Fusion de données spectroscopiques appliquée aux sols

Lauric Cécillon

Outline

Intro – Data fusion: why ?

1 – Outer product analysis: OP-PLS / OP-PCA

2 – Principal component transform

Conclusion of our SPIRSOL paper

« Coupling NIR spectral libraries with other diffuse reflectance measurements of soils, such as mid-infrared reflectance spectra, will probably be the next step towards spectral sensing of soil quality worldwide »

Soil spectroscopy

- VIS-NIR vs MIR
- Different analyzers, techniques, labs, teams

Spectroscopic data fusion

- 2D correlation spectroscopy (I. Noda et al., since 1986)
- Outer product analysis (A. Barros, D. Rutledge et al., since 1997)

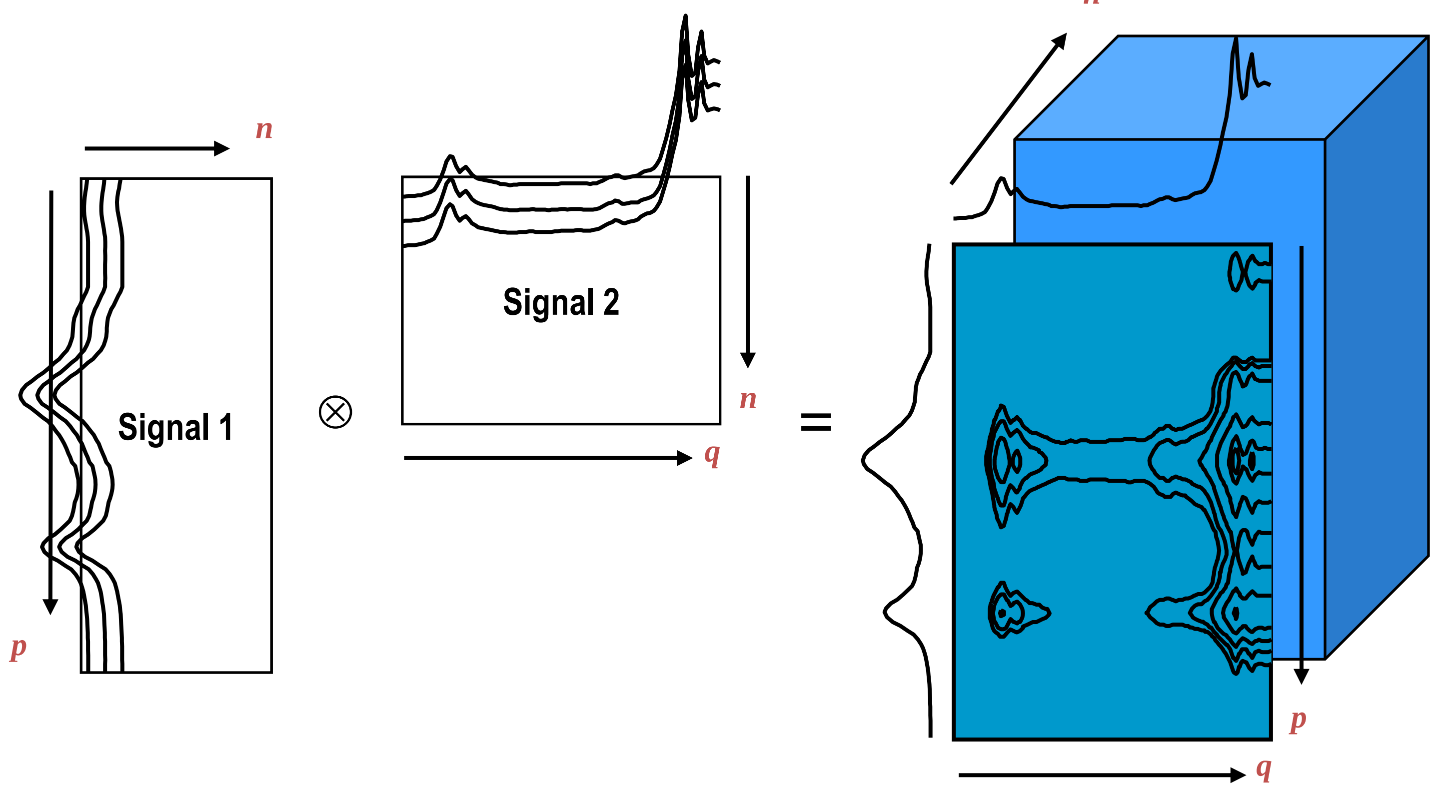
2D COS

$$\mathbf{\Phi}_{AB} = \mathbf{A}^{T}\mathbf{B}$$

$$\mathbf{\Psi}_{AB} = \mathbf{A}^{T}\mathbf{N}\mathbf{B}$$

Data fusion: an example with outer product analysis (OPA)

First: compute OP matrices for each sample



Mutual weighting of each signal by the other:

- if intensities simultaneously high in the two domains, the product is higher;
- if intensities simultaneously low in the two domains, the product is lower;
- if one intensity high and the other low, the product tends to an intermediate value

(Figure from D. Rutledge, Pers. Com.)

Detail of outer product computing

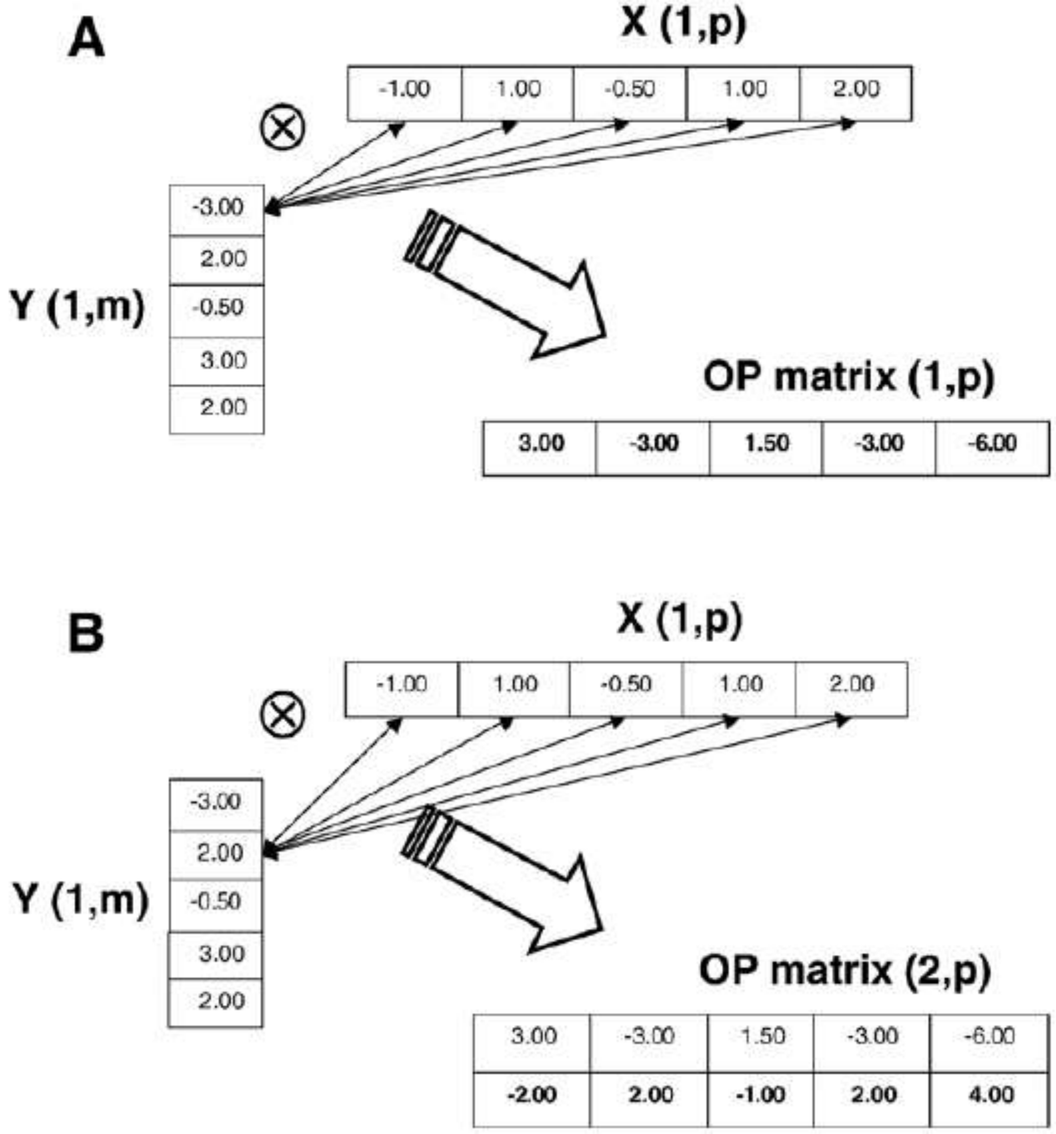
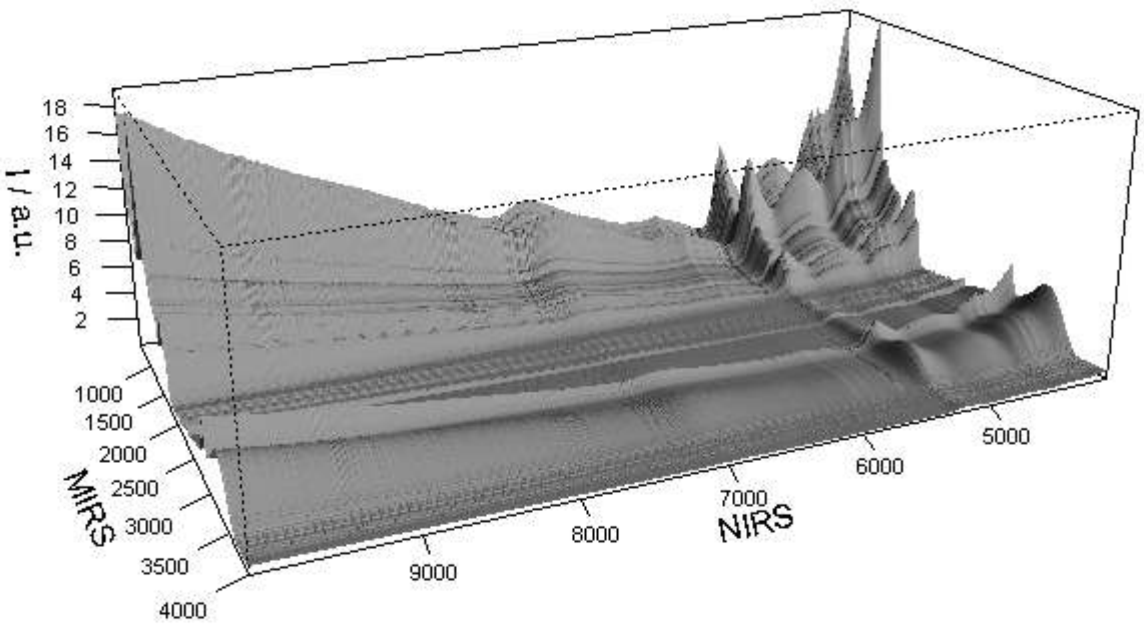
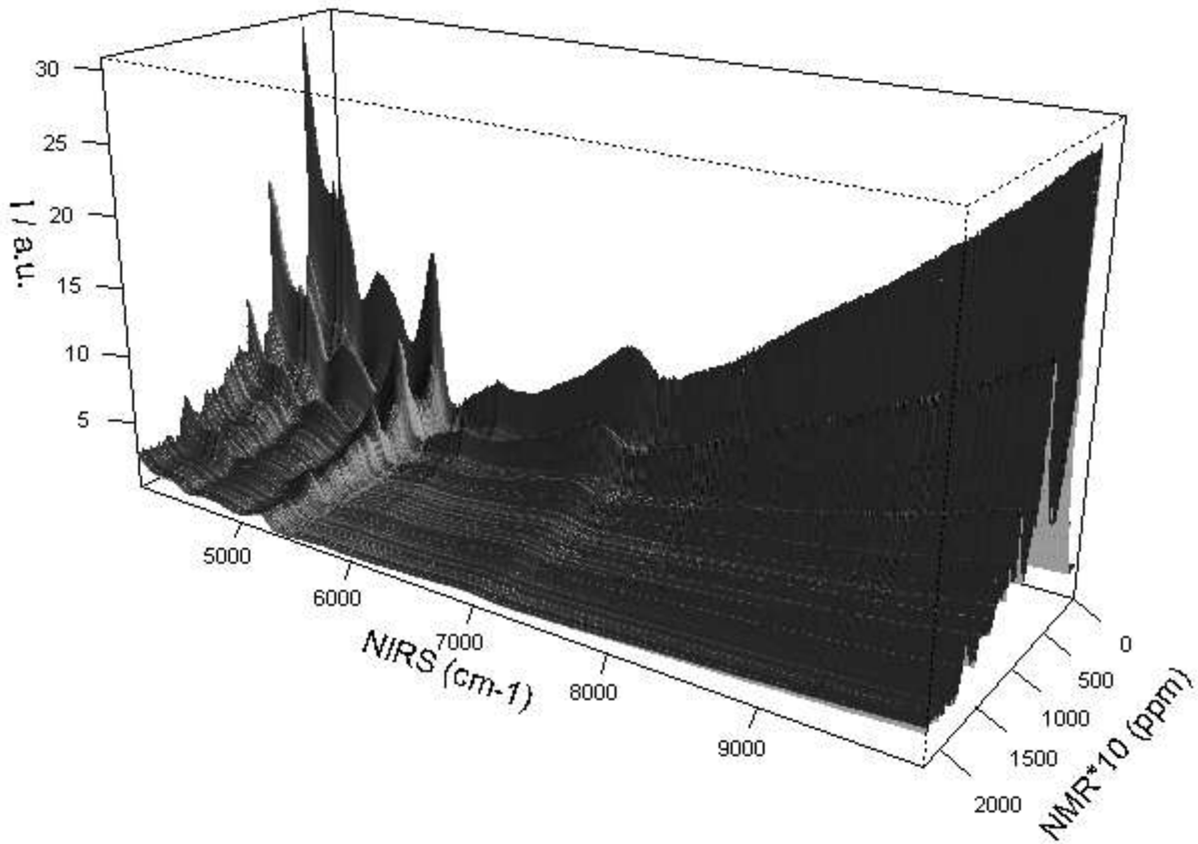


Figure from Jaillais et al., 2007 (ChemoLabS)

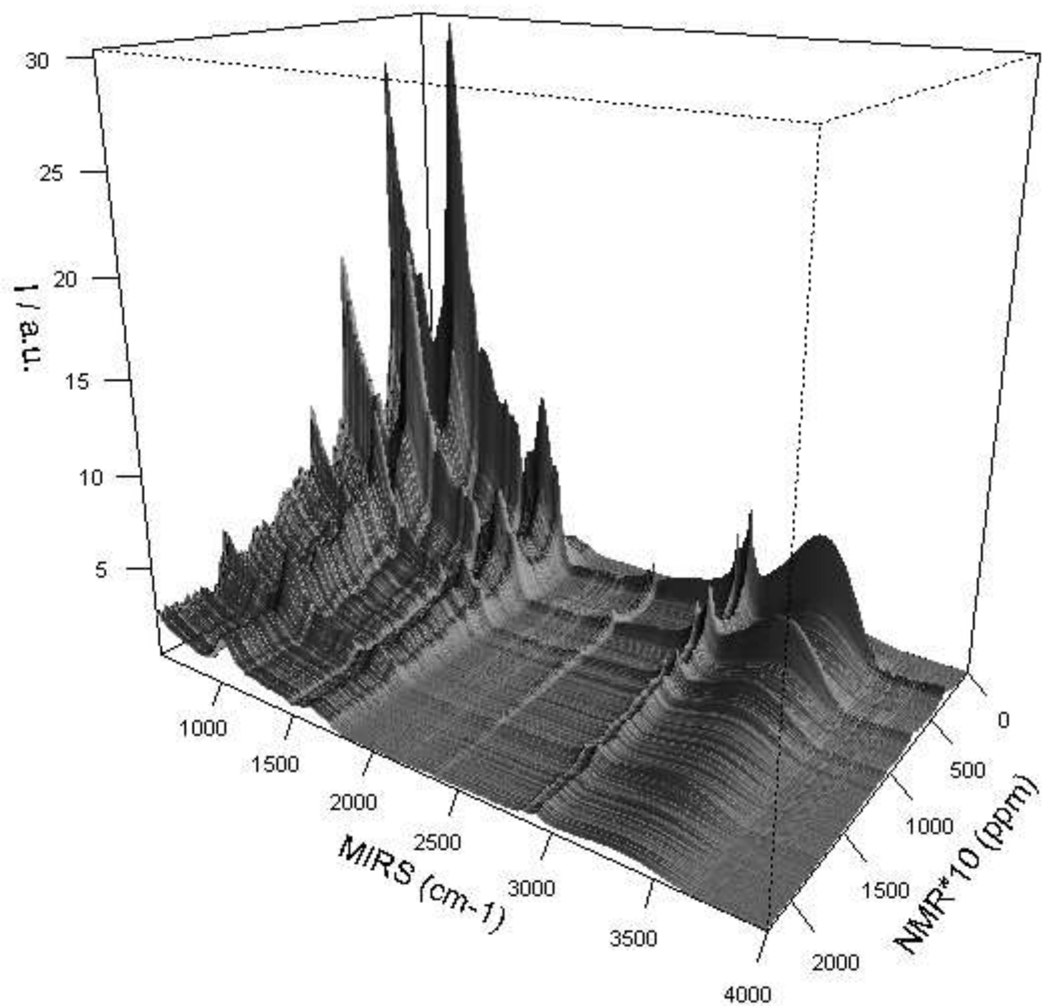
Example of the NIR-MIR OP matrix for 1 soil sample



NIR-NMR OP matrix for the same soil sample



MIR-NMR OP matrix for the same soil sample



Data fusion: an example with outer product analysis (OPA)

Unfold OP matrices, perform analysis on OP vectors, then fold-back result vector

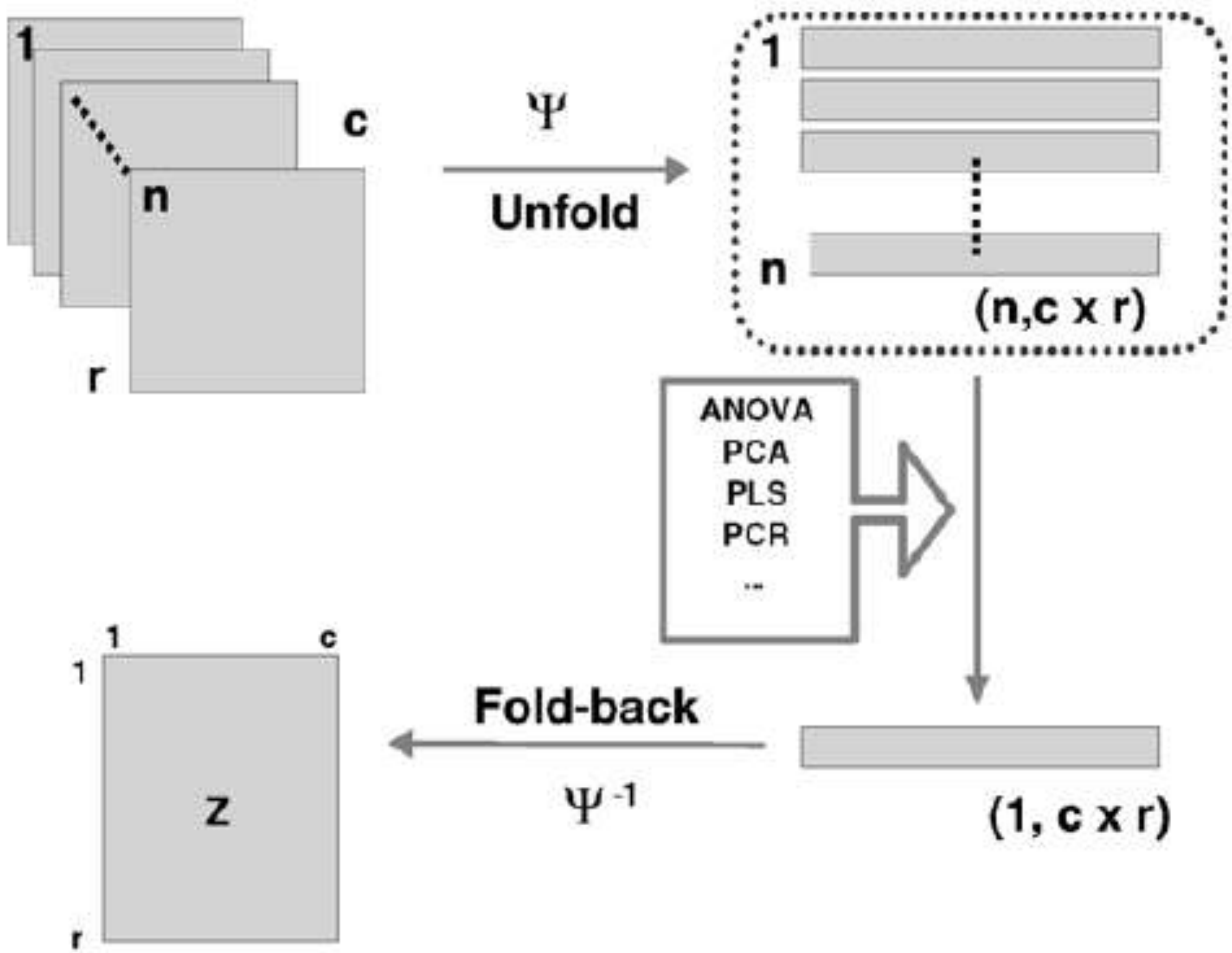


Figure from Jaillais et al., 2006 (Vibr. Spec.)

Predicting soil organic matter composition with IR-OPA

→ Outer product – partial least squares: OP-PLS

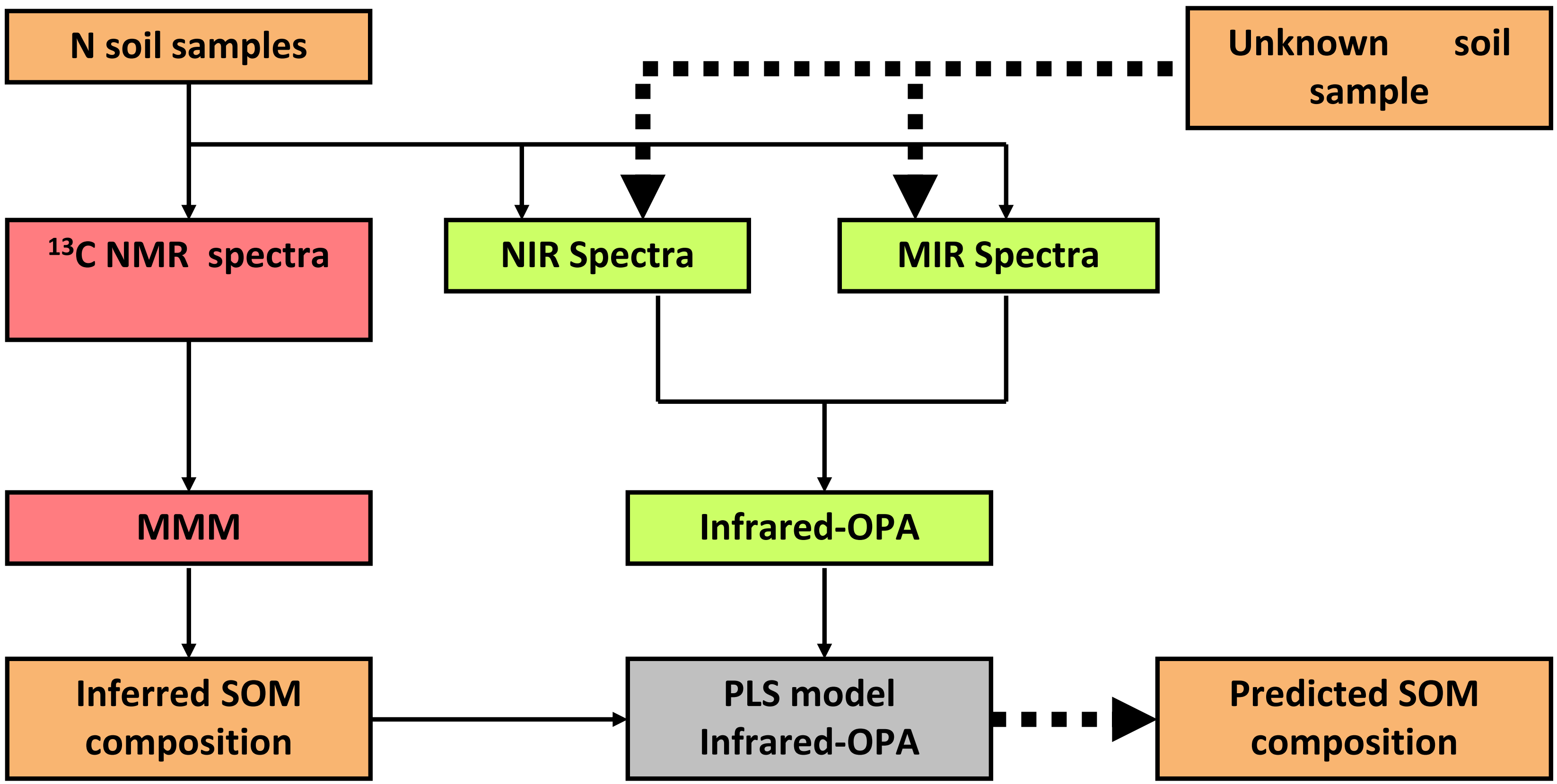


Figure from Cécillon et al., in revision

Predicting soil organic matter composition with IR-OPA

Study site

- Test on soil samples from Storgama catchment (Norway)

Photo from Live Semb Vestgarden

Predicting soil organic matter composition with IR-OPA

Study site: soils and vegetation

- Test on soil samples from Storgama catchment (Norway)

Molinia

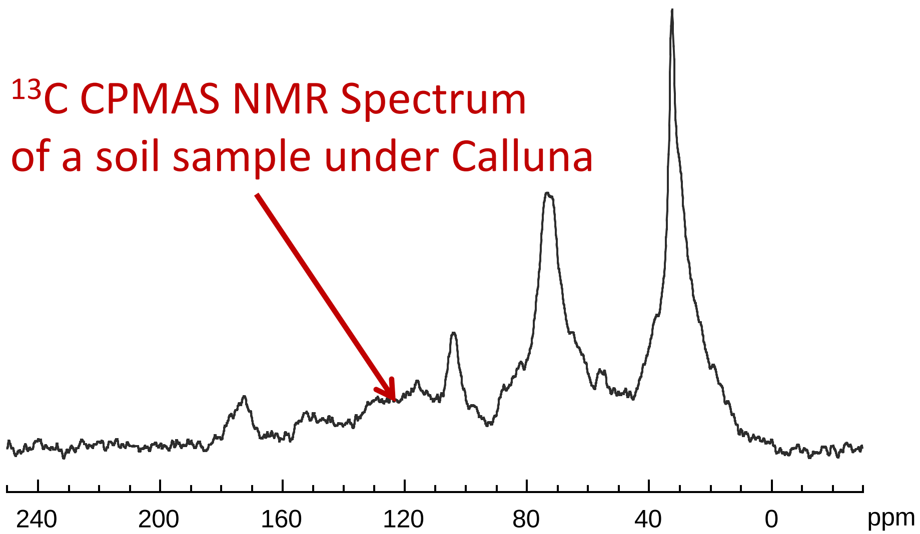
Sphagnum

Calluna

Photos from Strand et al., Ambio (2008)

Predicting soil organic matter composition with IR-OPA

^{13}C CPMAS NMR Spectrum
of a soil sample under Calluna



Alchile (0-45ppm)	O-alchile (45-110ppm)	Aromatico (110-160ppm)	Carbonile (160-220ppm)
37	**45**	**13**	**5**

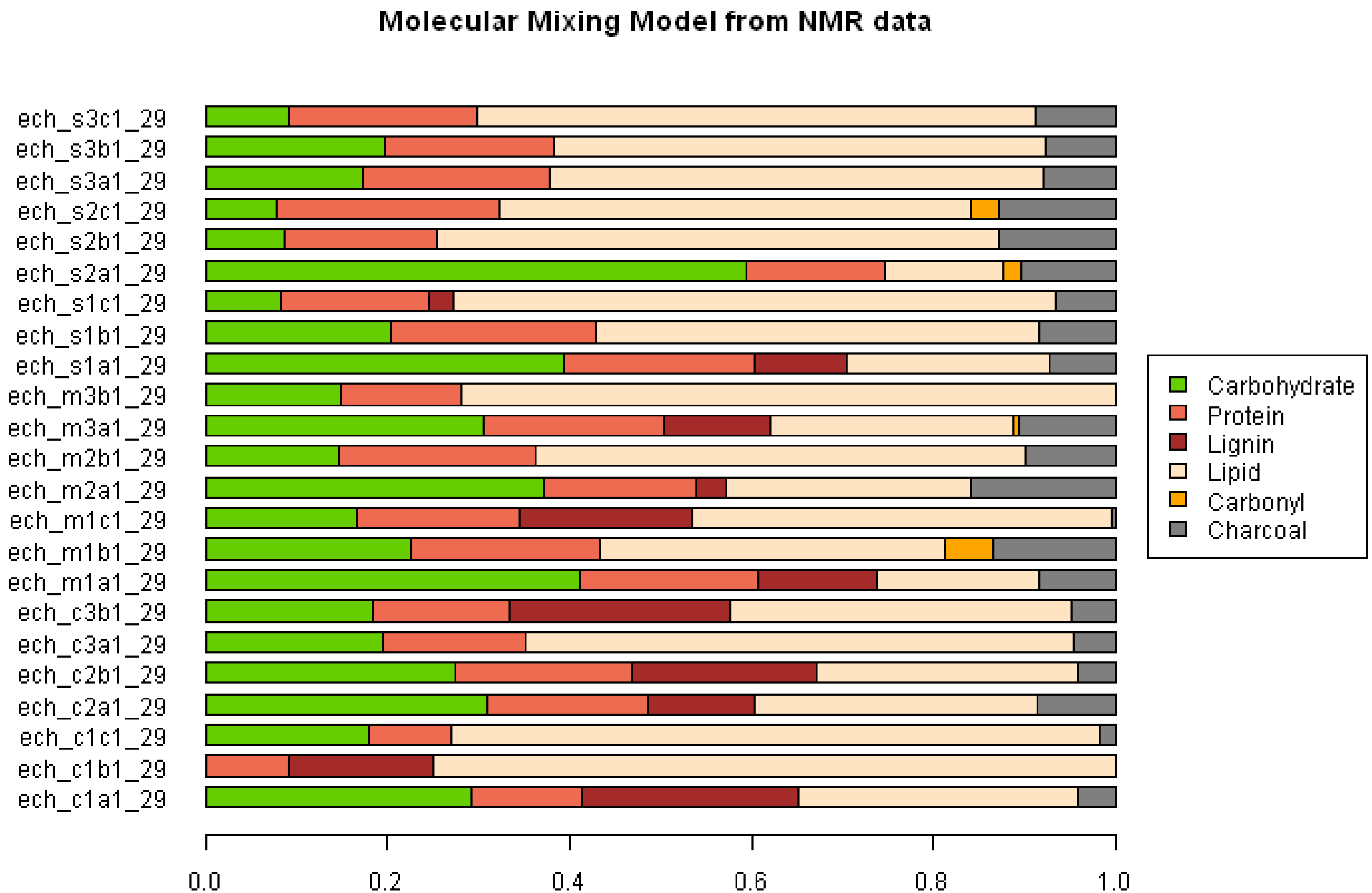
Classical interpretation

NMR – Molecular Mixing Model

Nelson & Baldock, 2005

Data from Cécillon et al., in revision

Predicting soil organic matter composition with IR-OPA



Data from Cécillon et al., in revision

Predicting soil organic matter composition with IR-OPA

- Cross-validated PLS model with 21 soil samples

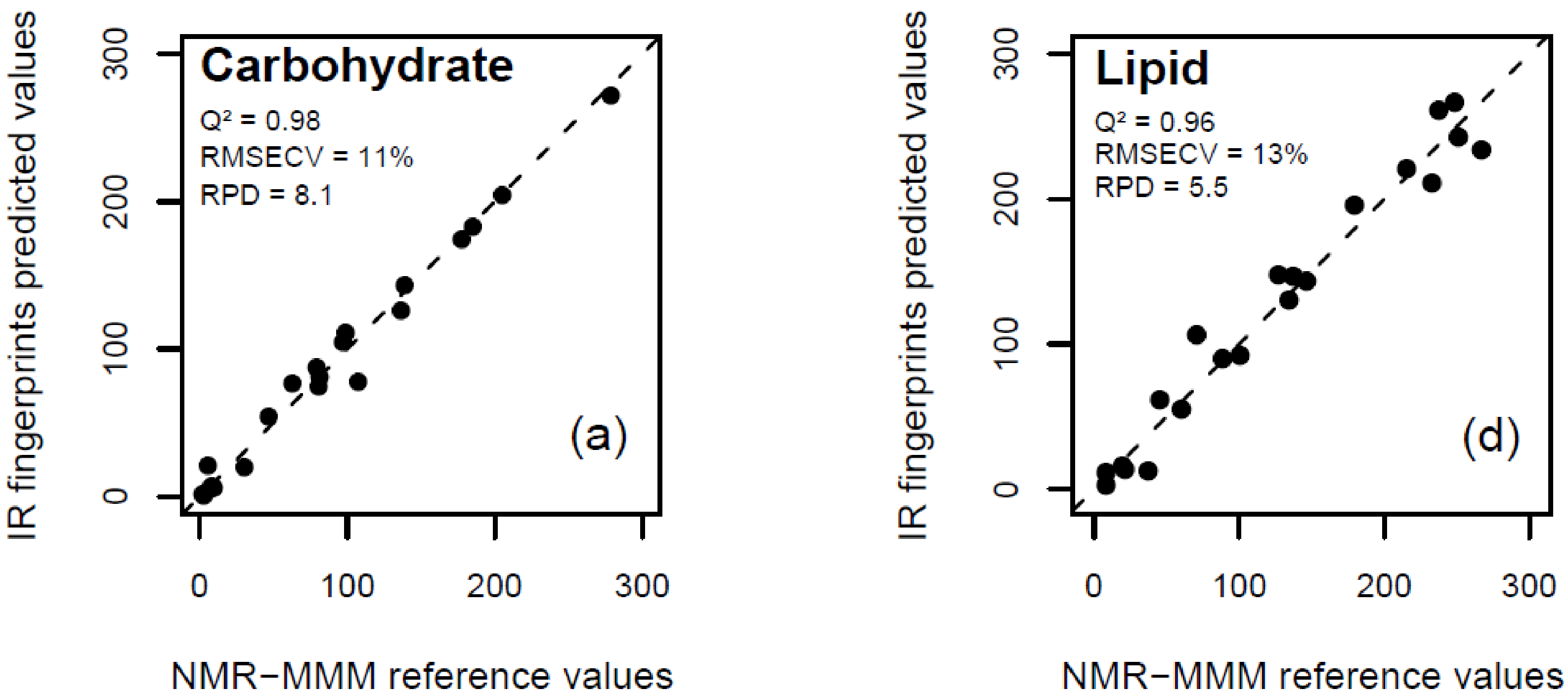
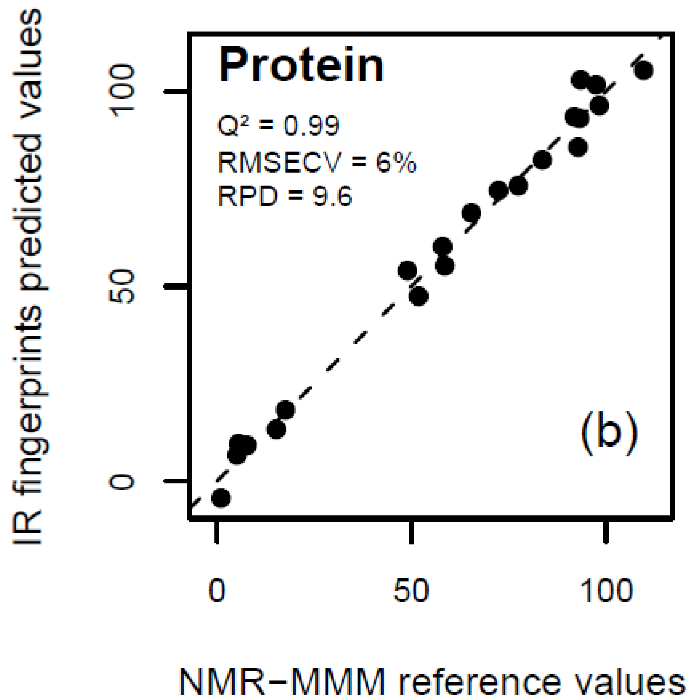


Figure from Cécillon et al., in revision

Predicting soil organic matter composition with IR-OPA

- Cross-validated PLS model with 21 soil samples



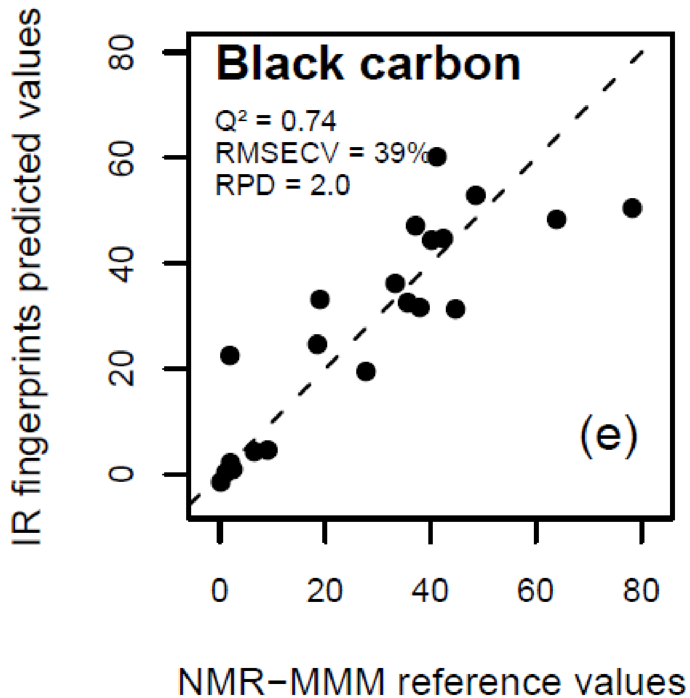


Figure from Cécillon et al., in revision

Predicting soil organic matter composition with IR-OPA

- Cross-validated PLS model with 21 soil samples

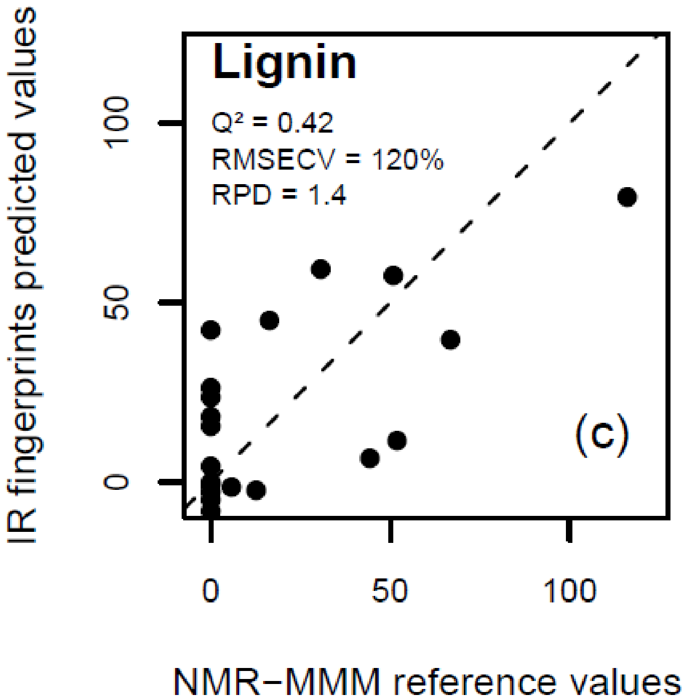
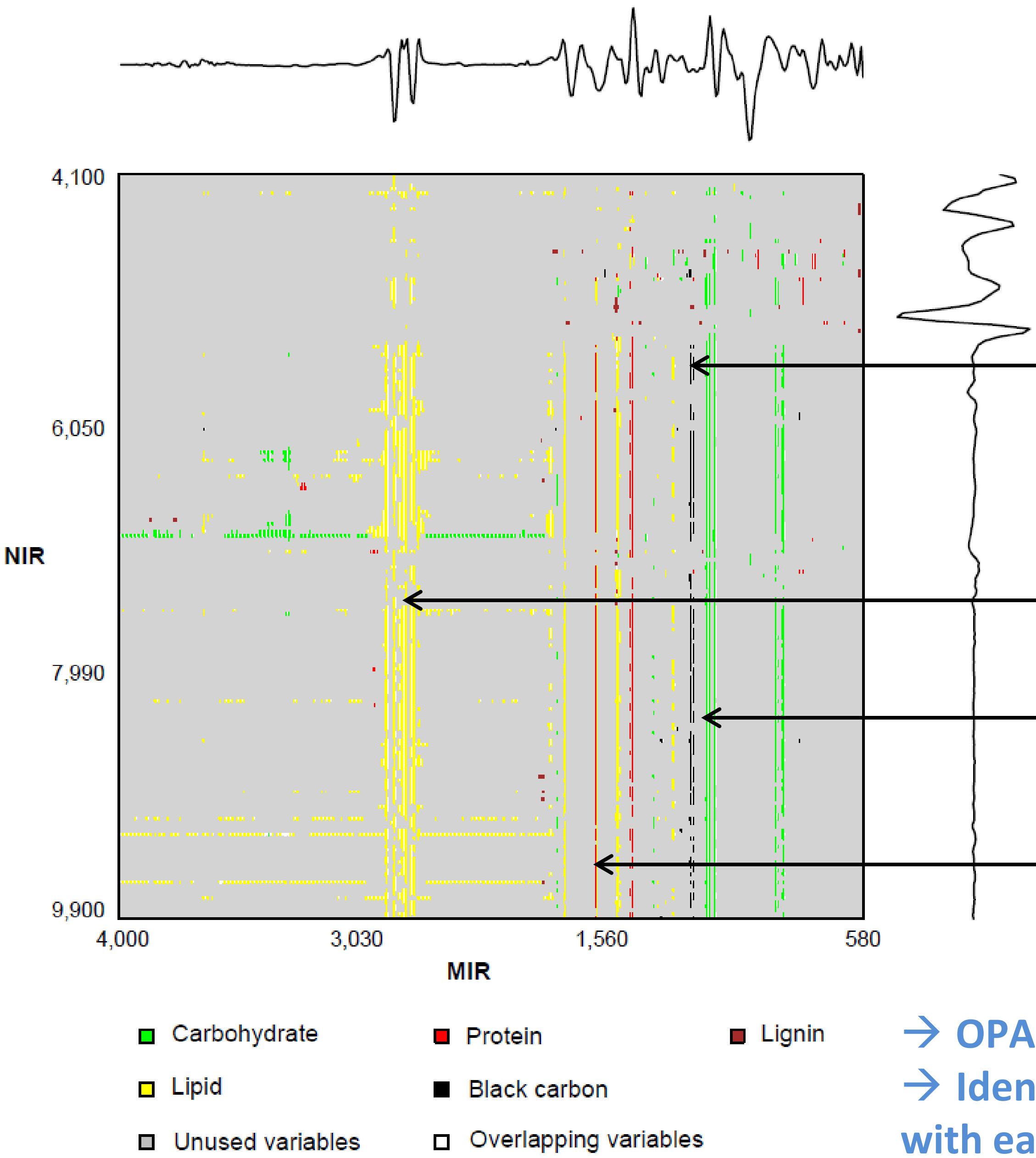


Figure from Cécillon et al., in revision



Phenol group (C-C-O or C-O stretching) and/or Aryl group (C-O-C or C-C-C stretching)

Aliphatic CH_2 & CH_3 groups (CH stretching)

Tertiary Alcohol (C-O stretching) and/or Ether group (C-O-C)

Primary, secondary, tertiary Amides (C=O stretching) and/or primary Amine and Amide groups (NH_2 scissoring)

→ **OPA-PLS with variable selection (VIP)**

→ **Identification of wavebands associated with each biochemical component**

Figure from Cécillon et al., in revision

Predicting soil organic matter composition with IR-OPA

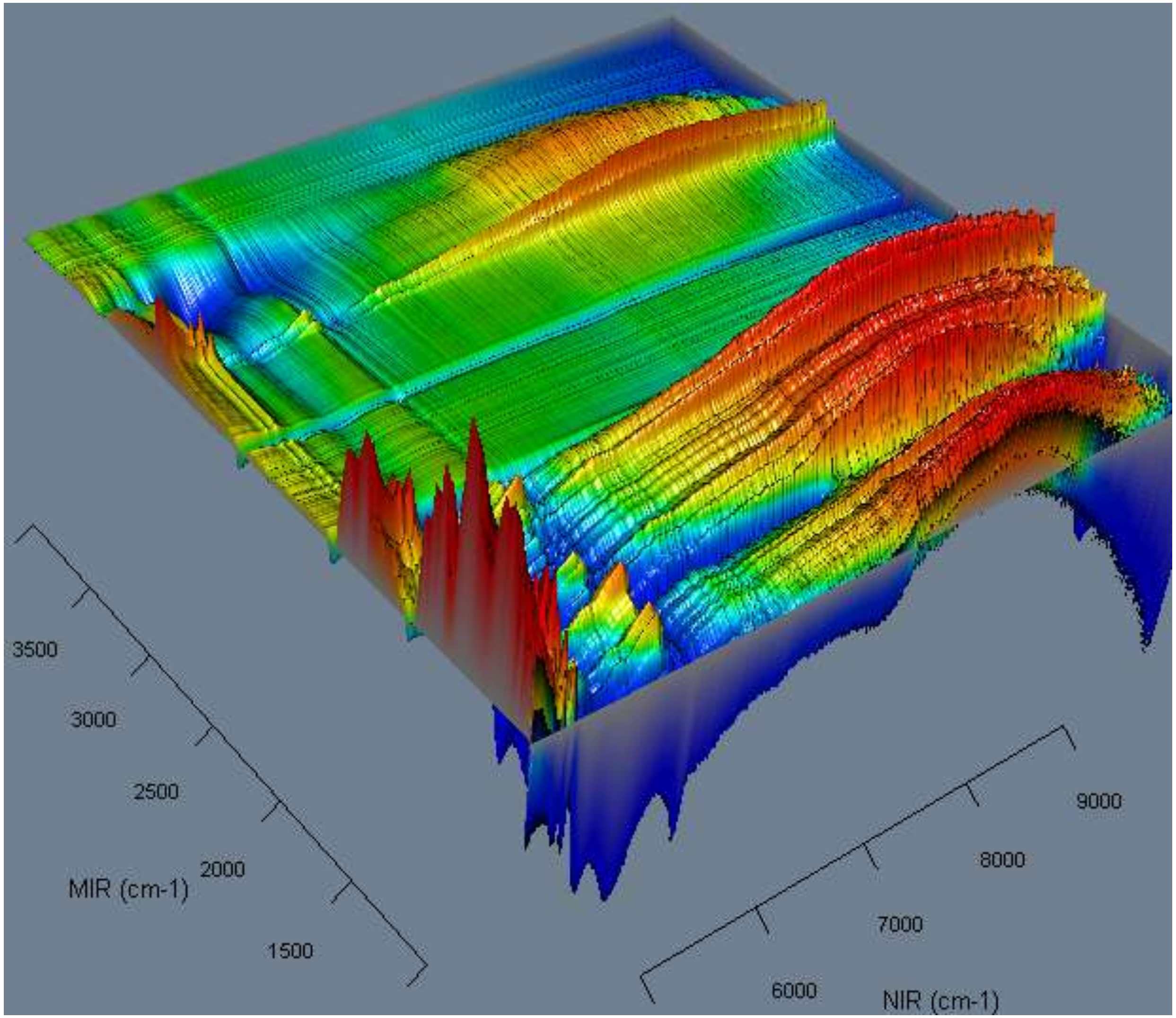
Biochemical component	NIR		Soon RPIQ!	MIR			IR fingerprints (NIR-MIR)		
	Q^2	RMSECV	RPD	Q^2	RMSECV	RPD	Q^2	RMSECV	RPD
Carbohydrate	0.95	18	4.8	0.88	29	3.1	0.98	11	8.1
Protein	0.92	17	3.7	0.98	8	7.6	0.99	6	9.6
Lignin	0.01	161	1.0	0.06	154	1.1	0.42	120	1.4
Lipid	0.79	31	2.3	0.78	32	2.2	0.96	13	5.5
Black carbon	0.55	51	1.5	0.70	42	1.9	0.74	39	2.0

Table from Cécillon et al., in revision

OP-PLS (prediction of lipid content with NIR-MIR OP matrices) → Map of B-coefficients

Model Stat. (CV-LOO):
4LV / Q^2 = 0.77 / RPIQ = 4.1

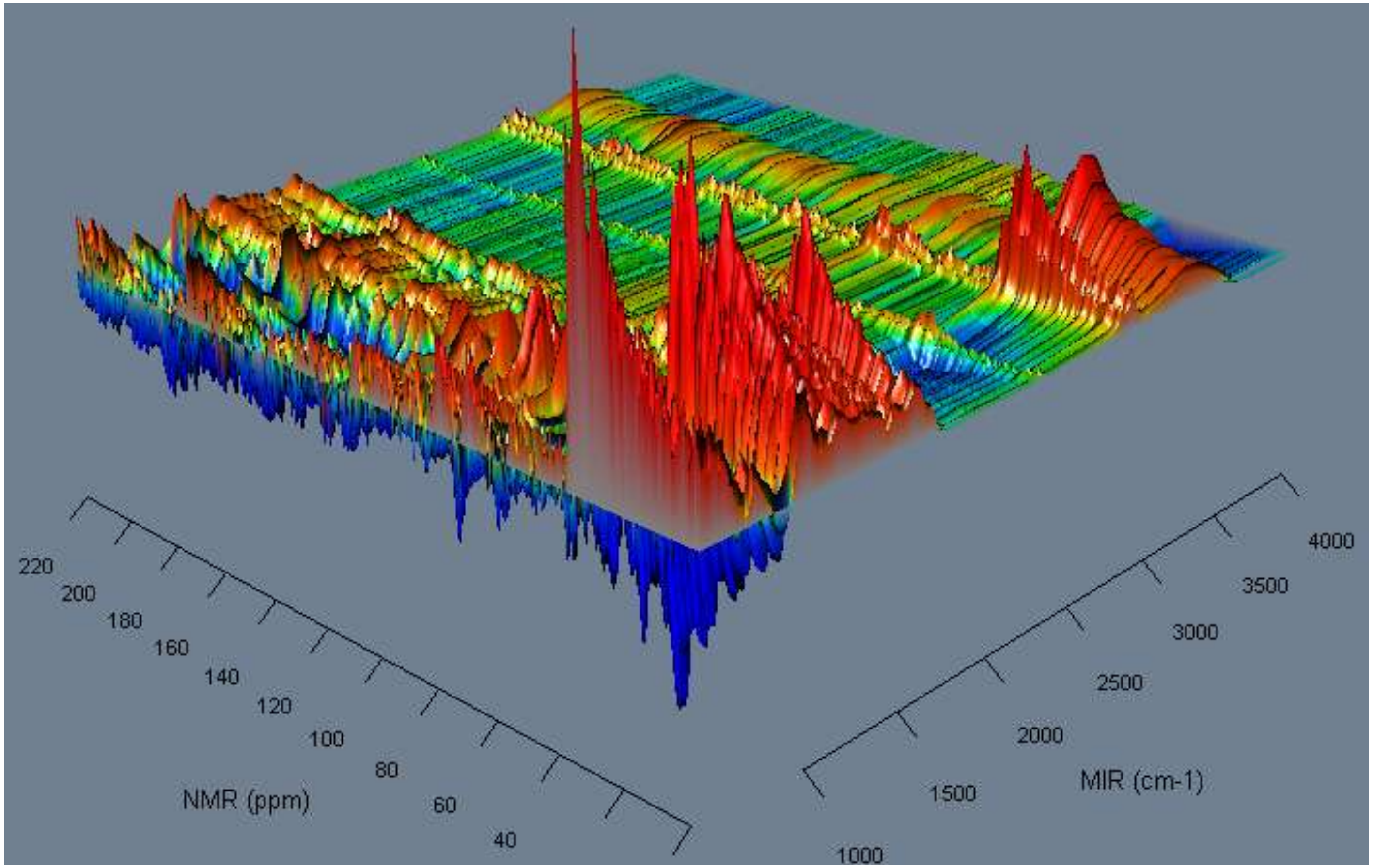
Raw spectra



OP-PLS (prediction of lipid content with MIR-NMR OP matrices) → Map of B-coefficients

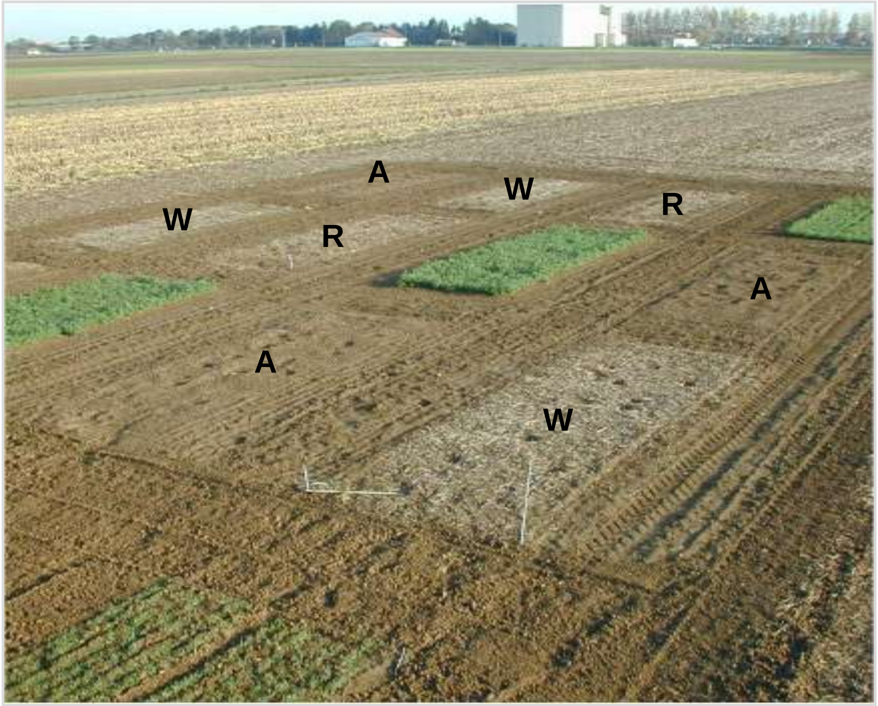
Model Stat. (CV-LOO):
5LV / Q^2 = 0.92 / RPIQ = 7.0

Raw spectra



11 months decomposition experiment: crop residues in soil

→ Principal component analysis of NIR-NIR OP matrices : OP-PCA



IR fingerprints obtained from the outer product of two NIR spectral domains:

4,220–4,835 cm^{-1} (160 wavenumbers)
&
5,685–6,100 cm^{-1} (109 wavenumbers)

=17,440 outer product variables per sample

These two NIR regions show numerous absorbance peaks related to carbohydrates, lipids, proteins, or aromatic structures

	Wheat	Rape	Alfalfa
C/N ratio	74	60	27

W : Wheat
R : Rape
A : Alfalfa

Pascault et al., in press (Mic. Ecol.)

11 months decomposition experiment: crop residues in soil

→ Principal component analysis of NIR-NIR OP matrices : OP-PCA

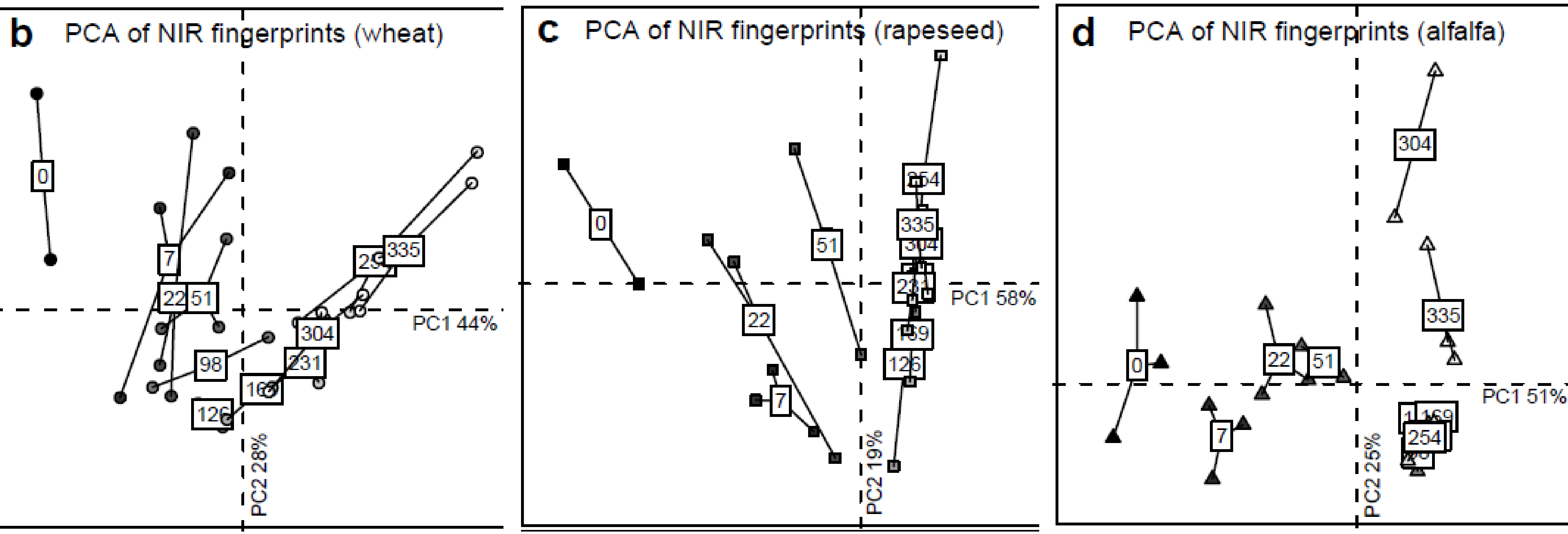


Figure from Pascault et al., in press (Mic. Ecol.)

11 months decomposition experiment: crop residues in soil

→ Principal component analysis of NIR-NIR OP matrices : OP-PCA

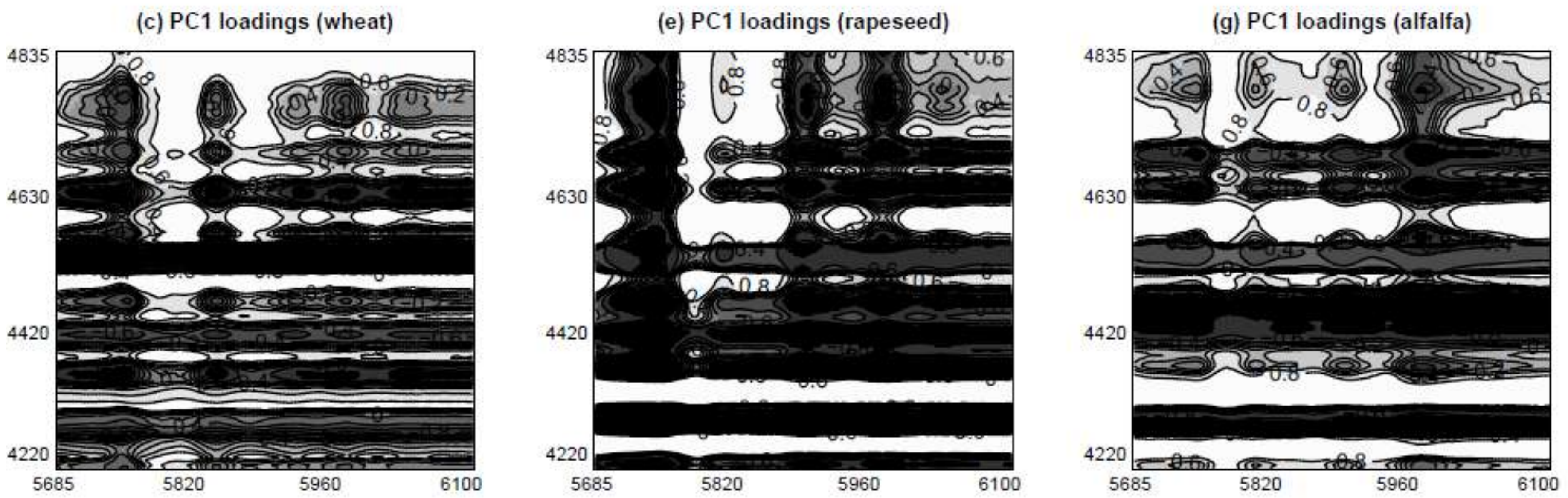


Figure from Pascault et al., in press (Mic. Ecol.)

11 months decomposition experiment: crop residues in soil

→ Principal component analysis of NIR-NIR OP matrices : OP-PCA

Table 1 NIR wavebands explanatory for PC1 and PC2 of principal component analysis of NIR fingerprints for all samples and for each residue and their corresponding chemical bonds and biochemical components

		All residues	Wheat	Rapeseed	Alfalfa
PC1	Wavebands (cm^{-1})	4800–4825[a] 4710–4725[a] 4620–4640[a] 4485–4505[a]	4790–4835[a] 4515–4550[a] 5785–5820[b] 4620–4635[b] 4485–4505[b] 4355–4370[b]	4320–4350[a] 4280–4305[a] 4245–4270[a] 5825–5865[b] 5770–5795[b] 5730–5750[b] 4575–4605[b] 4365–4380[b]	4425–4475[a] 4315–4355[a] 4275–4295[a] 4230–4265[a] 5765–5790[b] 4705–4765[b] 4560–4610[b]
	Chemical bonds[c]	NH, OH, CO, CN, aromatic CH[a]	NH, OH, aromatic CH; CHO[a]; aliphatic CH; aromatic CH; NH; $CONH_2$[b]	Aliphatic CH, aromatic CH, and CH[a]; aliphatic CH, CH, aromatic CH, NH, $CONH_2$[b]	CH, aliphatic CH, and aromatic CH[a]; CH, CO, and NH[b]
	Biochemical components[b]	Protein, alcohol, water, and aryl[a]	Protein, water, alcohol, aryl, and carbohydrate[a]; lipid[b]	Lipid, carbohydrate[a] aryl, protein, and carbohydrate[b]	Lipid and carbohydrate[a]; aryl, protein, carbohydrate[b]

Table from Pascault et al., in press (Mic. Ecol.)

Principal component transform

Speeding-up processing without loosing any information

R script available !

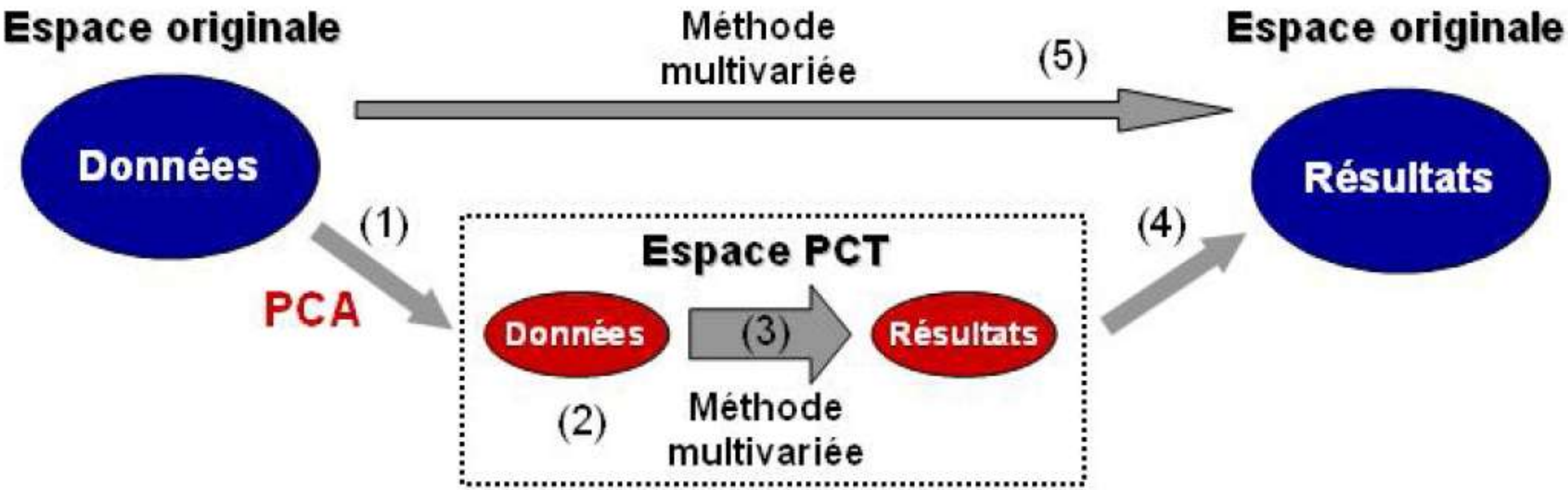


Figure from Pinto, 2009 (PhD Thesis)

Principal component transform

Speeding-up processing without loosing any information

R script available !

High-level description for the OP-PCT-PCA algorithm

$$\mathbf{X} = \mathbf{T} \cdot \mathbf{P}^{\mathrm{T}}$$

Step	Computation	Comments
1	$\mathbf{X},\mathbf{Y}$	Input of **X** and **Y** matrices
2	$[\mathbf{T}_X, \mathbf{P}_X] \longleftarrow \mathrm{PCA}(X)$	Full rank PCA of **X**
3	$[\mathbf{T}_Y, \mathbf{P}_Y] \longleftarrow \mathrm{PCA}(Y)$	Full rank PCA of **Y**
4	$\mathbf{K} = \mathbf{OP}(\mathbf{T}_X, \mathbf{T}_Y)$	Outer product (OP) between $\mathbf{T}_X$ and $\mathbf{T}_Y$
5	$[\mathbf{T}, \mathbf{P}_{\mathrm{PCT}}] = \mathrm{PCA}(\mathbf{K})$	Full rank PCA of **K**. **T** represents the scores of the original space. PPCT represents the PCT loadings (compressed space)
6	for $a = 1$:PC unfold $\mathbf{P}_{\mathrm{PCTa}}$ $\mathbf{P}_a = \mathbf{P}_Y\, \mathbf{P}_{\mathrm{PCTa}}\, \mathbf{P}_X^{\mathrm{T}}$ end for	Rebuild each Principal Component's loadings (a) – Eq. (13)

Table from Barros et al., 2008 (ChemoLabS)

Principal component transform

Speeding-up processing without loosing any information

R script available !

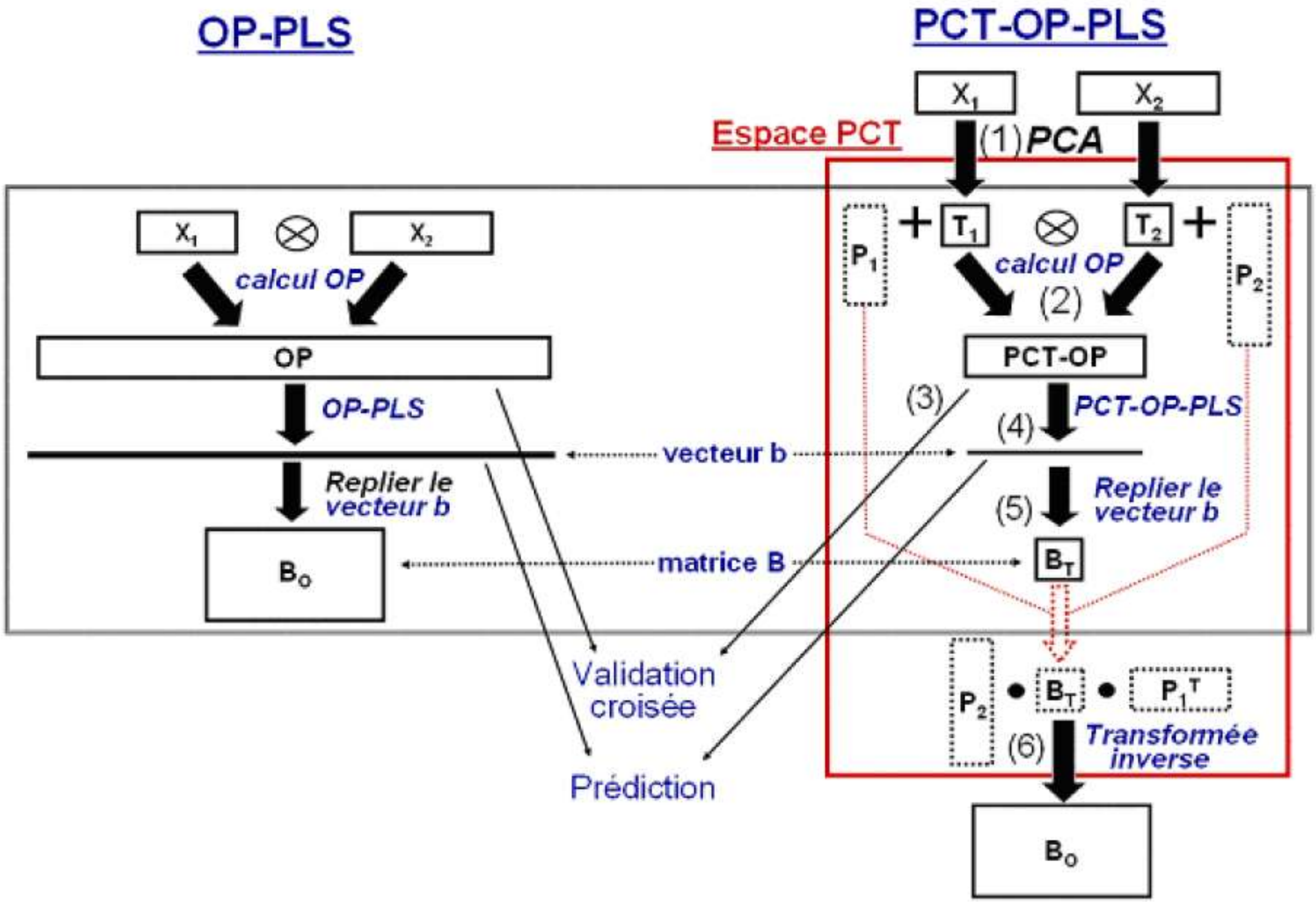
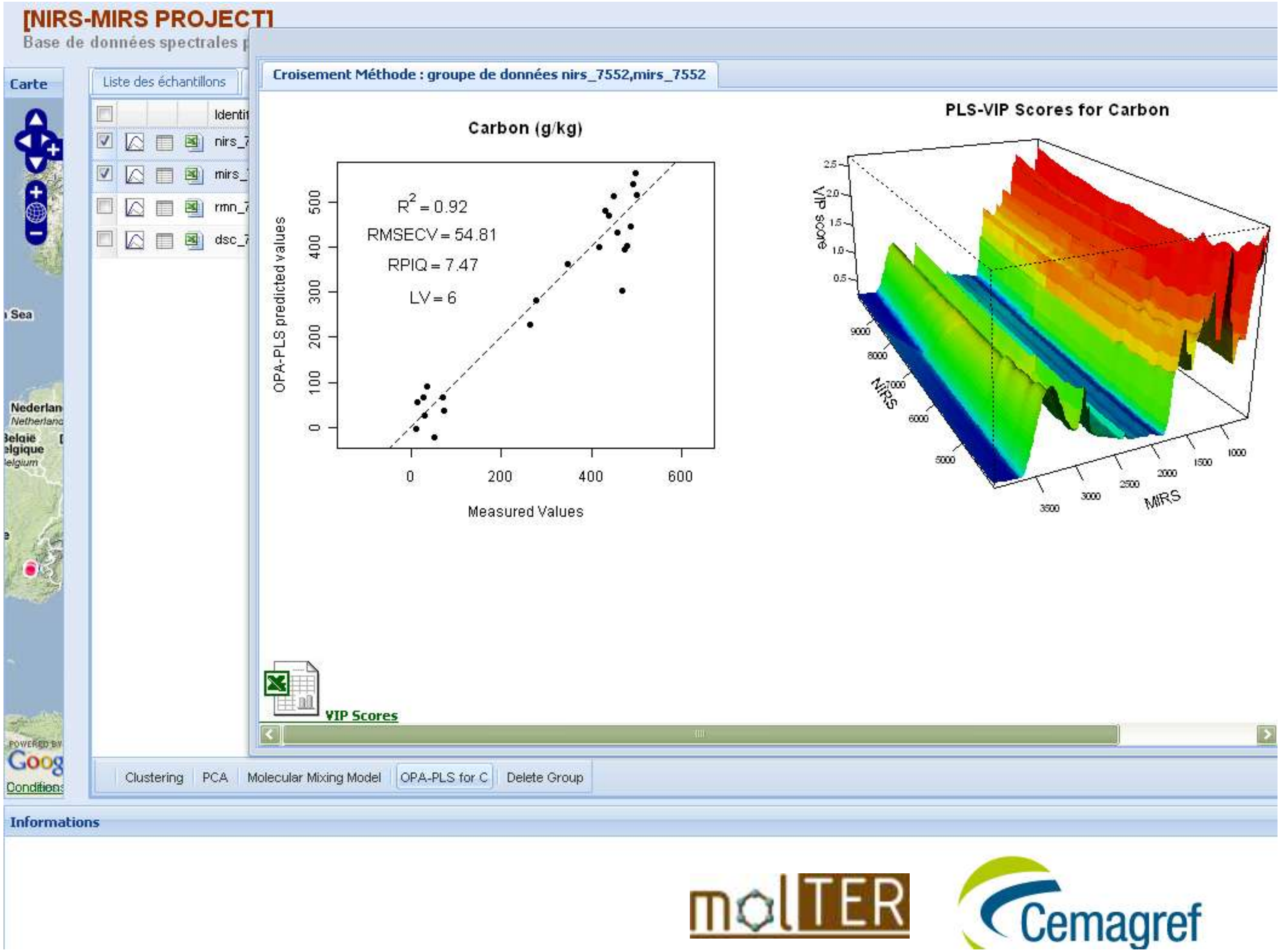


Figure from Pinto, 2009 (PhD Thesis)

OP-PCT-PLS implemented in Cemagref DB (MOLTER initiative)



Merci !

Remerciements particuliers :

B. Barthès

L.T. Strand, G. Certini, C. Forte, H. Lange

N. Pascault, P.A. Maron

D. Rutledge, C. Cordella, A. Barros

F. Bray, S. De Danieli, E. Ancelet, A. Saenger, J.J. Brun

D. Rasse

Contact :

http://lauric.cecillon.free.fr/ cecillonlauric@yahoo.fr