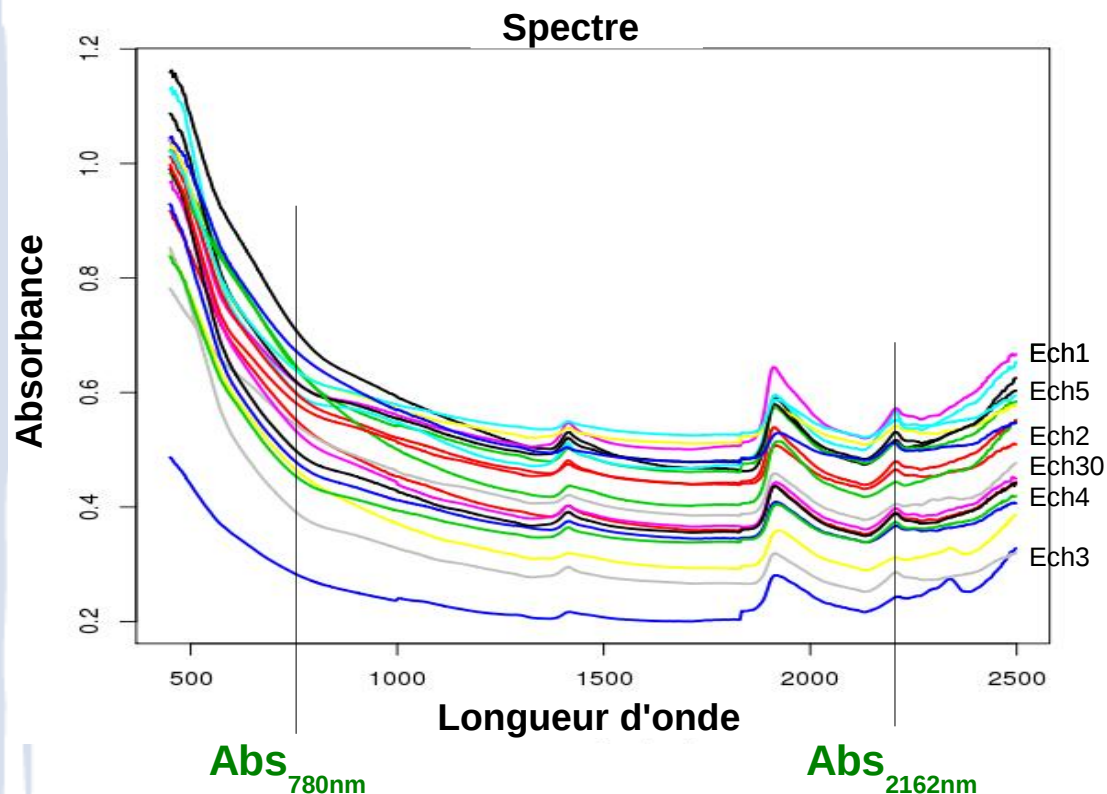


Locally-Weighted Partial Least Squares Regression for Infra-red spectra analysis

- Spectrométrie infrarouge : intérêt en sciences du sol
 - Spectre IR: mesure rapide et peu couteuse
 - Mesure de référence lente et couteuse



Modèle




	Carbone Organique
Ech 1	12.72
Ech 2	8.52
Ech 3	26.81
Ech 4	62.31
Ech 5	21.57
...	...
...	...
...	...
...	...
Ech 30	36.26

Spectrométrie Infra-rouge

- Objectif

- Exprimer un constituant du sol en fonction de l'absorbance de l'échantillon de sol à chaque longueur d'onde

	Abs 1100 nm	Abs 1102 nm	Abs 1104 nm	Abs 1106 nm	...	Abs 2496 nm	Abs 2498 nm	
Ech 1	0.3034	0.3031	0.3027	0.3024	...	0.3171	0.3174	Corg
Ech2	0.2856	0.2852	0.2848	0.2843	...	0.2941	0.2943	12.72
Ech3	0.3362	0.3358	0.3354	0.3349	...	0.3293	0.3297	8.52
Ech4	0.3375	0.3372	0.3369	0.3366	...	0.3458	0.3462	26.81
Ech5	0.4074	0.4071	0.4068	0.4065	...	0.4053	0.4056	62.31
...	:::	21.57
...
...
...
Ech30	0.4878	0.4878	0.4877	0.4877	...	0.5380	0.5382	36.26



$$\text{Carbone organique} = a_{1100} \text{Abs}_{1100\text{nm}} + a_{1102} \text{Abs}_{1102\text{nm}} + \dots + a_{2498} \text{Abs}_{2498\text{nm}} + a_0$$

=> déterminer $a_{1100}, a_{1102}, \dots, a_{2498}, a_0$

- Outils chimiométriques

- Régression des moindres carrés partiels (PLS)

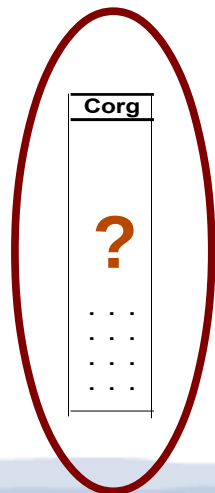
=> Méthode de construction de la matrice de composantes latentes

Régression PLS - Données spectrales

	Abs 1100 nm	Abs 1102 nm	Abs 1104 nm	Abs 1106 nm	...	Abs 2496 nm	Abs 2498 nm	Corg
MPL1	0.3034	0.3031	0.3027	0.3024	...	0.3171	0.3174	12.72
MPL2	0.2856	0.2852	0.2848	0.2843	...	0.2941	0.2943	8.52
MPL3	0.3362	0.3358	0.3354	0.3349	...	0.3293	0.3297	26.81
MPL4	0.3375	0.3372	0.3369	0.3366	...	0.3458	0.3462	17.21
MPL5	0.4074	0.4071	0.4068	0.4065	...	0.4053	0.4056	31.65
MPL6	0.3334	0.3330	0.3325	0.3321	...	0.3271	0.3275	21.57
...
...
...
...
...
MPL94	0.4878	0.4878	0.4877	0.4877	...	0.5380	0.5382	46.81
MPL95	0.3215	0.3211	0.3207	0.3204	...	0.3248	0.3252	17.26
MPL96	0.3866	0.3863	0.3860	0.3857	...	0.3835	0.3838	11.21
MPL97	0.3076	0.3073	0.3070	0.3067	...	0.3264	0.3269	10.86
MPL98	0.2839	0.2834	0.2830	0.2826	...	0.2986	0.2986	15.23
MPL99	0.3393	0.3394	0.3394	0.3395	...	0.3681	0.3685	84.46
MPL100	0.3540	0.3537	0.3534	0.3530	...	0.3870	0.3874	62.31

Echantillons de calibration

MODELE PLS



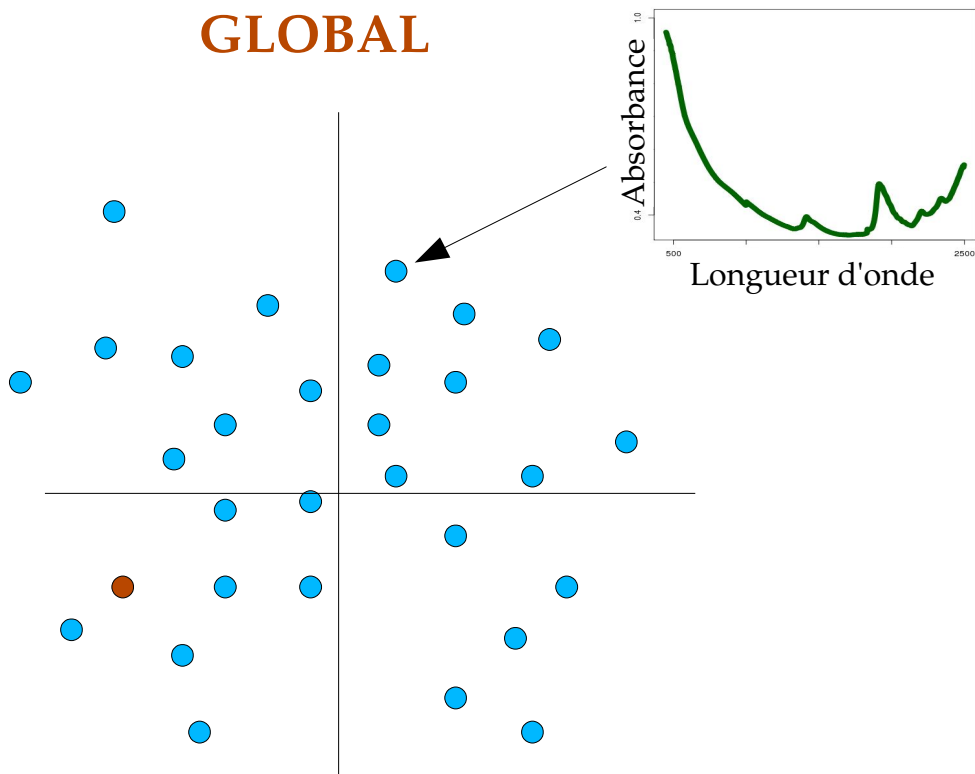
	Abs 1100 nm	Abs 1102 nm	Abs 1104 nm	Abs 1106 nm	...	Abs 2494 nm	Abs 2496 nm	Abs 2498 nm
ech1	0.3034	0.3031	0.3027	0.3024	...	0.3168	0.3171	0.3174
ech2	0.3362	0.3358	0.3354	0.3349	...	0.3289	0.3293	0.3297
ech3	0.3375	0.3372	0.3369	0.3366	...	0.3455	0.3458	0.3462
ech4	0.4074	0.4071	0.4068	0.4065	...	0.4049	0.4053	0.4056
ech5	0.3334	0.3330	0.3325	0.3321	...	0.3268	0.3271	0.3275
...
...
...
...
ech100	0.3540	0.3537	0.3534	0.3530	...	0.3866	0.3870	0.3874

Echantillons à prédire

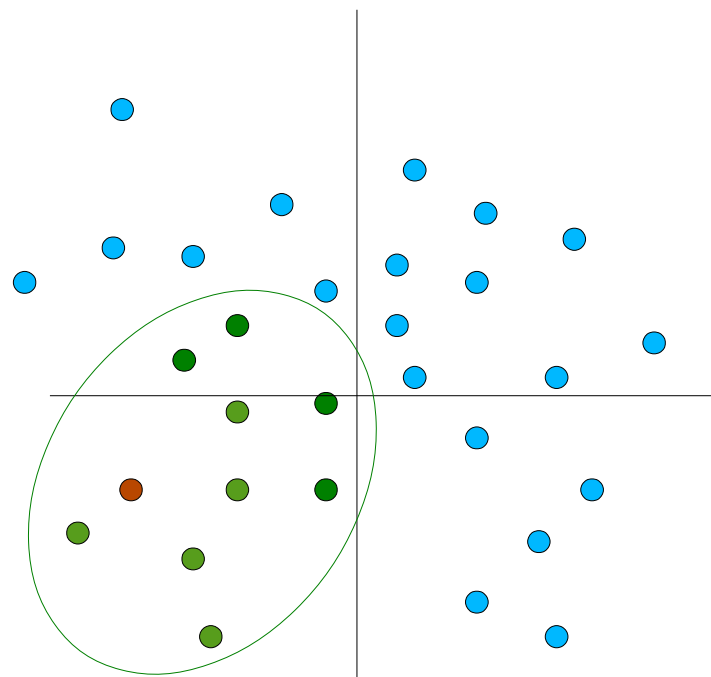


Régression PLS

GLOBAL



LOCAL



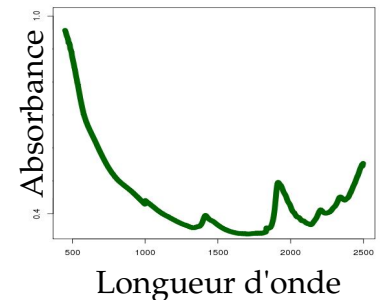
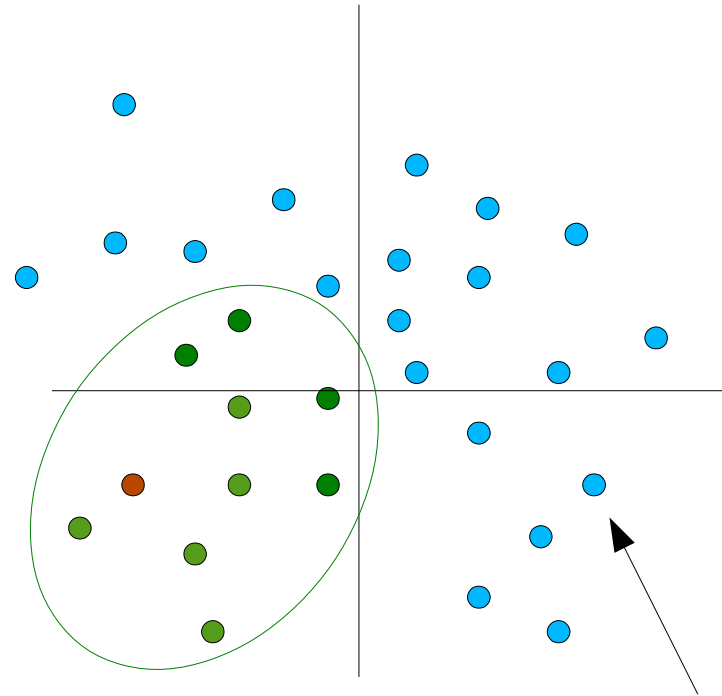
Tous les échantillons de calibration sont utilisés

Sélection des échantillons de calibration sur la base de la ressemblance spectrale

Construction du modèle : calibration locale

- Calibration locale
- - Distance – `localCal ()`
 - R^2 ou H
 - Distance limite `dlim`
ou nombre de voisins `Nb.neigh`
 - Nombre minimum de voisins `Nmin`
- - Cluster : Méthode bayésienne – `DPcal ()`
 - Clustering basé sur Dirichlet Process - `DP ()`
 - Voisins = cluster

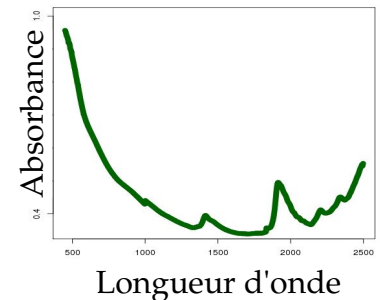
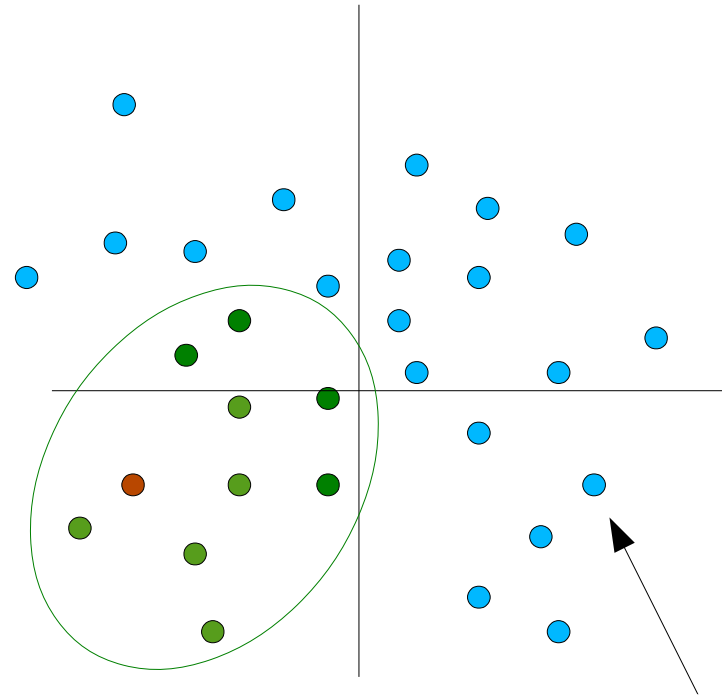
LOCAL



Construction du modèle : calibration locale

- Calibration locale
- - Distance – **localCal** ()
 - R^2 ou H
 - Distance limite **dlim**
ou nombre de voisins **Nb.neigh**
 - Nombre minimum de voisins **Nmin**
- - Cluster : Méthode bayésienne – **DPcal** ()
 - Clustering basé sur Dirichlet Process - **DP** ()
 - Voisins = cluster

LOCAL



PLS locale : Prédiction

- PLS locale - `lwPLS()`

- 1 modèle par échantillon z à prédire

- Echantillons de calibration N_{spec} et valeurs de références associées N_{refcal}

-

PONDERATION ?



NON

Tous les échantillons de calibration ont le même poids

method = "LocalPLS"
mvr() - package pls

OUI

Poids lié à la distance à l'échantillon cible

=> **les échantillons les plus proches spectralement ont plus de poids**

method = "LW-PLS"
lm_pls() - package plsRglm

PLS locale : Prédiction

- PLS locale - `lwPLS()`

- 1 modèle par échantillon z à prédire

- Echantillons de calibration N_{spec} et valeurs de références associées N_{refcal}

-

PONDERATION ?



NON

Tous les échantillons de calibration ont le même poids

method = "LocalPLS"
mvr() - package pls

OUI

Poids lié à la distance à l'échantillon cible
=> **les échantillons les plus proches spectralement ont plus de poids**

method = "LW-PLS"
lm_pls() - package plsRglm

- Nombre de composantes PLS à prendre en compte, n_{comp}

=> Agrégation des prédictions, `aggreg`

PLS locale : Agrégation des prédictions

- Agrégation
 - Agrégation des prédictions obtenues avec les modèles à 1, 2, 3,..., k composantes (`ncomp = k`)
 - Possibilité d'enlever les 1^{ères} composantes (`kmin`)
 - Choix de la méthode d'agrégation
 - Moindres erreurs quadratiques ("`RMS`")
(*Shenk et al., 1997*)
 - Méthode à poids exponentiels ("`MSE`")
(*Rigollet and Tsybakov. Sparse estimation by exponential weighting. arXiv:1108.5116*)

PLS locale : Agrégation des prédictions

- Agrégation

- Agrégation des prédictions obtenues avec les modèles à 1, 2, 3,..., k composantes ($n_{comp} = k$)

- Possibilité d'enlever les 1^{ères} composantes (k_{min})

- Choix de la méthode d'agrégation

- Moindres erreurs quadratiques ("RMS")

- (Shenk et al., 1997)

- Méthode à poids exponentiels ("MSE")

- (Rigollet and Tsybakov. Sparse estimation by exponential weighting. arXiv:1108.5116)

Nb.comp	Prediction	Poids
1.comp	26.89	0.0206
2.comp	30.13	0.1183
3.comp	31.05	0.1165
4.comp	27.36	0.1403
5.comp	23.56	0.0903
6.comp	20.67	0.0565
7.comp	24.32	0.0849
8.comp	15.53	0.0539
9.comp	16.17	0.0564
10.comp	18.33	0.0542
11.comp	15.53	0.0509
12.comp	17.13	0.0530
13.comp	14.33	0.0358
14.comp	11.82	0.0356
15.comp	9.54	0.0329

Valeur prédite = $\Sigma(\text{prediction} * \text{poids})$

Valeur prédite = 22.61

Routine PLS locale

Jeu de données (X)
valeurs de références associées (y)



Nettoyage : outliers
spectraux



Séparation
Calibration/Validation



Set Calibration



Set Validation

PRE-TRAITEMENT (réduction, transformation)



BOUCLE DE PREDICTION



PREDICTIONS

```
plotSpec(X)
```

```
outSpec(X, k=3, n=7)
```

```
calValSet(X, y, method = c("KS", "refval"),  
          per.n = 0.25)
```

```
transSpec(X, trans = c("SNV", "deriv1", ...),  
          Wmin=450, Wmax=2500, Wstep=4)
```

Routine PLS locale

Set Calibration

Set Validation

PRE-TRAITEMENT (réduction, transformation)

```
transSpec(X,trans = c("SNV","deriv1",...),  
          Wmin=450, Wmax=2500, Wstep=4)
```

BOUCLE DE PREDICTION

for (i in 1 =nrow(SetVal)) {

Selection des voisins spectraux

PLS Locale
Non pondérée
ou
Pondérée

```
localCal(z, X, Y, method=c("R2", "H"),  
        neigh=80, dlim=0.90, Nmin=50)
```

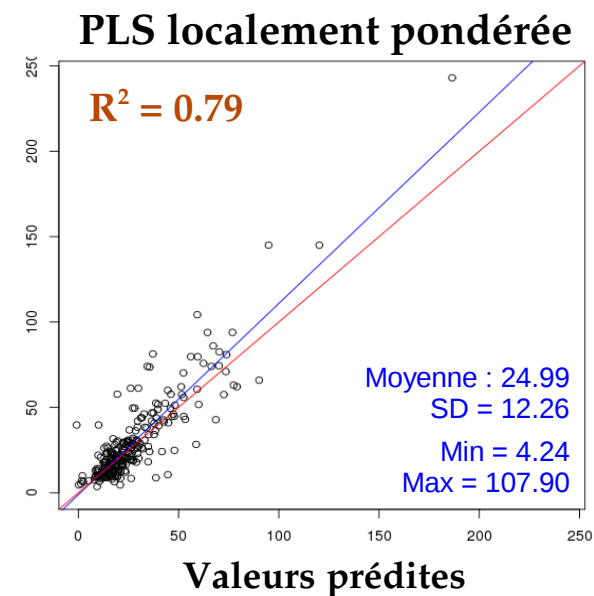
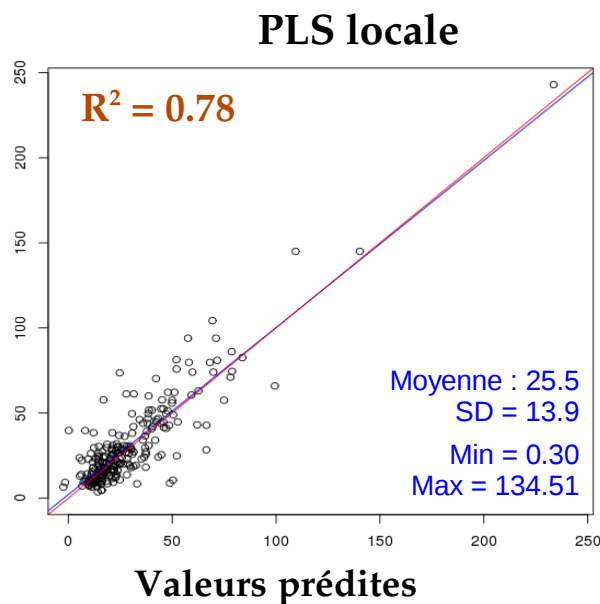
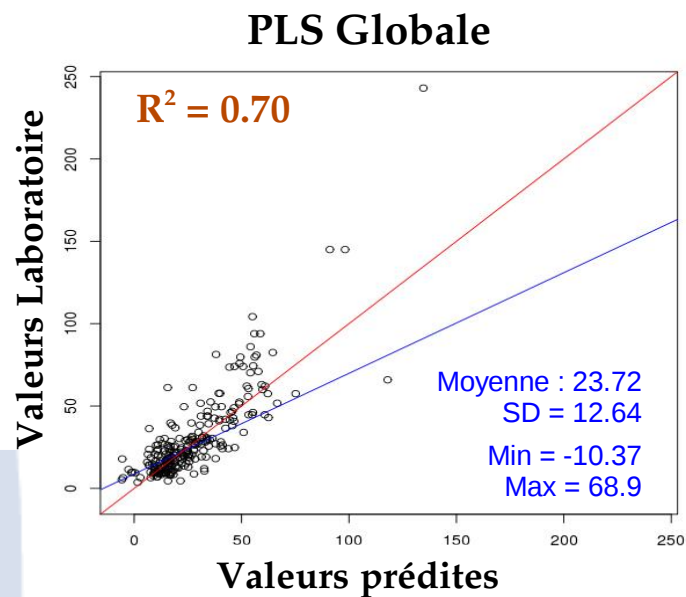
```
LwPLS (z, X, Y, weight=W, ncomp=20,  
      aggreg = c("RMS", "MSE"),
```

```
method = "LocalPLS"
```

```
method = "LW-PLS")
```

PREDICTIONS

Application à la prédiction du Carbone du Sol



Routine PLS disponible :
aurelie.thebault@supagro.inra.fr

Remerciements : Programme GESSOL - ADEME
Bénédicte Fontez, Nicolas Verzelen & Damien Juery – INRA Montpellier

