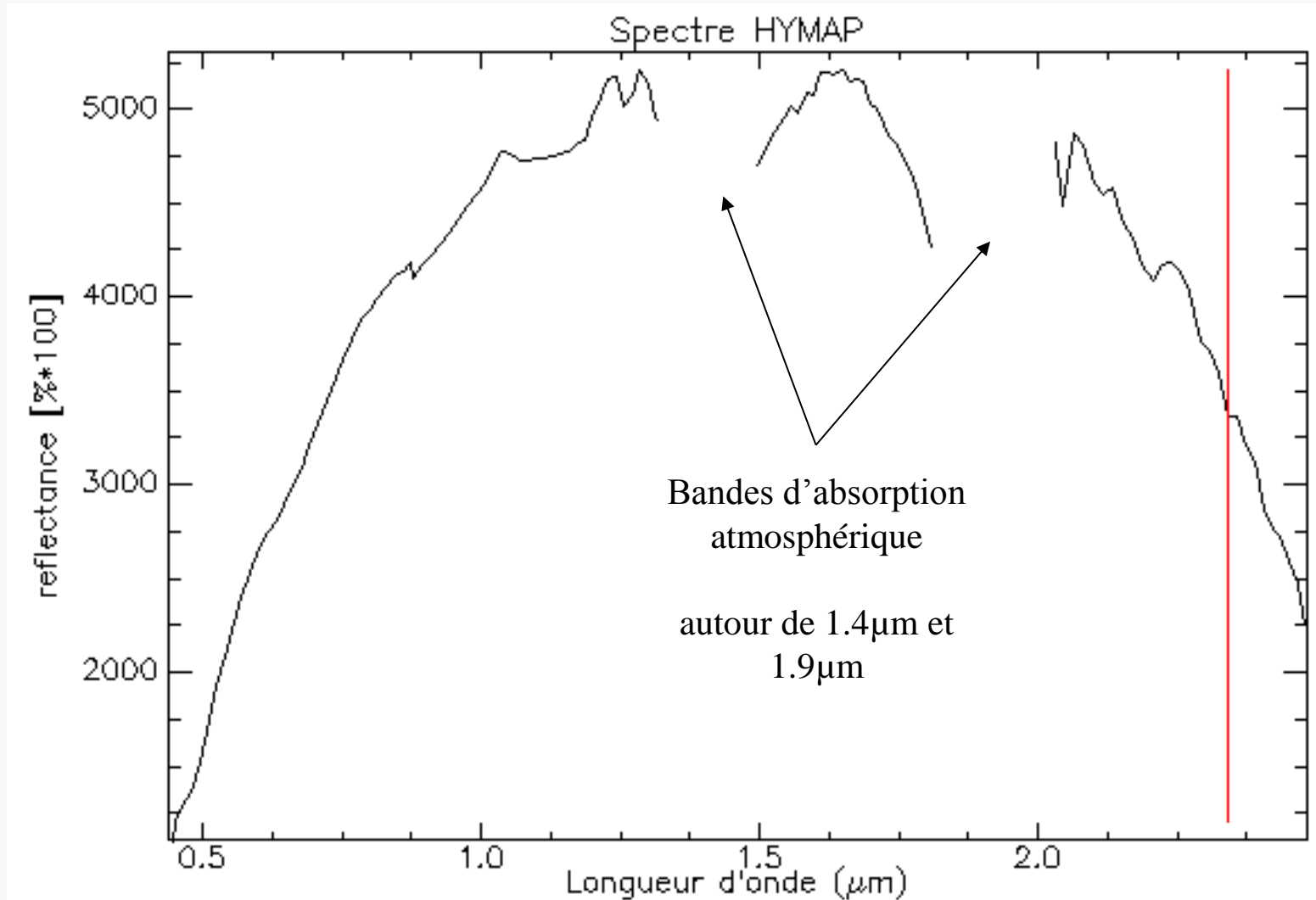


L'imagerie Visible Proche Infra Rouge 400 – 2500 nm



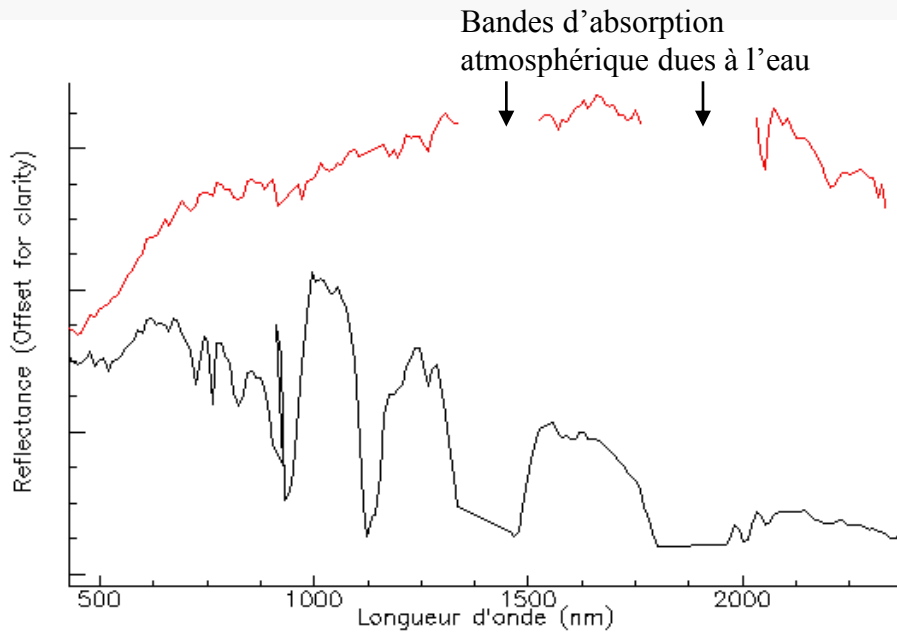
- Vision synoptique
- Permet d'observer des terrains difficiles d'accès
- Capteur embarqué sur différentes plateformes (drone, avion, satellite)

Pourquoi dit-on spectre « quasi-continu » ?

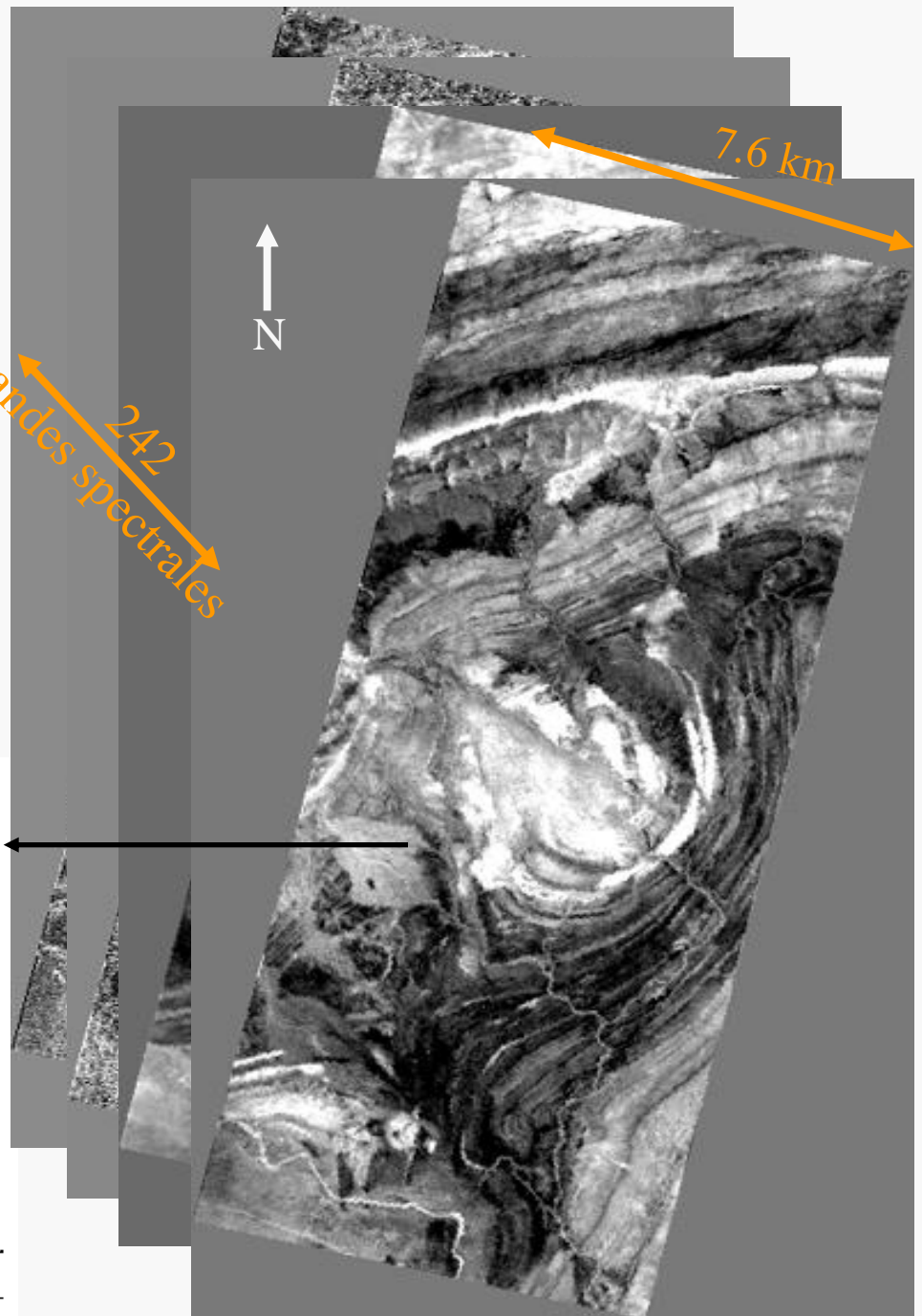


Données satellitaires HYPERION

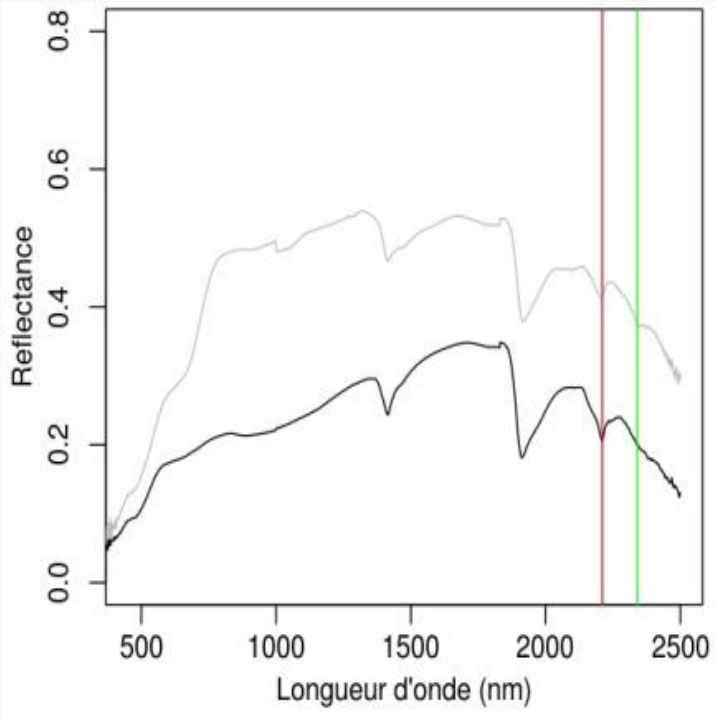
- Résolution spectrale : 10nm
- Domaine spectral : 0,4 à 2,5 μm
- Résolution spatiale : 30m
- Rapport Signal sur Bruit $\sim 50:1$



242
bandes spectrales



Le mélange spectral



$$X_{\text{sol}} = f(X_{\text{composé}_1}, X_{\text{composé}_2}, \dots, X_{\text{composé}_n})$$

avec f supposée linéaire

$$X_{\text{sol}} = f(X_{\text{Argile}}, X_{\text{fer}}, \dots, X_{\text{CaCO}_3})$$

Mais :

- les spectres de composé pur tel que l'argile granulométrique ou le carbone organique ($X_{\text{composé}}$) n'existent pas.
- les interactions entre composés déforment les spectres et les bandes d'absorption
- Influences physiques externes interviennent (lumières, température, ...)

Règle de prédiction

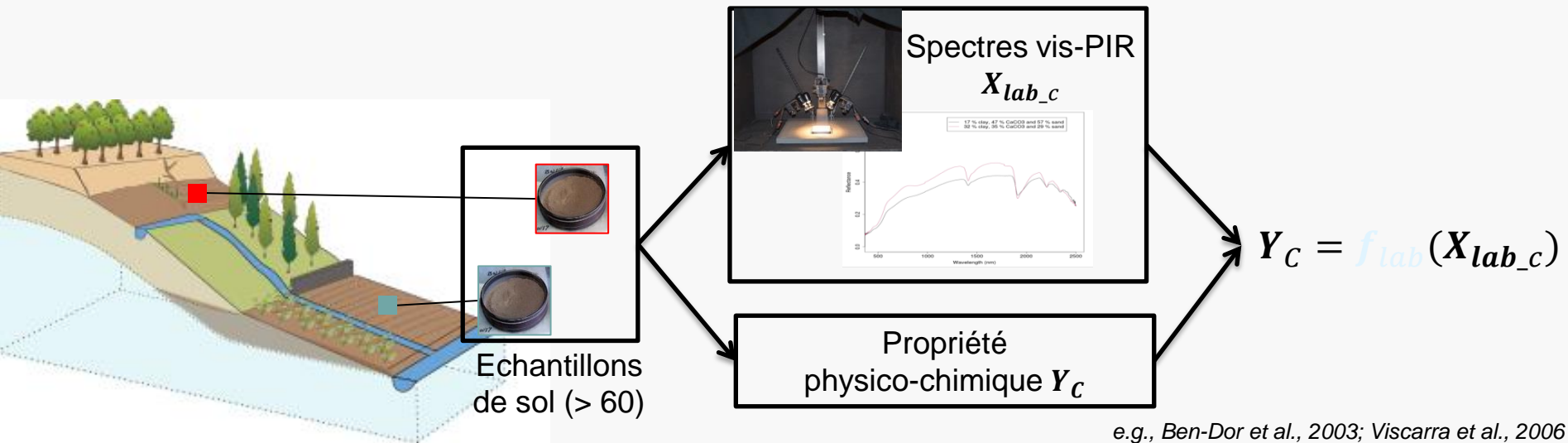
D'après Ben-Dor et al., 2002 :

Une propriété de sol Y_C est prédictible par imagerie Vis-PIR si :

Elle a une spécificité chimique

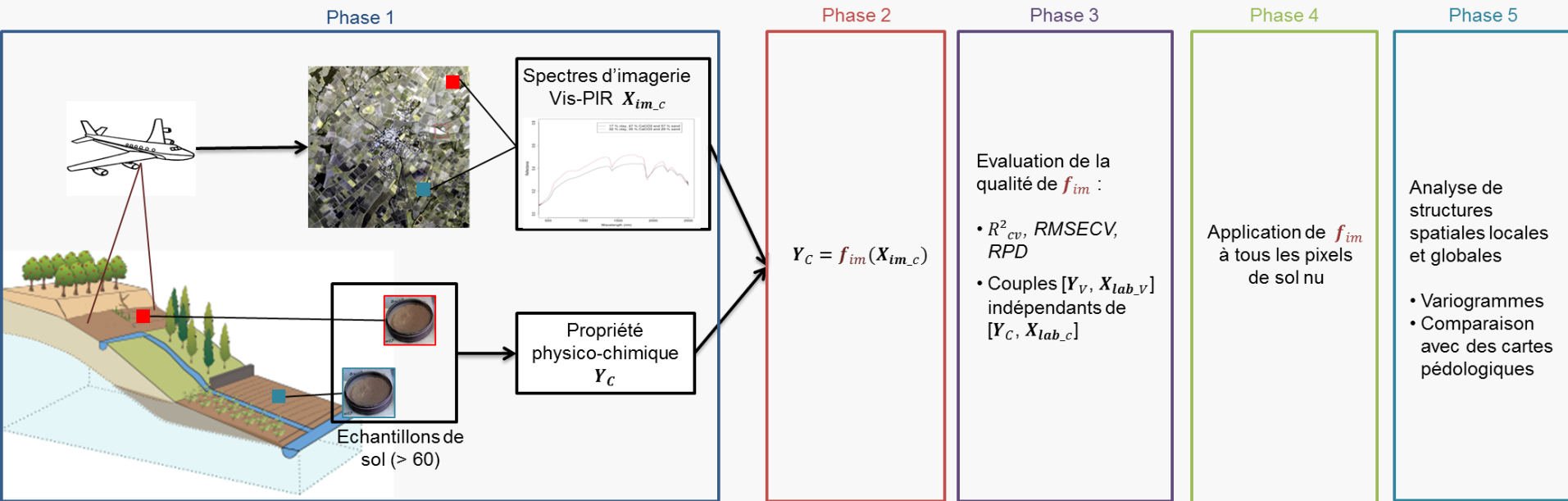
OU

Elle est corrélée à une propriété ayant une spécificité chimique



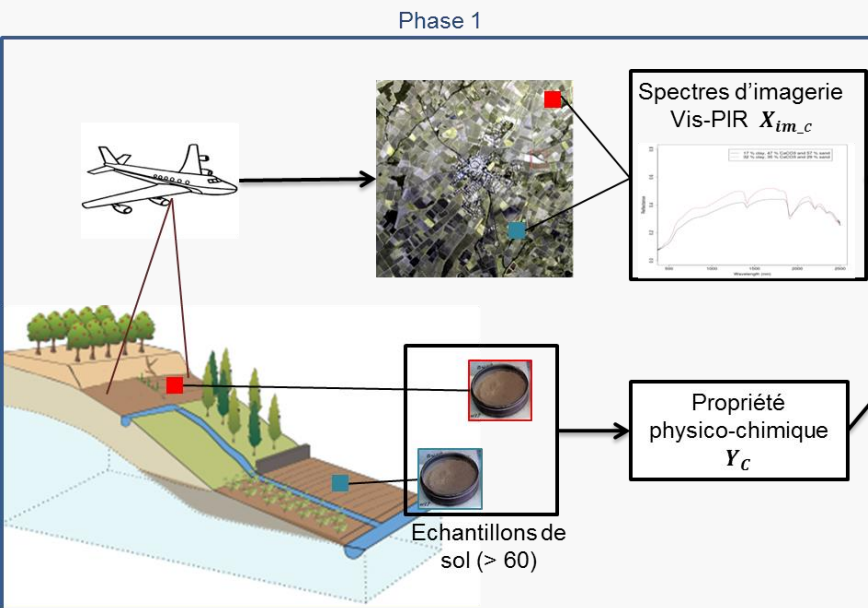
e.g., Ben-Dor et al., 2003; Viscarra et al., 2006

Schéma général



Phase 1 : Préparation des données

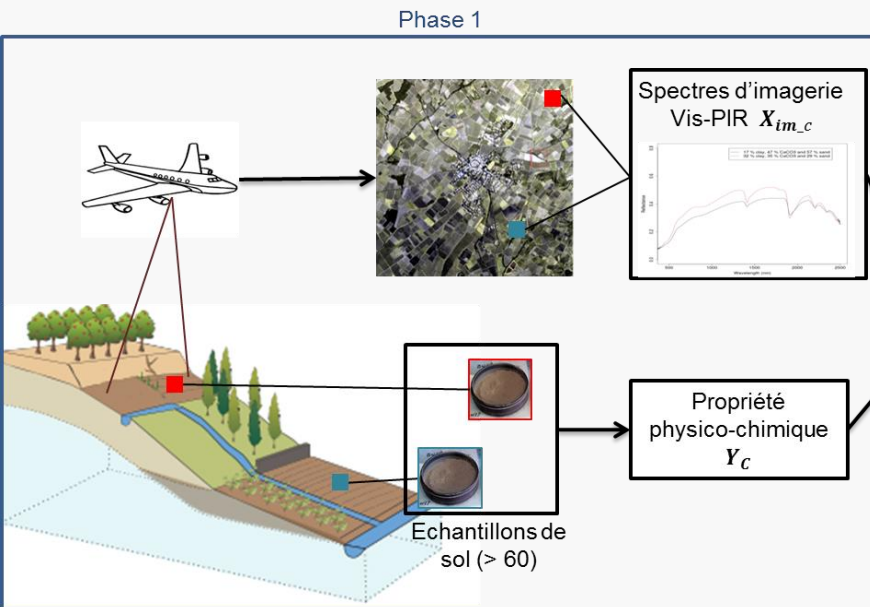
1. *Préparation* de la Base de Données de Calibration (BD_Calib)



- Minimum $N = 60$ couples [Y_c , X_{im_c}]
- Echantillons collectés sur des pixels de sol nu
- Echantillons représentatifs de la pédologie de la zone d'étude

Phase 1 : Préparation des données

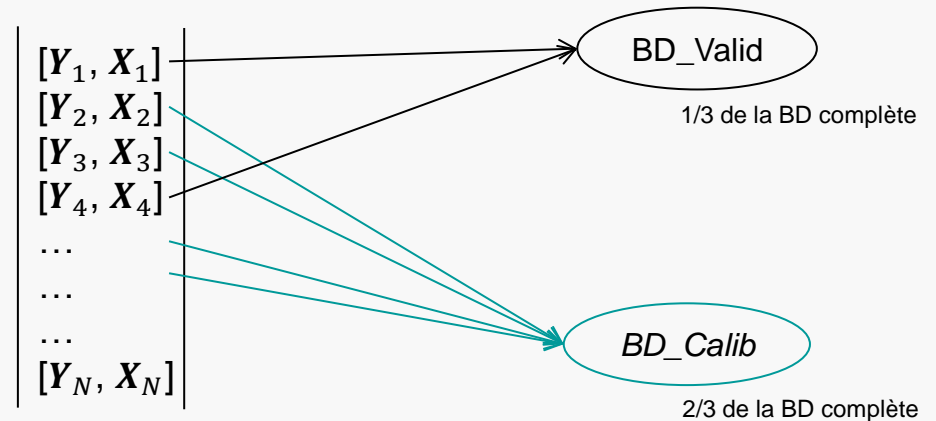
1. *Préparation* de la Base de Données de Calibration (BD_Calib) et la Base de Données de Validation (BD_Valid).



Si $N > 80$: Division de la base de données (BD) en 2 lots :

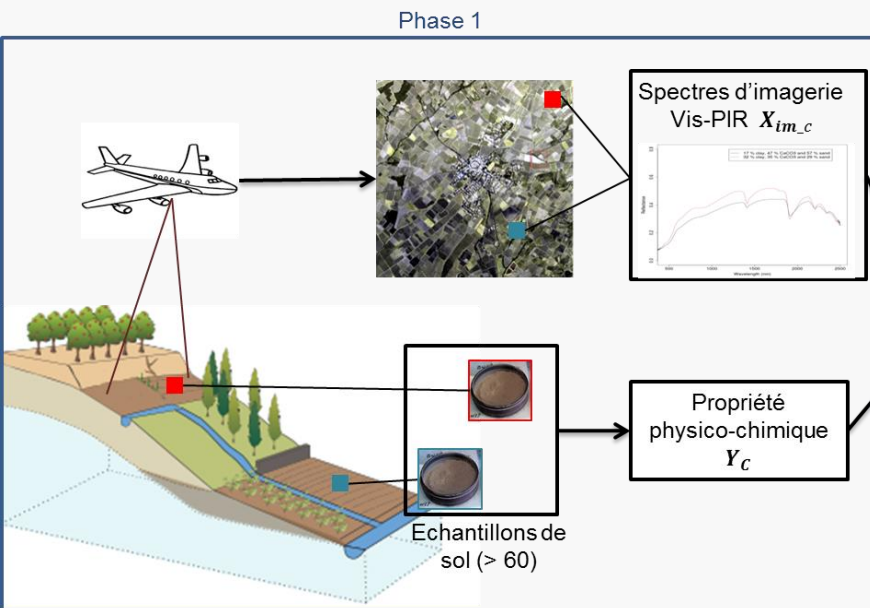
- BD de calibration (servira à la construction du modèle [Y_c, X_{im_c}])
- BD de validation (servira à la validation du modèle [Y_v, X_{im_v}])

La BD_Valid est indépendante de la BD_Calib



Phase 1 : Préparation des données

- Préparation** de la Base de Données de Calibration (BD_Calib) et la Base de Données de Validation (BD_Valid).



Corrélation inter-propriété

R ²	Argile	CaCO ₃	pH
Argile	1		
CaCO ₃	0.6	1	
pH	0.2	0.5	1

Analyse des bandes d'absorption

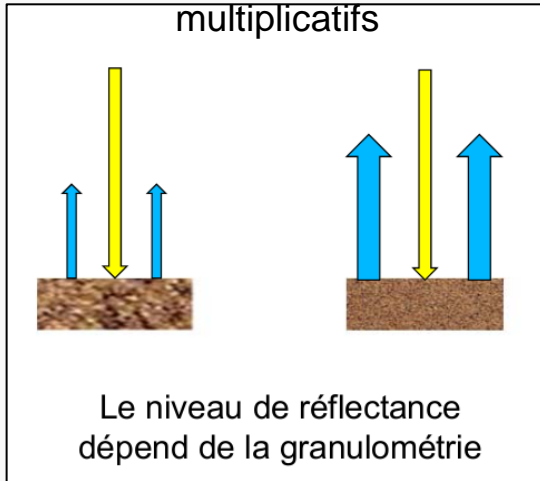
	Argile	CaCO ₃	Fer
Bande absorption	Oui	Oui	Non

Phase 1 : Préparation des données

1. *Préparation* de la Base de Données de Calibration (BD_Calib) et la Base de Données de Validation (BD_Valid).

2. *Prétraitements* des spectres

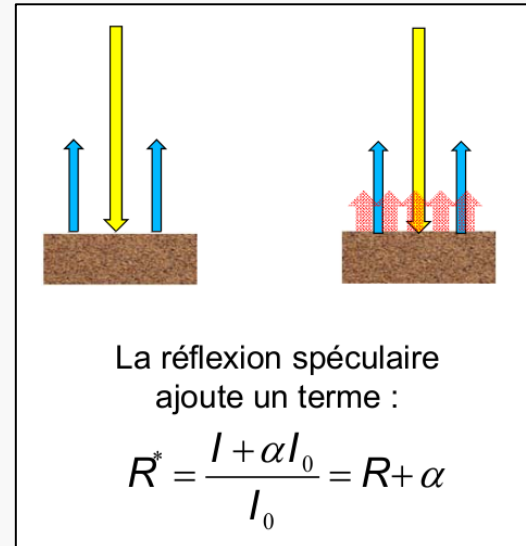
Dédiés aux effets
multiplicatifs



Le niveau de réflectance
dépend de la granulométrie

- **Division du spectre** par une quantité affectée à l'effet multiplicatif mais non informative : *somme ou écart type*
- **Transformation** de l'effet multiplicatif en effet additif :
Logarithme (1/R)

Dédiés aux effets additifs



La réflexion spéculaire
ajoute un terme :

$$R^* = \frac{I + \alpha I_0}{I_0} = R + \alpha$$

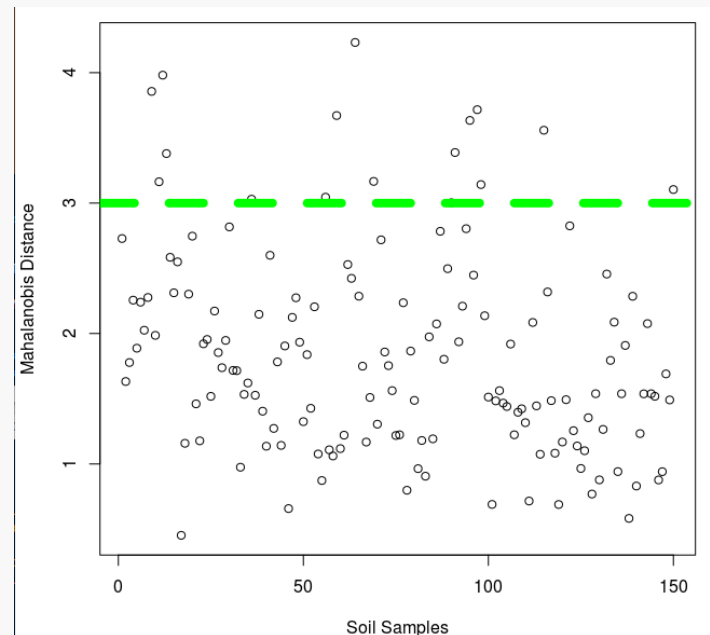
- **Soustraction** d'une valeur affectée par l'effet additif : *somme ou moyenne*
- **Filtrage** du signal : *dérivation du spectre*

Standard normal Variate :

Soustraction de la moyenne + **Division** du spectre l'écart type

Phase 1 : Préparation des données

1. Préparation de la Base de Données de Calibration (BD_Calib) et la Base de Données de Validation (BD_Valid).
2. Prétraitements des spectres
3. Suppression d'**outliers** (uniquement dans BD_Calib)
Distance de Mahalanobis (écart à la moyenne, en prenant en compte les corrélations entre variables)



Phase 1 : Préparation des données

1. Préparation de la Base de Données de Calibration (BD_Calib) et la Base de Données de Validation (BD_Valid).
2. Prétraitements des spectres
3. Suppression d'outliers (uniquement dans BD_Calib)
4. **Identification** d'outliers dans BD_Valid (ne pas enlever ou vous surestimerez les performances de vos modèles)

Phase 2 : Construction du modèle



- Multiple Linear Regression
- Partial Linear Square Regression
- Leave-one-out Cross Validation :

$$Y_C = f_{im}(X_{im_c}) = b_0 + b_1 X_{\lambda_1} + \dots + b_m X_{\lambda_m}$$

On apprend sur (n-1) observations puis on valide le modèle sur la énième observation et l'on répète cette opération n fois

Pourquoi la PLSR en Spectroscopie Vis-PIR ?

En laboratoire : Nombre de Longueur d'onde >> nombre d'échantillon

Permet de gérer la colinéarité entre longueur d'onde.

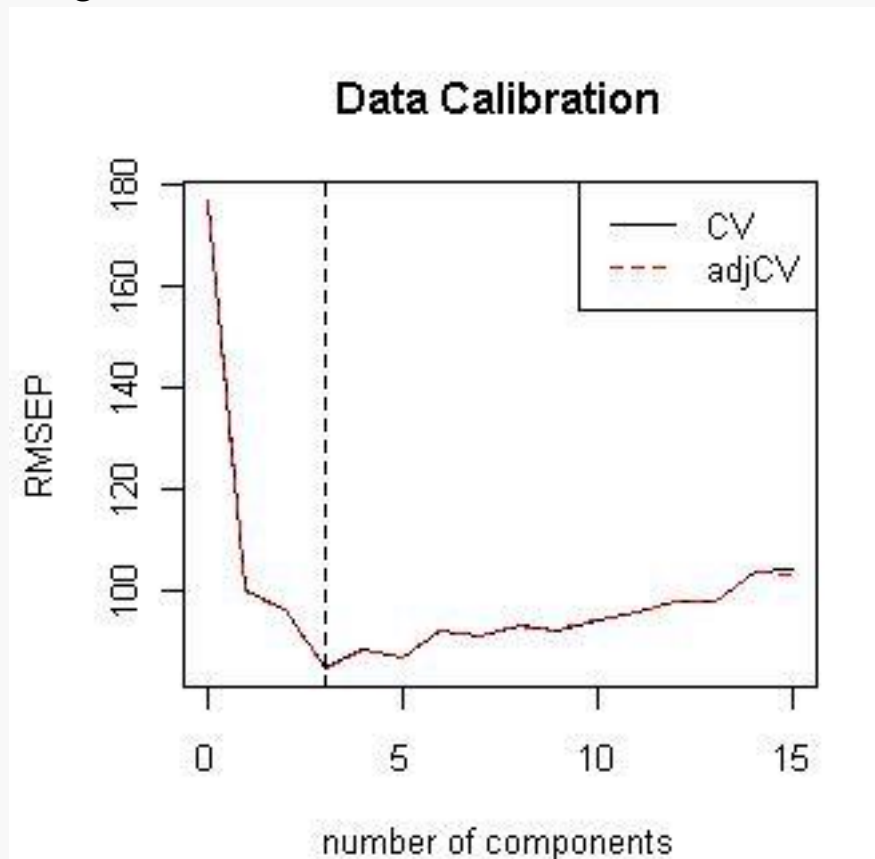
Phase 2 : Construction du modèle

Cas de la PLSR

Choix du nombre de variables latentes p .

$RMSEP$ ↘ tant qu'on améliore le modèle puis $RMSEP$ ↗

Si p = nombre de longueurs d'onde => MLR



Phase 3 : Evaluation du modèle

Sur la BD_Calib

- $R^2_{cal} = 1 - \frac{SSR_{cal}}{SST_{cal}} = 1 - \frac{\sum_{i=1}^n (y_{cal_i} - \hat{y}_{cal_i})^2}{\sum_{i=1}^n (y_{cal_i} - \bar{y}_{cal})^2}$
- Standard Error of Calibration (SEC) ou Root Mean Square Error of Calibration (RMSEC)

$$SEC = RMSEC = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_{cal_i} - y_{cal_i})^2}{n}} = RMSEP \text{ non corrigé du biais}$$

- $Biais = \sum_{i=1}^n \frac{\hat{y}_{cal_i}}{n} - \sum_{i=1}^n \frac{y_{cal_i}}{n}$

- Ratio of performance to deviation (RPD)

$$RPD = \frac{sd(y_{cal})}{RMSEP}$$

RPD < 1.4 : modèle « non performant »

1.4 < RPD < 2 : modèle « correct »

2 < RPD : modèle « bon ».

- Ratio of performance to interquartile distance (RPIQ)

$$RPIQ = \frac{IQ}{RMSEC} = \frac{Q3 - Q1}{RMSEC}$$

Phase 3 : Evaluation du modèle

Sur la BD_Valid

- $R^2_{val} = 1 - \frac{SSR_{val}}{SST_{val}} = 1 - \frac{\sum_{i=1}^n (y_{val_i} - \hat{y}_{val_i})^2}{\sum_{i=1}^n (y_{val_i} - \bar{y}_{val})^2}$

- Standard Error of Prediction (SEP) ou Root Mean Square Error of Prediction (RMSEP)

$$RMSEP = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_{val_i} - y_{val_i})^2}{n}}$$

- $Biais = \sum_{i=1}^n \frac{\hat{y}_{val_i}}{n} - \sum_{i=1}^n \frac{y_{val_i}}{n}$

- Ratio of performance to deviation (RPD)

$$RPD = \frac{sd(y_{val})}{RMSEP}$$

RPD < 1.4 : modèle « non performant »

1.4 < RPD < 2 : modèle « correct »

2 < RPD : modèle « bon ».

- Ratio of performance to interquartile distance (RPIQ)

$$RPIQ = \frac{IQ}{RMSEP} = \frac{Q3 - Q1}{RMSEP}$$

Phase 3 : Cohérence du modèle

- **Variable Importance in the Projection (VIP)** (Wold et al., 2001, Chong & Jun, 2005)

$$VIP_i = N \cdot \sum_{p=1}^{p_{opt}} w_{p,i}^2 \cdot R_p^2$$

Où $w_{p,n}^2$ est le poids du loading pour la p^{th} variable latente à la longueur d'onde i , p_{opt} est le nombre optimal de variables latentes, R_p^2 est le coefficient de détermination obtenu pour la p^{th} variable latente et N est le nombre total de longueur d'onde.

- **b-coefficients** (Haaland & Thomas, 1988)

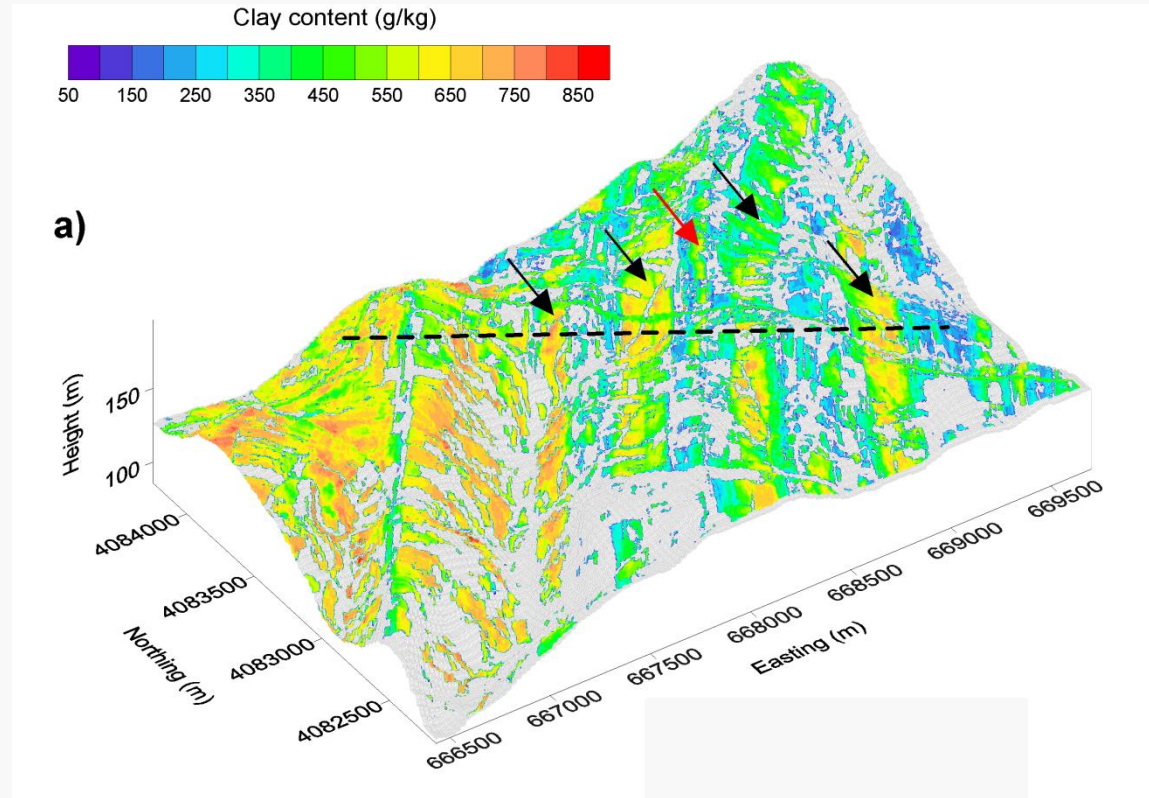
λ_i est considérée comme importante si :

$$b - \text{coefficient}_{\lambda_i} > sd(b - \text{coefficient})$$

$$VIP_i > 1$$

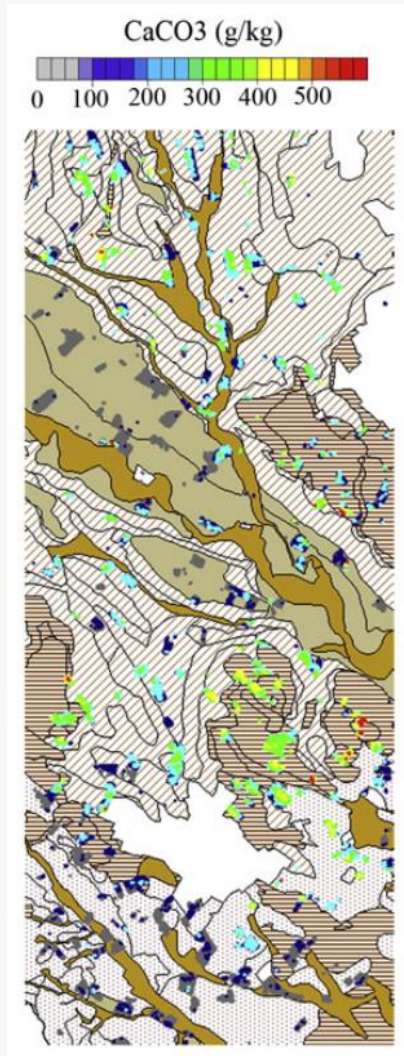
Phase 5 : Analyse des cartes

Analyse Visuelle









Phase 5 : Analyse des cartes

Analyse vis-à-vis d'une carte pédologique



Soilscapes:

-  S1
 -  S2
 -  S3
 -  S4
 -  S5
 -  Urban
- Sols calcaires et sédimentaires datant du Miocène
- Sols ayant subi un lessivage de carbonate du à des processus pédogénétique

Phase 5 : Analyse des cartes

Variogrammes

Idée fondamentale:

La nature n'est pas entièrement "imprévisible". Deux observations situées l'une près de l'autre devraient, en moyenne, se ressembler davantage que deux observations éloignées.

Variogramme : Décrit la continuité spatiale et la régularité du phénomène. Illustre la dépendance spatiale entre 2 points espacés d'une distance h .

$$\gamma(h) = \frac{1}{2n} \sum_{i=1}^n \{p(x_i) - p(x_i + h)\}^2$$

Où

n : nombre de point d'échantillonnage

$p(x_i)$: valeur de l'échantillon à la position x_i

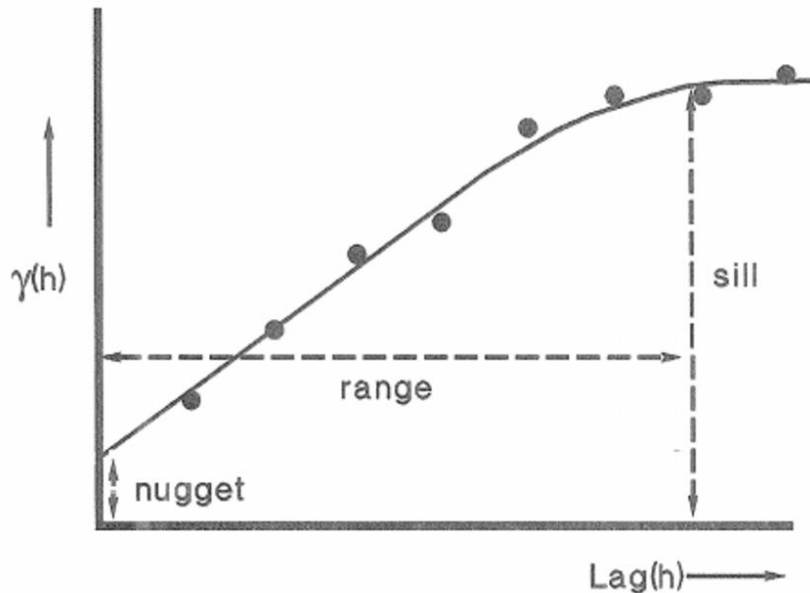
Phase 5 : Analyse des cartes

Variogrammes

Idée fondamentale:

La nature n'est pas entièrement "imprévisible". Deux observations situées l'une près de l'autre devraient, en moyenne, se ressembler davantage que deux observations éloignées.

Variogramme : Décrit la continuité spatiale et la régularité du phénomène
Illustre la dépendance spatiale entre 2 points espacés d'une distance h .



Nugget (=Effet de pépité C_0): Variation à très courte échelle, erreurs de localisation, erreurs d'analyse et précision analytique.

Range (=Portée a): Distance où deux observations ne se ressemblent plus du tout en moyenne, elles ne sont plus liées (covariance nulle) linéairement. À cette distance, la valeur du variogramme correspond à la variance de la variable aléatoire.

Sill (Palier σ^2) = $C_0 + C$: Variance de la v.a. ($\text{Var}(Z(x))$)



Gestion des
ressources en eau
dans les milieux
ruraux tunisiens



Merci au CNCT et à la DG/ACTA

Merci au soutien financier reçu du :

- ✓ **LMI NAILA « gestion des ressources en eau dans les milieux ruraux »**
- ✓ **Projet ANR ALMIRA « Adaptation des mosaïques paysagères dans les agrosystèmes pluviaux Méditerranéens pour une gestion durable de la production agricole, des ressources en eau et en sol »**

