



Darégal Bretagne
Service approvisionnement
ZA de Penhoat
29800 Saint-Divy

Expérimentation : Test d'une méthode d'évaluation du ratio Limbes/Pétioles sur persil plat frais

Spectrométrie proche infra-rouge



Résumé

Evaluation du ratio Limbes/Pétioles sur persil plat frais par spectrométrie proche infrarouge

Pour la production du persil *Darégal*, ce sont majoritairement les feuilles qui sont utilisées; d'où l'importance d'avoir un bon pourcentage de feuilles sur la matière première. L'objectif de ce travail a été de tester la spectrométrie proche infra-rouge sur le calcul du rapport foliaire du persil plat frais. L'expérimentation montre que le ratio foliaire peut être prédit suite à l'analyse d'échantillons de persil plat, frais, brut de récolte, par un spectromètre proche infra-rouge. Néanmoins, une incertitude sur la mesure réside.

Cette étude montre qu'un modèle de calcul peut être créé et que la méthode de référence peut être améliorée.

La spectrométrie proche infrarouge s'avère être un outil efficace pour la détermination du rapport foliaire directement sur la ligne de production.

MOTS CLES: Ratio Limbes/Pétioles, persil plat, Spectrométrie proche infra-rouge

Abstract

Leaves/Stems ratio assessment on fresh parsley by near infrared spectrometry

For *Daregal's* parsley production, leaves are mainly used; hence the importance to have a raw material with a good leaves percentage. The objective of this experiment was testing the near infrared spectrometry on fresh parsley leaves' ratio calculation. This experiment shows that the leaves' ratio can be predicted following the fresh parsley sample analysis, fresh, gross harvest, thanks to a near infrared spectrometer. Nevertheless, there is a measurement uncertainty.

This study shows that a calculation model can be created and that the reference method can be improved.

Near-infrared spectroscopy is proving to be an effective tool for determining the leaves' ratio directly on line's production.

KEYWORDS: Ratio Leaves/Stems, parsley, near infrared spectrometry

Remerciements

Je tiens tout d'abord à remercier Mr Anthony RAUD, mon maître de stage, de m'avoir confié ce projet sur la spectrométrie proche infrarouge durant mes trois mois de stage et de m'avoir aidé et suivi afin de le mener jusqu'au bout.

Je remercie toute l'équipe de *Darégal Bretagne* pour son accueil chaleureux et son intégration au sein de l'entreprise. La convivialité qui règne dans l'entreprise m'a permis de travailler dans un cadre très agréable.

Je remercie également tout le *groupe Darégal* de m'avoir permis de réaliser ce stage et de m'avoir permis d'effectuer un déplacement au Centre Wallon de Recherches Agronomiques à Gembloux en Belgique afin de mieux comprendre et préparer l'expérimentation.

Suite à ce déplacement, je remercie Mr Philippe VERMEULEN, Mr Stéphane BRICHARD, Mr Nicaise KAYOKA, Mr Bernard LECLER, Mr Vincent BAETEN et Mr Olivier MINET de m'avoir accueilli dans leur Centre Wallon de Recherches Agronomiques. Je les remercie de m'avoir fait découvrir la technologie et l'application de la spectrométrie proche infra-rouge dans le domaine de l'agroalimentaire.

Je tiens également à remercier Mr Jean-Claude BOULOU et Mr Christophe TARDY de chez *Bruker* pour les conseils apportés durant l'expérimentation et pour l'utilisation et la configuration du spectromètre Matrix-F et du logiciel OPUS.

Je remercie également les professeurs de *Unilasalle*, notamment Mme Nathalie ROBERT de m'avoir aidé avec l'analyse des données et Mme Alicia AYERDI GOTOR pour les conseils sur la spectrométrie proche infrarouge et la mise en place de l'expérimentation.

Enfin je remercie toutes les personnes qui sont restées en contact tout au long du projet et qui m'ont aidées et conseillées pour les différentes étapes de l'expérimentation. Je pense notamment à Mme Bernadette JULIER de l'*INRA*.

Table des matières

Résumé	2
Abstract	2
Remerciements	3
1. Le rayonnement proche infrarouge	6
1.1. Définition	6
2. La spectroscopie proche infrarouge	6
2.1. La méthode	6
Définition	6
2.2. Principe	6
3. Centre wallon de Recherches agronomiques de Gembloux (Belgique)	7
3.1. Premières analyses du persil plat frais par spectrométrie proche infra-rouge	7
Objectifs	7
Matériel.....	7
3.2. Protocole expérimental.....	8
Préparation des échantillons	8
3.3. Résultats obtenus.....	9
Spectrométrie proche infrarouge	9
3.4. Modèles développés sur base des échantillons de mélanges (0%, 25%, 50%, 75% et 100% feuille) et validé sur l'échantillon brute	11
Calibration/Validation croisée	11
Imagerie proche infra-rouge	15
3.5. Conclusions et perspectives	15
4. Expérimentation <i>Darégal</i>	16
4.1. Méthode de référence de mesure du rapport Limbes/pétioles	16
4.2. Spectromètre proche infrarouge.....	16
4.3. Protocole expérimental (phase de collecte des données)	16
Matériel.....	16
Réglages et paramétrage du spectromètre	17
Protocole de mesure	17
4.4. Etablissement de l'Equation de prédiction (phase de modélisation).....	17
4.5. Test de l'équation (Validation de la méthode)	18
5. Résultats de l'expérimentation	18
5.1. Premiers résultats (au 19.07.2016)	18
5.2. Création des modèles et analyse	19
5.3. Méthode de référence	19

Mesures.....	19
Statistiques descriptives	20
5.4. Comparaisons des modèles de mesure du ratio L/P par spectrométrie proche infrarouge.....	20
La méthode PLS (Partial Least Square)	20
Statistique descriptive.....	21
Statistique inférentielle	21
5.5. Graphiques obtenus pour quelques modèles	23
5.6. Comparaisons valeurs vraies/valeurs prédites	26
6. Modèle avec la totalité des échantillons	32
7. Discussion des résultats.....	34
7.1. Interprétation des résultats	34
7.2. Limites	34
7.3. Recommandations.....	35
Conclusion.....	35
Liste des abréviations et symboles.....	37
Liste des références bibliographiques	38
Annexes.....	39

1. Le rayonnement proche infrarouge

1.1. Définition

Le rayonnement proche infrarouge comprend les longueurs d'ondes de la lumière situées entre la lumière visible et l'infrarouge moyen (Figure1). Il s'étend environ de 770 à 2500 nm (12900 cm^{-1} à 4000 cm^{-1}).

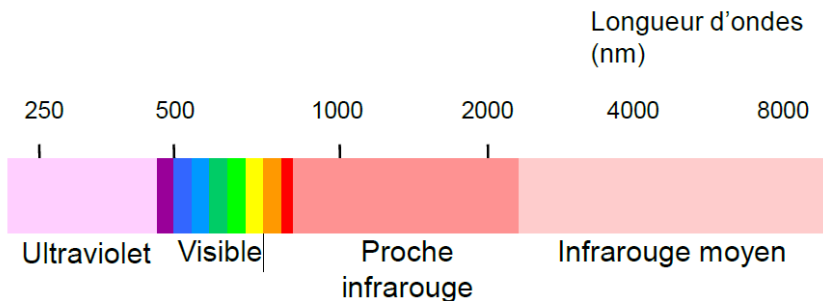


Figure 1 - Les différentes zones du spectre lumineux - *Areas of light spectra (CIRAD, 2013)*

2. La spectroscopie proche infrarouge

2.1. La méthode

Définition

La spectroscopie proche infrarouge est une technique analytique quantitative qui permet la mesure de l'absorption de la lumière proche infrarouge (longueur d'onde et intensité) par un échantillon. Le proche infrarouge permet notamment d'identifier la présence de molécules organiques dans la matière analysée. Le principe de la technique repose sur les propriétés vibrationnelles des molécules et leurs interactions avec la lumière. Chaque vibration peut être caractérisée avec une longueur d'onde.

2.2. Principe

Le faisceau de lumière est projeté sur l'échantillon. L'énergie lumineuse d'une fréquence de radiation donnée est absorbée par une liaison moléculaire qui a la même fréquence de vibration. On peut donc relier une longueur d'onde donnée à une liaison chimique donnée et sa vibration. Dans le proche infrarouge, une liaison chimique peut être caractérisée par plusieurs longueurs d'ondes contrairement au moyen infrarouge. C'est donc tout le spectre qui est pris en compte lors du traitement des données.

L'absorption de l'énergie est régie par la Loi de Beer Lambert : absorbance proportionnelle à la concentration du constituant.

$$A_{\lambda} = \epsilon \cdot l \cdot C$$

A : Absorbance (sans unité)

ϵ : coefficient d'extinction molaire ($\text{L} \cdot \text{mol}^{-1} \cdot \text{cm}^{-1}$ ou $\text{cm}^2 \cdot \text{mol}^{-1}$)

l : trajet optique (cm)

C : concentration de la substance dans la solution (mol/L)

Cette méthode, basée sur l'absorption par la matière organique du rayonnement proche infrarouge, permet de déterminer soit des composantes chimiques, soit des propriétés physiques d'un composant/d'un mélange. Elle nécessite le développement d'équations et leur mise à jour régulière.

3. Centre wallon de Recherches agronomiques de Gembloux (Belgique)

Avant de débiter l'expérimentation j'ai effectué un déplacement au CRA-W. L'objectif principal était de découvrir le fonctionnement du dispositif Matrix-F et son utilisation sur des échantillons de persil plat frais. Une première étude de faisabilité a été effectuée.

3.1. Premières analyses du persil plat frais par spectrométrie proche infra-rouge

Objectifs

Dans un premier temps, l'objectif était de réussir à discriminer les feuilles des tiges grâce à 2 méthodes :

- Imagerie proche infra-rouge hyperspectrale
- Spectroscopie proche infra-rouge (Spectroscopie Matrix-F de chez *Bruker*)

Dans un second temps, l'objectif était de trouver le pourcentage de feuilles d'un échantillon brut de persil plat frais.

Matériel

- Echantillon d'environ 600g de persil plat frais. (ratio L/P moyen : 51%)
- Spectroscopie NIR Matrix-F de chez *Bruker*
- Sonde d'émission Q412 du Matrix-F
- Tapis roulant
- Caméra linescan hyperspectrale NIR

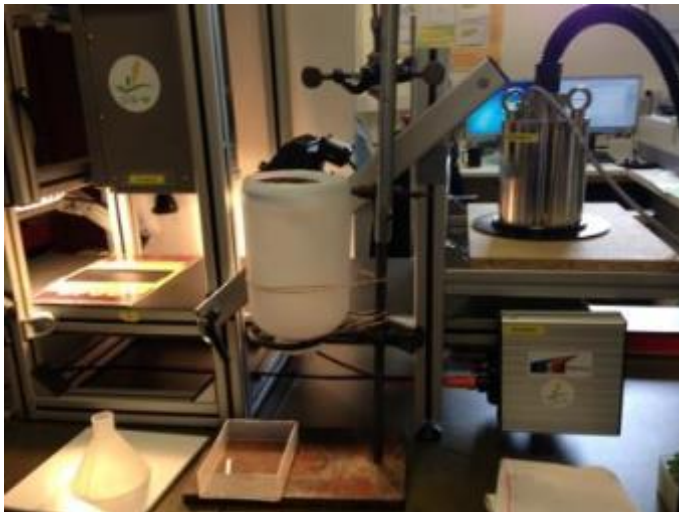


Figure 3 - Dispositif complet : tapis roulant + caméra + sonde Q412



Figure 2 - Matrix-F relié à l'ordinateur avec Logiciel OPUS 7.5

3.2. Protocole expérimental

Deux analyses sont réalisées en parallèle sur chaque échantillon de persil plat frais :

- L'échantillon est placé sur le tapis roulant qui est ensuite mis en mouvement
- L'échantillon passe d'abord sous la caméra linescan hyperspectrale NIR
- Puis sous la sonde PIR Q412
- Les images et spectres sont recueillis dans des logiciels et ordinateurs différents (propres à chaque méthode de mesure)
- Analyse des spectres et des images

Préparation des échantillons

1. Séparation feuilles/tiges persil plat frais

Ces feuilles et tiges vont permettre de créer manuellement des ratios L/P variés à analyser. Les différents ratios permettent de créer une droite de calibration (un modèle).

2. Préparation des différents échantillons à analyser

Tableau 1- Echantillons à analyser

Echantillon (% feuilles)	Nom attribué	Poids total	Nombre de spectres recueillis via la sonde du Matrix-F
T (0)	Tige	50g	10
F (100)	Feuille	_*	10
75F/25T (75)	T1/F3	20g	10
50F/50T (50)	T2/F2	20g	10
25F/75T (25)	T3/F1	20g	10
Brut (à déterminer)	Brut	20g	10

*Le poids de l'échantillon de feuilles « F » était approximatif car, pour les analyses, il suffisait que les feuilles recouvrent entièrement le tapis et 50g de feuilles représentait un volume trop important.

3. Analyses en 2 phases



Figure 4 - Persil sous caméra NIR hyperspectrale



Figure 5 - Sonde Q412

Pour chaque passage sous la caméra et sous la sonde NIR, chaque échantillon est disposé de la même manière sur le tapis roulant. Afin d'obtenir assez de spectres chaque échantillon est passé sur le tapis de 3 à 4 fois de suite (Remplacement aléatoire de l'échantillon sur le tapis avant chaque nouveau passage sous la sonde et la caméra NIR).

3.3. Résultats obtenus

Spectrométrie proche infrarouge

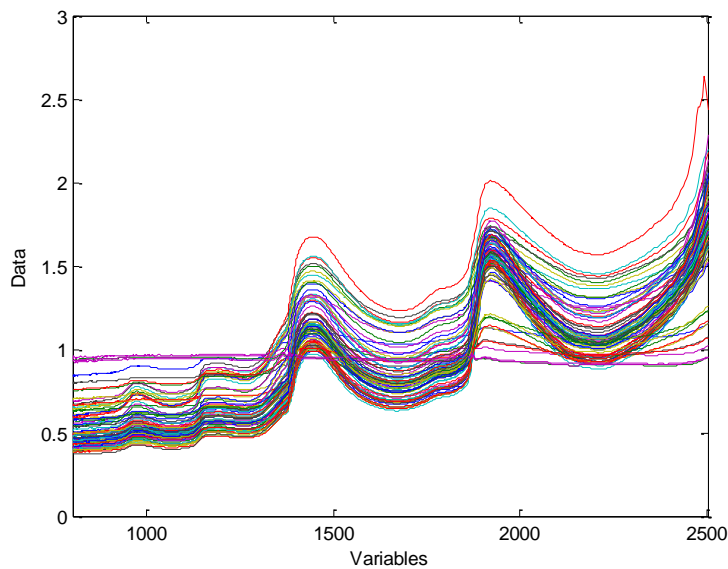


Figure 6 - Totalité des spectres

Quelques spectres (1, 9, 12, 59, 75, 82) ne sont pas retenus car le tapis n'a pas été discriminé. D'où la nécessité d'avoir une épaisseur de persil suffisante ou un contenant inorganique (inox par exemple).

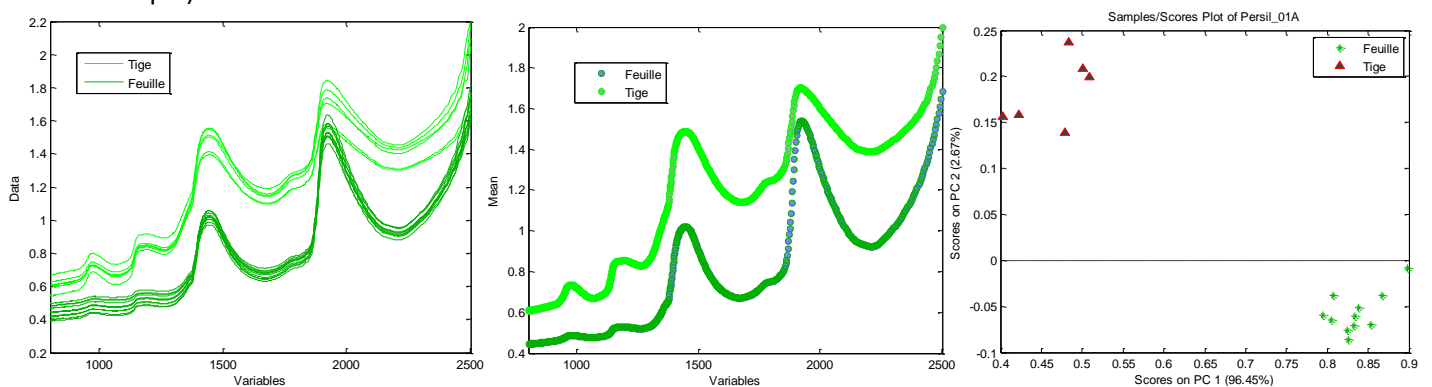


Figure 7 – Spectres de feuilles et de tiges pures/Normalisation de ces spectres/Discrimination graphique des feuilles et tiges par ACP

Les spectres retenus pour les feuilles et les tiges « pures » sont bien différents. Il y a une très bonne discrimination entre les feuilles et les tiges. Nous pouvons remarquer une différence de concentration en eau (970, 1450 et 1920 nm) et la teneur en composés lipidiques (1150, 1760 nm). Les tiges contiennent plus d'eau que les feuilles. Cette distinction laisse penser que des pourcentages différents de feuilles et de tiges pourraient être mis en évidence par spectrométrie PIR.

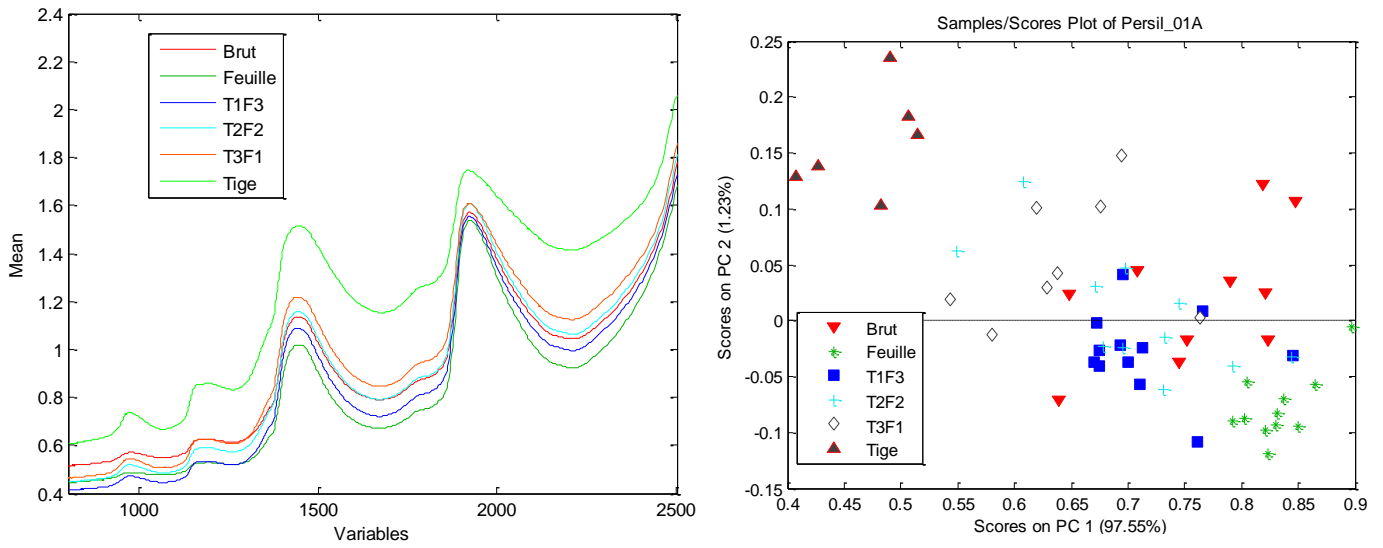


Figure 8 - Normalisation des spectres de chaque type d'échantillon/ACP

Les spectres des mélanges se situent entre ceux des tiges et des feuilles pures. Cela montre bien que les tiges et les feuilles sont discriminées et que les autres spectres obtenus sont issus d'un mélange de feuilles et de tiges.

3.4. Modèles développés sur base des échantillons de mélanges (0%, 25%, 50%, 75% et 100% feuille) et validé sur l'échantillon brute

Calibration/Validation croisée

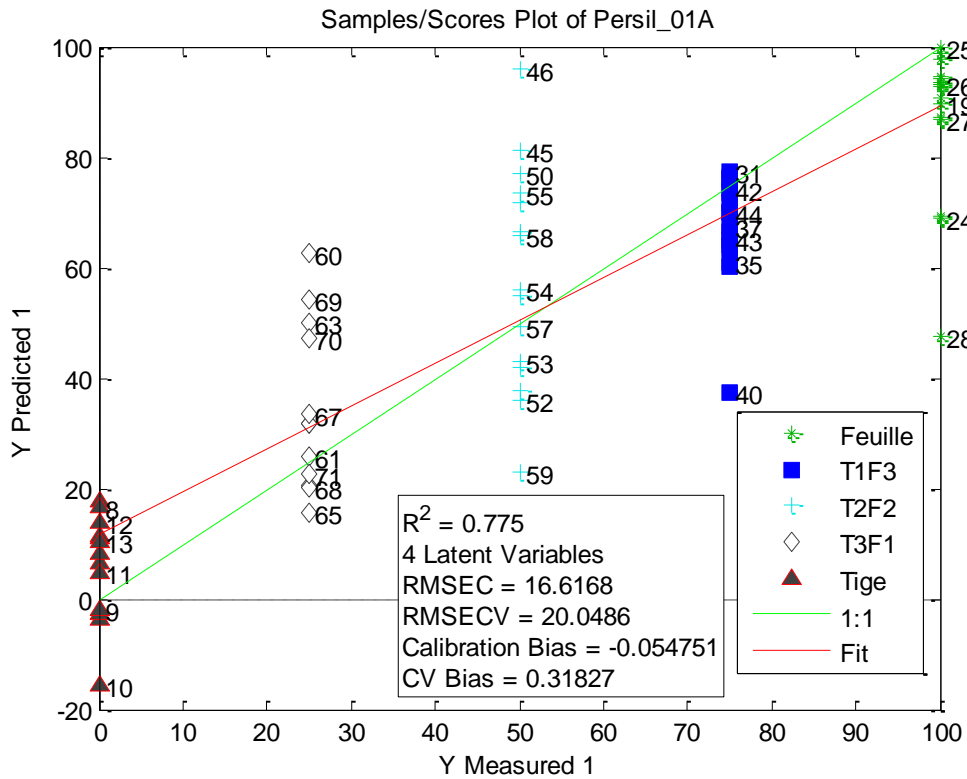


Figure 9 - Modèle 1

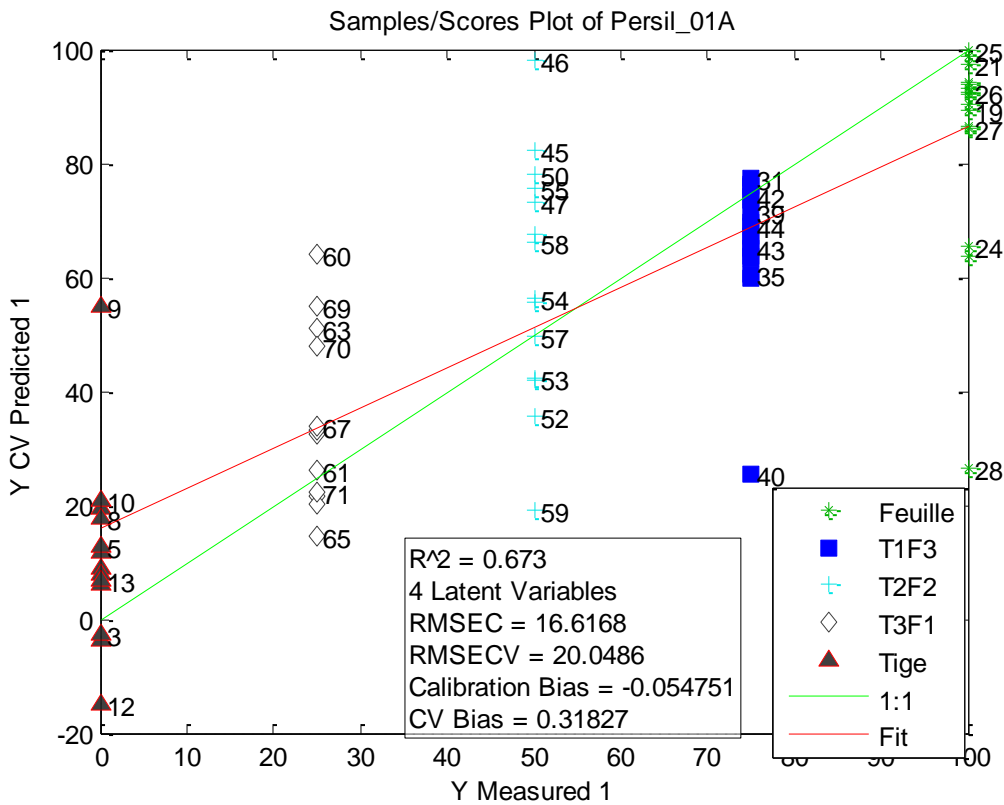


Figure 10 - Modèle 2

Deux modèles ont été créés avec la totalité des spectres. Leur fiabilité est similaire. Le premier modèle a un R² supérieur (0,775) au second (0,673).

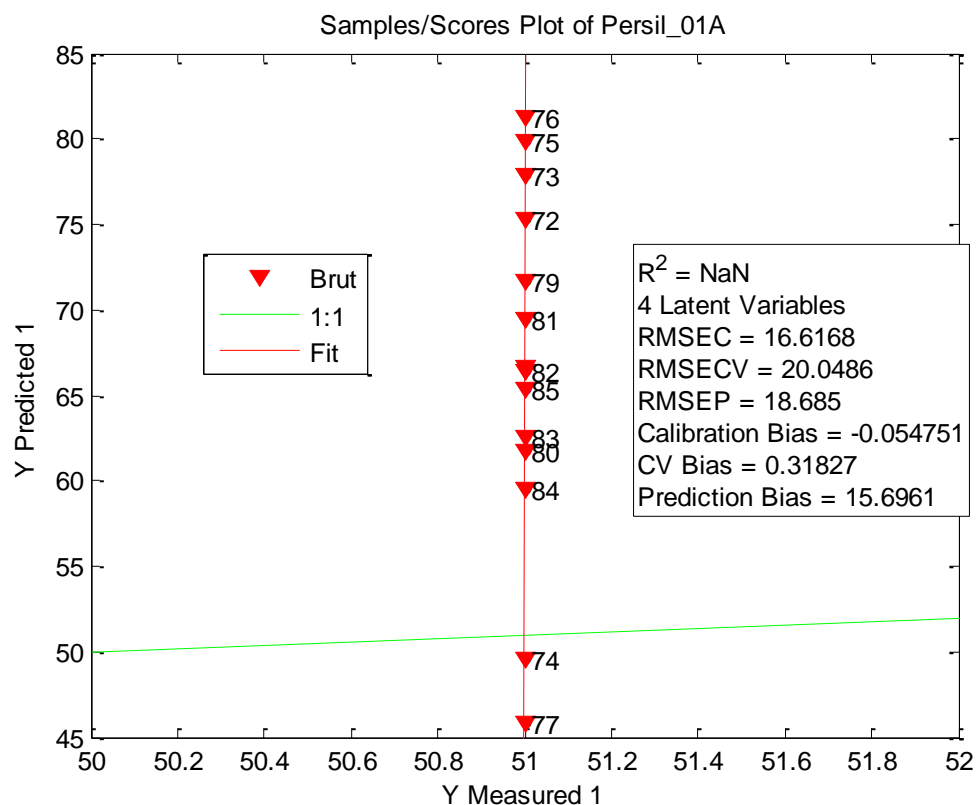


Figure 11 – Prédiction du ratio LP de l'échantillon brut

Tableau 2 - Analyse des modèles 1 et 2

Ratio L/P réelle de l'échantillon	Ratio L/P prédit par le modèle	Moyenne	67
72	75	Ecart-type	11
73	78	Minimum	46
74	50	Maximum	81
75	80		
76	81		
77	46		
78	67		
79	72		
80	62		
81	70		
82	66		
83	63		
84	60		
85	65		

➔ 95% des échantillons sont compris entre (67-22 et 67+22)

Calibration en enlevant la dernière prise de mesures (sur 3-4 séries de 64 scans par échantillon) qui, souvent, incluait le tapis

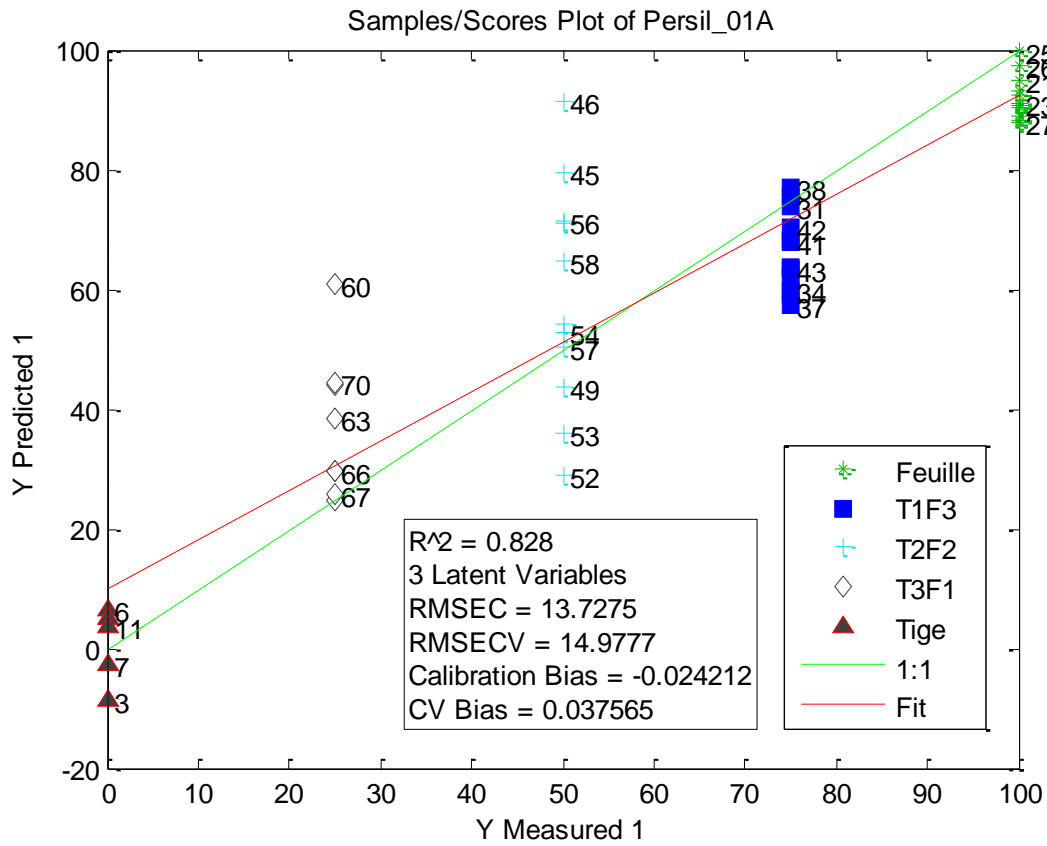


Figure 12 - Modèle 3

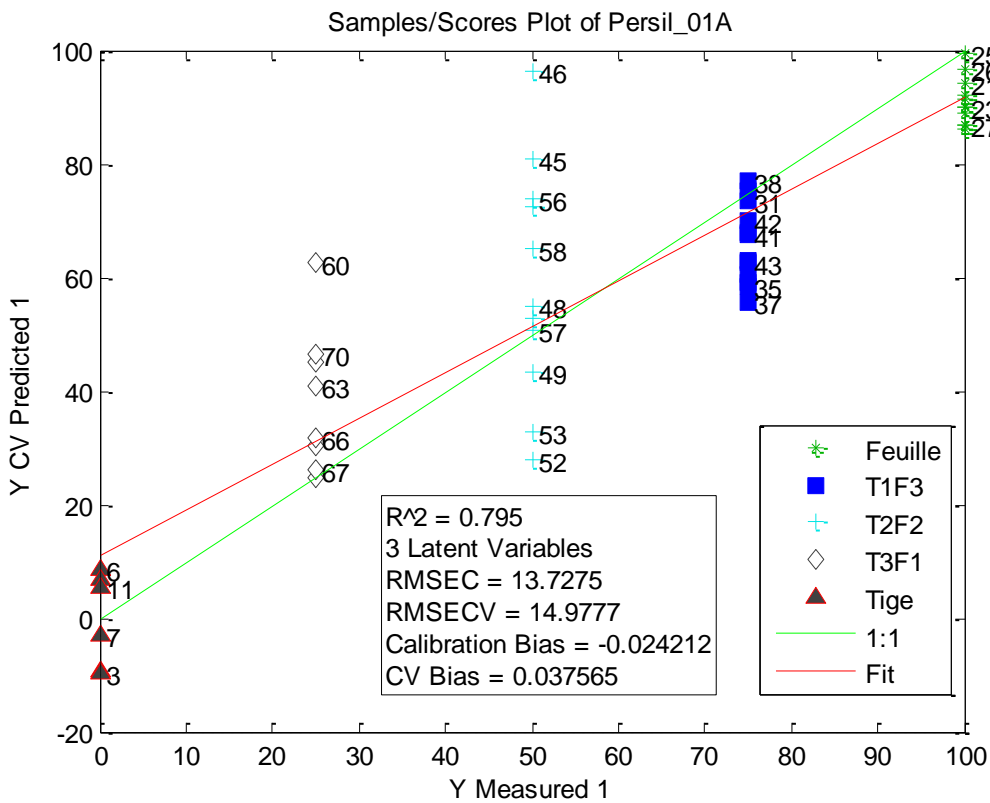


Figure 13 - Modèle 4

Nous obtenons des R^2 de 0.828 et de 0.795.

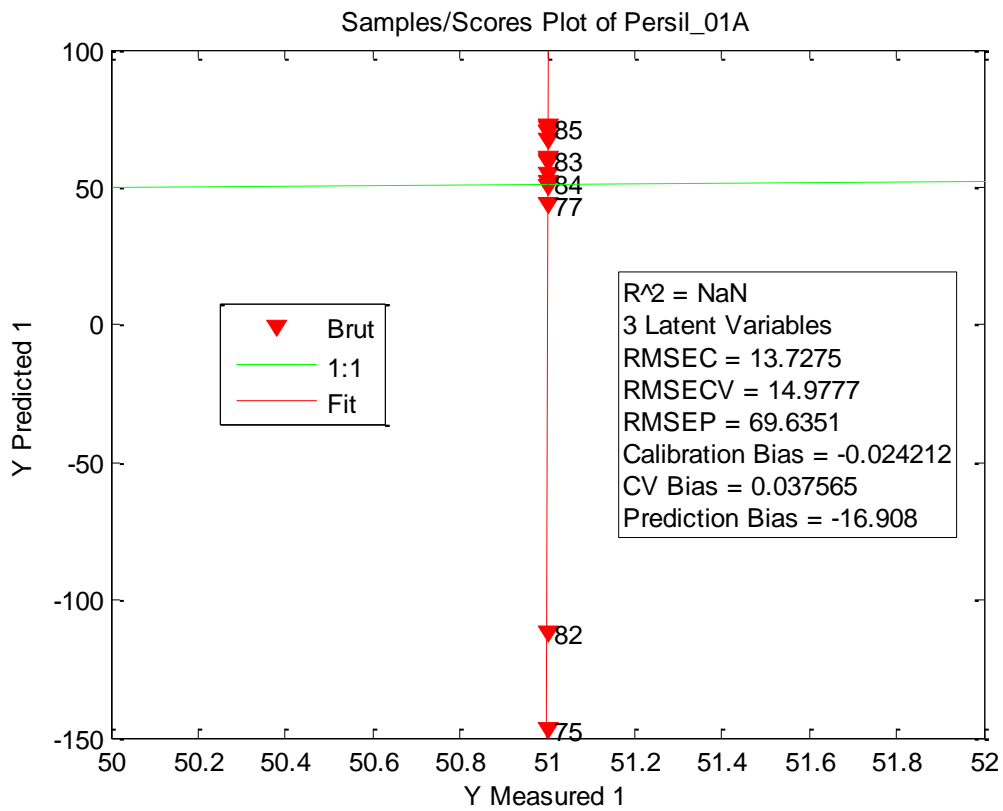


Figure 14 - Prédiction du ratio LP de l'échantillon brut

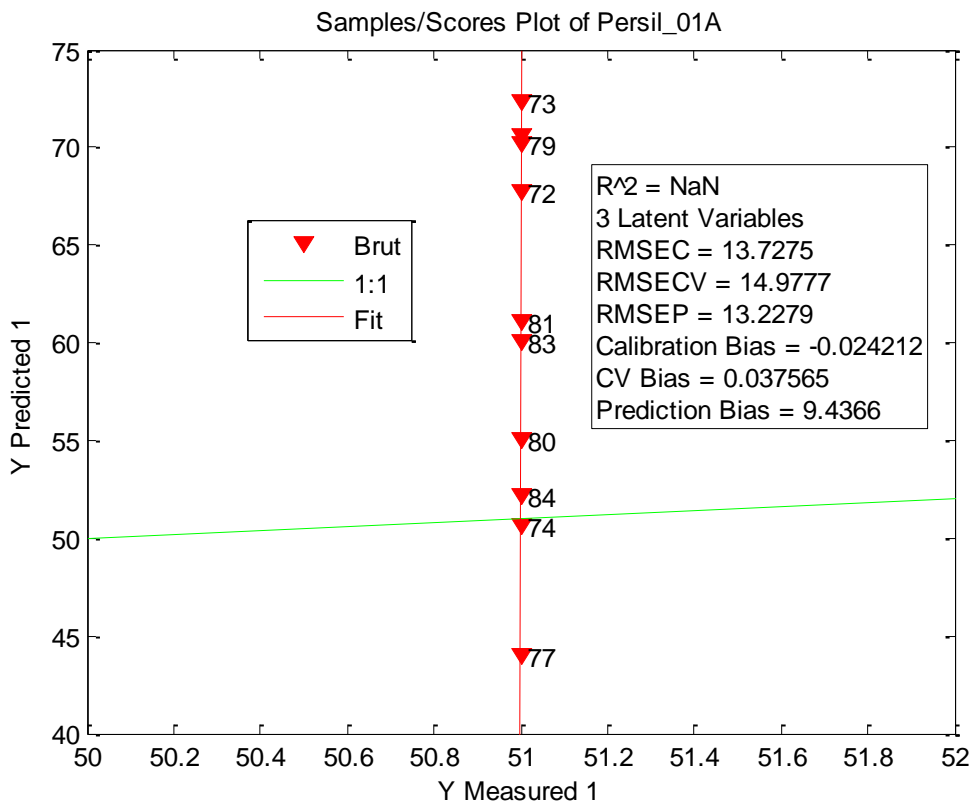


Figure 15 - Prédiction du ratio LP de l'échantillon brut sans les mesures sur les fins d'échantillon

Tableau 3 - Analyse des modèles 3 et 4

Ratio L/P réelle de l'échantillon	Ratio L/P prédit par le modèle	Remarques	Moyenne	60
72	68		Ecart-type	10
73	72		Minimum	44
74	51		Maximum	72
75	-147	fin d'échantillon		
76	71			
77	44			
78	59	fin d'échantillon		
79	70			
80	55			
81	61			
82	-112	fin d'échantillon		
83	60			
84	52			
85	72	fin d'échantillon		

➔ 95% des échantillons sont compris entre (60-20 et 60+20)

Imagerie proche infra-rouge

L'analyse des images donne de bons résultats. Néanmoins, ils peuvent être faussés par la position de l'échantillon (si les tiges sont cachées sous les feuilles, par exemple, Cf. Annexe 3). Avec un placement homogène d'un échantillon effeuillé (Cf. Annexe 4), on obtient une bonne prédiction du ratio L/P en imagerie : A partir de l'image hyper spectrale (« noir et blanc ») un modèle peut être créé.

Avant ma venue, Monsieur Kayoka (spécialisé dans l'imagerie PIR) a analysé du persil plat frais par imagerie. Il avait ainsi créé des modèles de prédiction. Avec les images hyper spectrales obtenues avec les « échantillons *Darégal* », il a appliqué le modèle de prédiction trouvé. Après quelques ajustements et améliorations du modèle, il pouvait prédire le ratio L/P de l'échantillon analysé. Sur une image RVB on peut bien distinguer le tapis, les tiges et les feuilles après création d'un modèle fiable. (Cf. Annexe 3).

3.5. Conclusions et perspectives

L'ACP et la PLS montrent une plus grande variabilité des échantillons de mélange et une plus grande homogénéité pour les échantillons purs (de tiges ou de feuilles).

Dans le meilleur des cas, la valeur prédite de l'échantillon brute varie entre 44% et 72% de feuilles pour une valeur de référence de 51% de feuilles et pour le même échantillon.

Cette étude montre que la variabilité intra échantillon est assez importante. Cette variabilité est la variabilité attendue entre différents échantillons au vu de l'incertitude de la méthode de référence. Il faudrait améliorer la précision de la méthode de référence et créer des modèles avec un nombre important d'échantillons, avec des rapports feuilles/tiges différents.

A noter que les mélanges de feuilles et de tiges ont été faits manuellement. Les modèles créés ne représentent donc pas tout à fait la réalité. De plus, les échantillons dans le modèle ne sont pas assez nombreux afin de donner de vraies conclusions.

Cette étude de faisabilité montre tout de même que l'utilisation de la spectrométrie proche infra-rouge est possible afin de déterminer le ratio L/P du persil plat frais.

4. Expérimentation *Darégal*

4.1. Méthode de référence de mesure du rapport Limbes/pétioles

L'agrégage manuel permet de définir le pourcentage de feuilles et de tiges présent sur le lot de matière première réceptionnée. Une analyse par spectrométrie proche infra-rouge s'avère intéressante pour analyser beaucoup d'échantillons rapidement et sans les détruire.

4.2. Spectromètre proche infrarouge

Durant cette expérimentation, nous allons utiliser un spectromètre Matrix-F avec une sonde Q412 pour l'analyse sans contact des corps solides:

- 2 sources de lumière NIR éclairent l'échantillon (10mm de diamètre)
- De la mono fibre (fibre optique) est utilisée afin de renvoyer la lumière vers le spectrophotomètre



Figure 16 - Spectroscopie Matrix-F

4.3. Protocole expérimental (phase de collecte des données)

Matériel

- Echantillons variés de persil plat frais (150 à 500 échantillons en tout) : 100g chacun. (Certains pour la modélisation, d'autres pour tester les modèles créés)
- Spectromètre Matrix-F de chez *Bruker*
- Coupelle en inox (30cm de diamètre et 8cm de profondeur) en rotation (moteur) + séparateur (réducteur)
- Sonde Q412 d'émission Bruker lumière proche infrarouge + son support
- Logiciel d'analyse (OPUS 7.5) + Ordinateur de pilotage du spectromètre



Figure 17 - Dispositif complet

Réglages et paramétrage du spectromètre

- Vitesse de rotation de la coupelle = 29.80 Hz (unité du moteur)
- Hauteur entre lumière de la sonde et fond de la coupelle (distance optique) : ~14 cm
- **Chaque mesure : 1min : environ 180 scan : environ 2 tours à V = 29.80 Hz**

Protocole de mesure

- **Pour chaque échantillon de 100g :**
 - Enregistrement de 3 spectres de l'échantillon intact (Cf. Annexe 2)
 - Effeuilage manuel : valeur L/P
 - 3 spectres de l'échantillon effeuillé (pas pour tous les échantillons)
- 3 mesures d'un même échantillon : 3 remplissages différents (changement de place de l'échantillon 3 fois dans la coupelle en inox)
- Normaliser les spectres d'un même échantillon et observer
- Effeuille l'échantillon à la main
- Noter le rapport L/P obtenu manuellement pour chaque échantillon

4.4. Etablissement de l'Equation de prédiction (phase de modélisation)

Un modèle peut être créé si un lien existe entre les ratios L/P obtenus manuellement et les différents spectres correspondants. L'établissement d'un modèle nécessite une analyse préalable d'environ 300 échantillons. Certains spectres sont utilisés pour la calibration/validation et d'autres pour l'analyse des modèles.

Pour la régression partielle des moindres carrés (PLS), la sélection du nombre de facteurs optimum a son importance pour la qualité du modèle. Plus le nombre de facteurs est faible plus le modèle est fiable. Ce nombre correspond au Rang pris en compte dans le modèle.

La calibration univariée est souvent inefficace. Il faut utiliser les calibrations multivariées comme :

- La régression linéaire multiple (ACM)
- La régression en composantes principales (ACP)

- La régression partielle des moindres carrés (PLS)

RMSEV (Erreur quadratique à la moyenne) doit être la plus faible possible (la plus proche possible de 0) après comparaison des données brutes et des valeurs issues des analyses -> Bonne qualité et fiabilité du modèle si elle est faible. Le RPD devrait être supérieur à 3 afin de valider directement une méthode. (RPD = RMSECV/SD). Pour un solide, le R^2 doit idéalement être supérieur à 90% mais dès 70% le modèle gagne en fiabilité.

4.5. Test de l'équation (Validation de la méthode)

On peut évaluer la calibration en utilisant un autre jeu de données externe au modèle (validation externe). On ne peut pas prendre les échantillons qui font déjà partie de l'étalonnage.

1. Prélever un ensemble d'échantillons différents des précédant
2. Placer chaque échantillon intact, brut de récolte, dans la coupelle en rotation
3. Analyser au spectromètre ces différents échantillons
4. Enregistrer tous les spectres obtenus
5. Utiliser l'équation de prédiction pour calculer les rapports L/P
6. Vérifier l'analyse et que l'équation permet une bonne estimation du rapport L/P réel (obtenu grâce à la méthode de référence). Il faut voir si en travaillant sur des moyennes on peut améliorer les performances des équations.

5. Résultats de l'expérimentation

5.1. Premiers résultats (au 19.07.2016)

Après plus d'un mois de collecte de données spectrales (voir protocole de mesure) nous avons obtenu les spectres de 234 échantillons de persil plat frais. Chaque spectre est associé à une valeur L/P calculée grâce à la méthode de référence. Le tout est répertorié dans un tableau Excel en attendant de rentrer les données dans le logiciel OPUS afin de créer un modèle.

Une première modélisation donne une régression qui semble en accord avec la méthode de référence compte tenu de l'incertitude des données de référence. A ce stade de l'expérimentation, il n'y avait pas assez de données pour confirmer la tendance et améliorer la statistique du modèle. Il faut, peut-être, utiliser la valeur moyenne des trois mesures d'un même échantillon comme valeur de référence.

Prédiction / Vrai

Ratio LP

Rang: 8

Rec: 8

18 -

R²:
78.42

RMSECV:
3.11

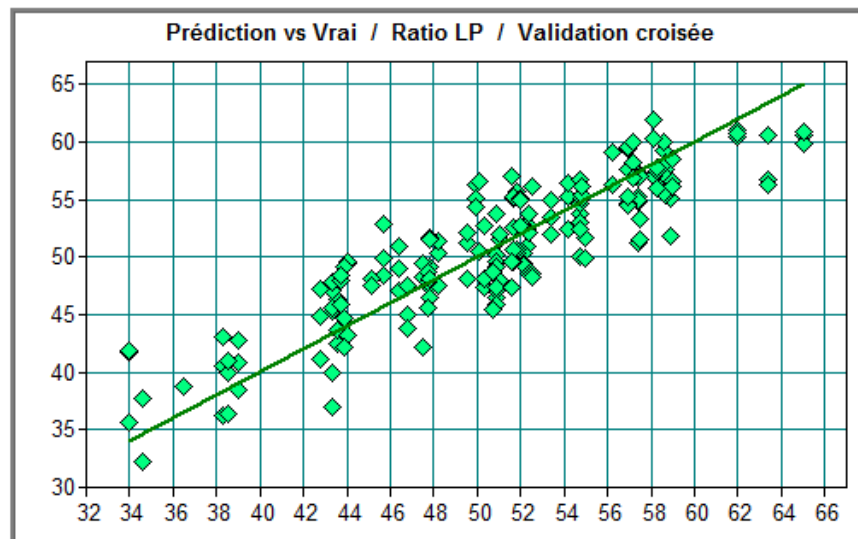
RPD:
2.15

Biais:
-2.42E-005

Ligne

Couleur

modélisation

Figure
Première

5.2. Création des modèles et analyse

La création d'un modèle de calcul fiable du ratio L/P semble possible. Néanmoins, l'incertitude ne sera jamais meilleure que celle de la méthode de référence.

Il est important que le RPD soit supérieur à 3 et que l'incertitude soit inférieure à la moitié de l'écart-type général pour lever l'incertitude sur la méthode de référence. Dans le cas contraire la méthode de référence est à revoir afin de l'améliorer. Dans cette expérimentation nous obtenons un écart-type moyen de 2,40. L'incertitude est aux alentours de 3 et est donc supérieur à la moitié de l'écart-type général. Le RDP quant à lui est inférieur à 3, situé majoritairement entre 2 et 2,5.

5.3. Méthode de référence

Mesures

En tout, 234 échantillons de 100g de persil plat frais ont été analysés au spectromètre puis effeuillés. Chaque échantillon est passé 3 fois au spectromètre avec un changement de position à chaque fois. Ces échantillons représentent 69 lots différents avec des coupes allant de la première à la troisième. Un à huit échantillons, de chaque lot, ont été mesurés. Ceci en fonction de la disponibilité du persil et en essayant d'avoir une moyenne plus représentative avec plus de 3 mesures d'un même lot.

Statistiques descriptives

Moyenne des Ecart-types de chaque lot	Ecart-type min	Ecart-type max	Moyenne totale	Ratio LP Min	Ratio LP Max
2,40	0,00	9,03	50,7	34	66,5

Moyenne Ecart type échant mesurés 2 fois	Moyenne Ecart type échant mesurés 3 fois	Moyenne Ecart type échant mesurés 4 fois	Moyenne Ecart type échant mesurés 5 fois	Moyenne Ecart type échant mesurés 6 fois
1,58	2,25	3,13	2,80	3,00

La différence entre chaque mesure d'un même lot est en moyenne de + ou - 6%. Cette incertitude sur la mesure va influencer la fiabilité du modèle. 95% des valeurs prédites par un modèle doivent se situer à + ou - 6% des valeurs réelles.

5.4. Comparaisons des modèles de mesure du ratio L/P par spectrométrie proche infrarouge.

La méthode PLS (Partial Least Square)

- Rang : [2 ; 8]
- R² : [51.04 ; 86.54]
- RMSECV : [1,9 ; 3.01]

Après observations de ces premiers critères, les méthodes les plus fiables sont :

Tableau 4 - Validation modèles créés

Modèle	Composante	Rang (méthode)	R ²	RMSECV	RPD
005.q2	Ratio LP	7	79,1	2,99	2,19
006.q2	Ratio LP	8	81,9	2,67	2,35
007.q2	Ratio LP	8	82,7	2,85	2,42
008.q2	Ratio LP	8	74,8	2,75	1,99
009.q2	Ratio LP	7	80,9	2,35	2,29
066.q2	Ratio LP	8	81,3	2,64	2,31
077.q2	Ratio LP	8	84,2	2,68	2,52
099.q2	Ratio LP	7	84,2	2,12	2,52
0012.q2	Ratio LP	6	86,5	1,9	2,74

Le rang de chaque méthode est majoritairement de 8. Ce rang peut paraître élevé mais cela semble normal compte tenu de l'incertitude de la méthode de référence et de la nature de l'échantillon (espèce végétale, vivante).

RMSE doit être le plus proche de 0. Il est juste en dessous de 3 pour la majorité des méthodes. Estimation de l'erreur = RMSE x2. Cette erreur est à comparer avec l'incertitude de la méthode de référence.

Les R² et RMSECV de ces méthodes sont proches. Il est donc difficile de savoir laquelle est plus fiable. Le nombre d'erreurs commises par le modèle et la marge d'erreur seront similaires au vu de ces premières statistiques. Si on se base sur ces chiffres, la fiabilité des méthodes est donc proche. Il n'y en a aucune qui se distingue clairement. Peut-être la 0012.q2. Ces méthodes sont donc à tester avec des lots d'échantillons d'analyse.

Statistique descriptive

- **Moyennes des écarts** entre les ratios L/P vrais et ceux prédits : [3,2 ; 5,3]
- **Ecart-types** de ces écarts : [2,4 ; 3,1]
- Moyennes des écarts entre ratios L/P moyens vrais et ceux prédits : [2,4 ; 4,9]
- Ecart-types de ces écarts : [1,6 ; 2,9]

Après observations de ces critères, les méthodes les plus fiables sont :

Tableau 5 - Statistique descriptive de l'analyse des modèles créés

Modèle	Moyenne des écarts entre ratios VRAIS et ratios PREDITS	Ecart-type des écarts entres ratios VRAIS et ratios PREDITS	MIN	MAX	Moyenne des écarts entre moyennes VRAIES et moyennes PREDITES	Ecart-type des écarts entres moyennes VRAIES et moyennes PREDITES	MIN	MAX
005.q2	3,2	2,4	0,0	10,2	2,5	1,6	0,0	5,4
006.q2	3,4	2,5	0,0	14,9	2,4	1,9	0,0	7,7
008.q2	3,4	2,4	0,0	15,6	2,5	1,6	0,2	6,3
066.q2	3,4	2,5	0,0	13,6	2,5	1,8	0,2	7,4
077.q2	3,5	2,5	0,0	18,4	2,6	1,6	0,1	7,4
088.q2	3,5	2,6	0,0	16,0	2,7	1,8	0,0	6,9

Les modèles restent proches avec des moyennes des écarts entre ratios LP vrais et ratios prédits restants, dans l'absolue, à plus ou moins 3,5%.

Statistique inférentielle

Calculs d'indicateurs d'écarts sur variables quantitatives (ratios L/P puis ratios L/P moyens par Lot) :

- **Somme des carrés des résidus** (SCR : *Sum of Squared Errors*) -> importance des grosses erreurs
- **Carré moyen des erreurs** ou **erreur quadratique moyenne** (MSE : *Mean Square Error*, MCE : moyenne des carrés des erreurs) -> Valeur à minimiser (variance résiduelle)
- **L'erreur-type** (RMSE) : RACINE(MCE)
- **L'erreur absolue moyenne** (MAE : *Mean Absolute Error*)
- **L'erreur absolue moyenne (%)** : moyenne des écarts en valeurs absolues par rapport aux valeurs observées (MAPE : *Mean Absolute Percentage Error*)

Intervalles pour les rapports L/P individuels :

- SCR : [4622 ; 8609]
- MCE : [15,99 ; 23,7]
- RMSE : [4 ; 4,9]
- MAE : [3,22 ; 3,99]
- MAPE (%) : [6,58 ; 8,61]

Après observations de ces critères sur les ratios L/P individuels, les méthodes les plus fiables sont :

Tableau 6 - Statistique inférentielle de l'analyse des modèles créés

Modèle	Nombre d'observations	SCR	MCE	RMSE	MAE	MAPE (%)
005.q2	291	4622	15,99	4	3,22	6,58
006.q2	429	7613	17,83	4,22	3,36	6,64
008.q2	426	7200	16,98	4,12	3,37	6,69
066.q2	402	7323	18,31	4,28	3,43	6,78

L'incertitude des ratios L/P calculé reste à plus ou moins 6%. Dans l'absolue il y a une différence de plus ou moins 3% chaque ratio L/P réel et chaque ratio L/P calculé.

Le meilleur modèle semble être le 005.q2 mais il est basé sur moins d'échantillons analysés. Pour les 3 autres modèles les résultats sont similaires. Le modèle 008.q2 semble faire moins d'erreurs.

Intervalles pour les rapports L/P moyens par lot :

- SCR : [265 ; 538]
- MCE : [9,39 ; 17,93]
- RMSE : [3,06 ; 4,23]
- MAE : [2,41 ; 3,58]
- MAPE (%) : [4,7 ; 7,1]

Après observations de ces critères sur les ratios L/P moyens, les méthodes les plus fiables sont :

Tableau 7 - Statistique inférentielle sur l'analyse des moyennes des ratios L/P

Modèle	Nombre d'observations	SCR	MCE	RMSE	MAE	MAPE (%)
005.q2	30	265	9,46	3,08	2,52	5,11
006.q2	37	340	9,71	3,12	2,41	4,7
008.q2	36	319	9,39	3,06	2,55	4,92
066.q2	36	348	10,23	3,2	2,53	4,96
077.q2	36	342	10,07	3,17	2,64	5,21

En ce qui concerne la prédiction des moyennes des lots, les résultats de ces méthodes sont également proches. Le modèle 008.q2 semble faire moins d'erreurs.

5.5. Graphiques obtenus pour quelques modèles

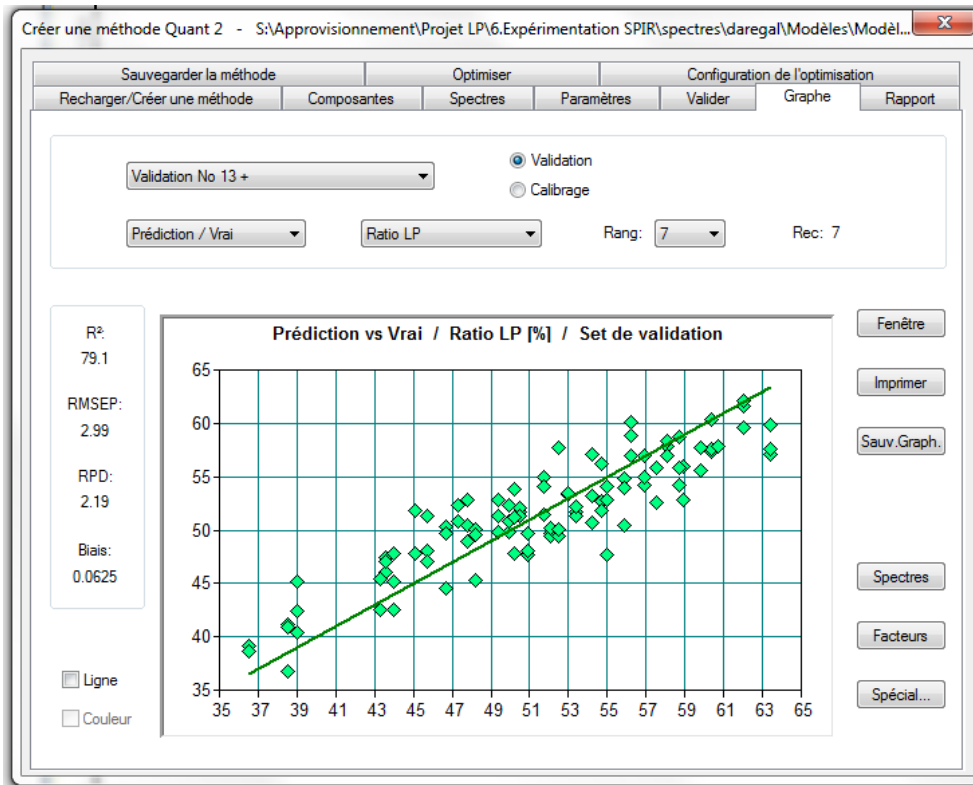


Figure 19- 005.q2

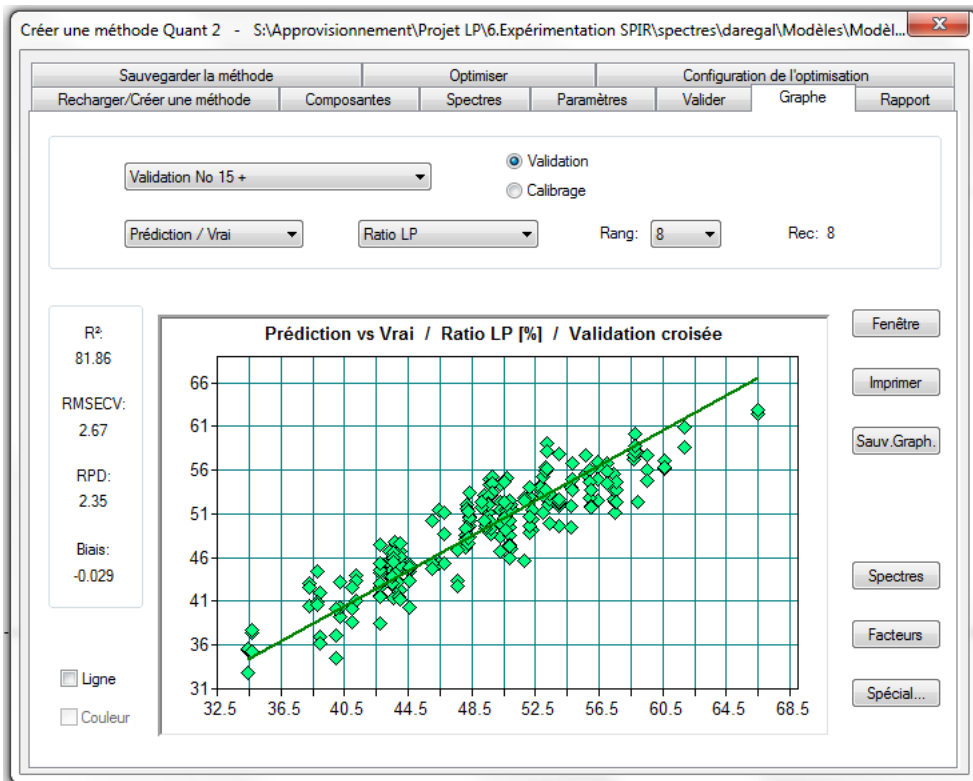


Figure 20 - 006.q2



Figure 21 - 008.q2

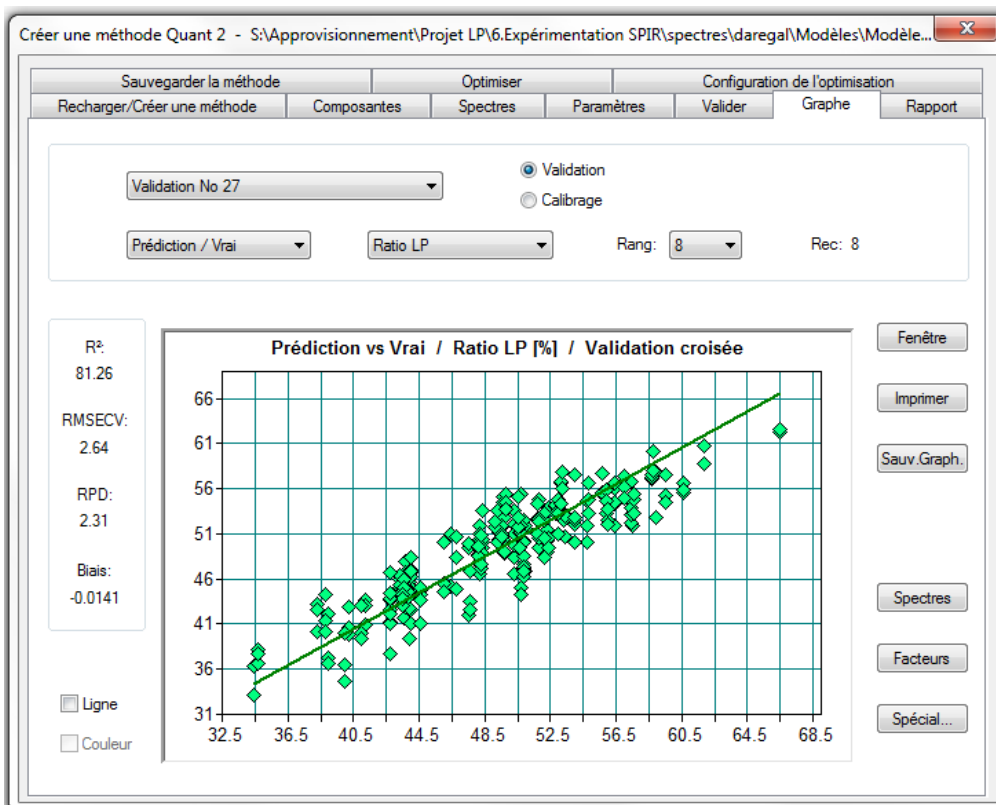


Figure 22 - 066.q2

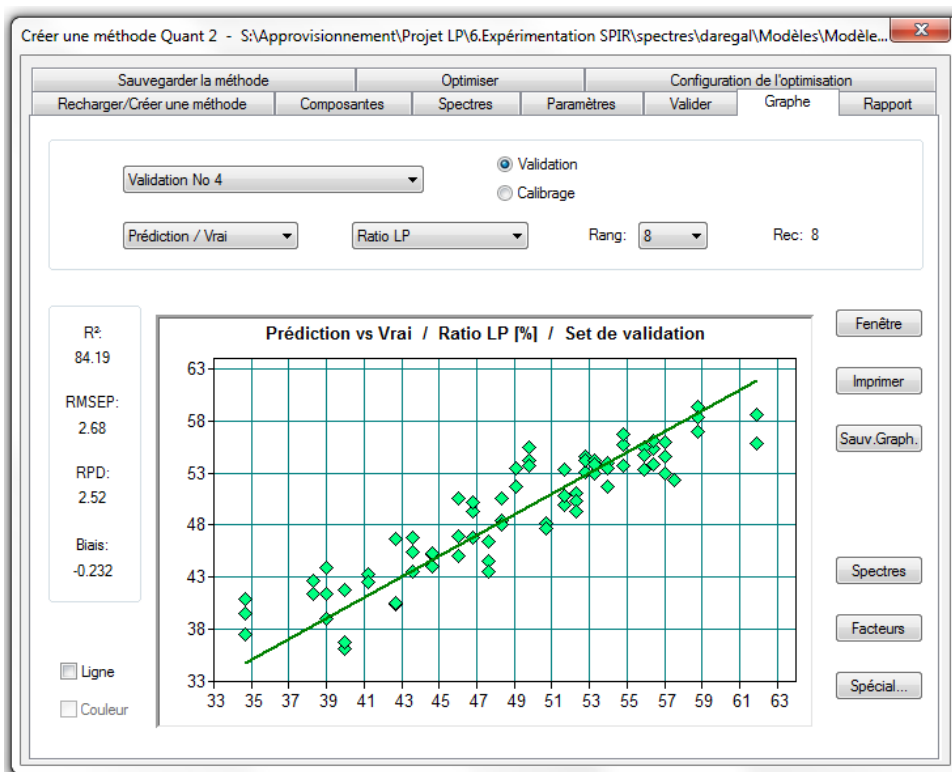


Figure 23 - 077.q2

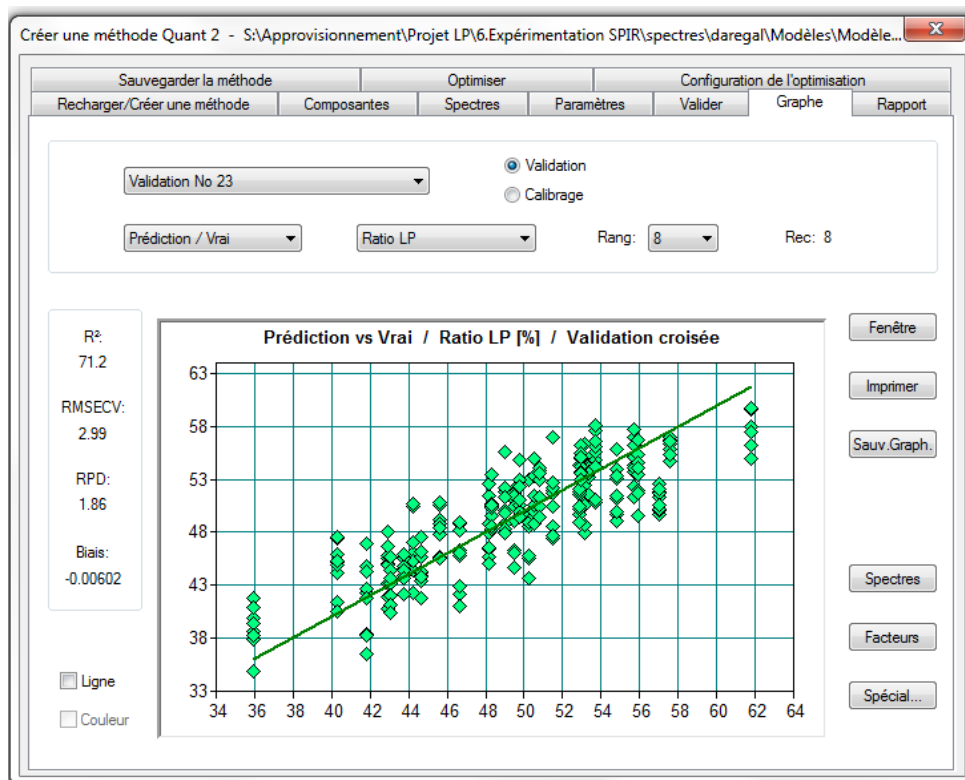


Figure 24 - 088.q2

5.6. Comparaisons valeurs vraies/valeurs prédites

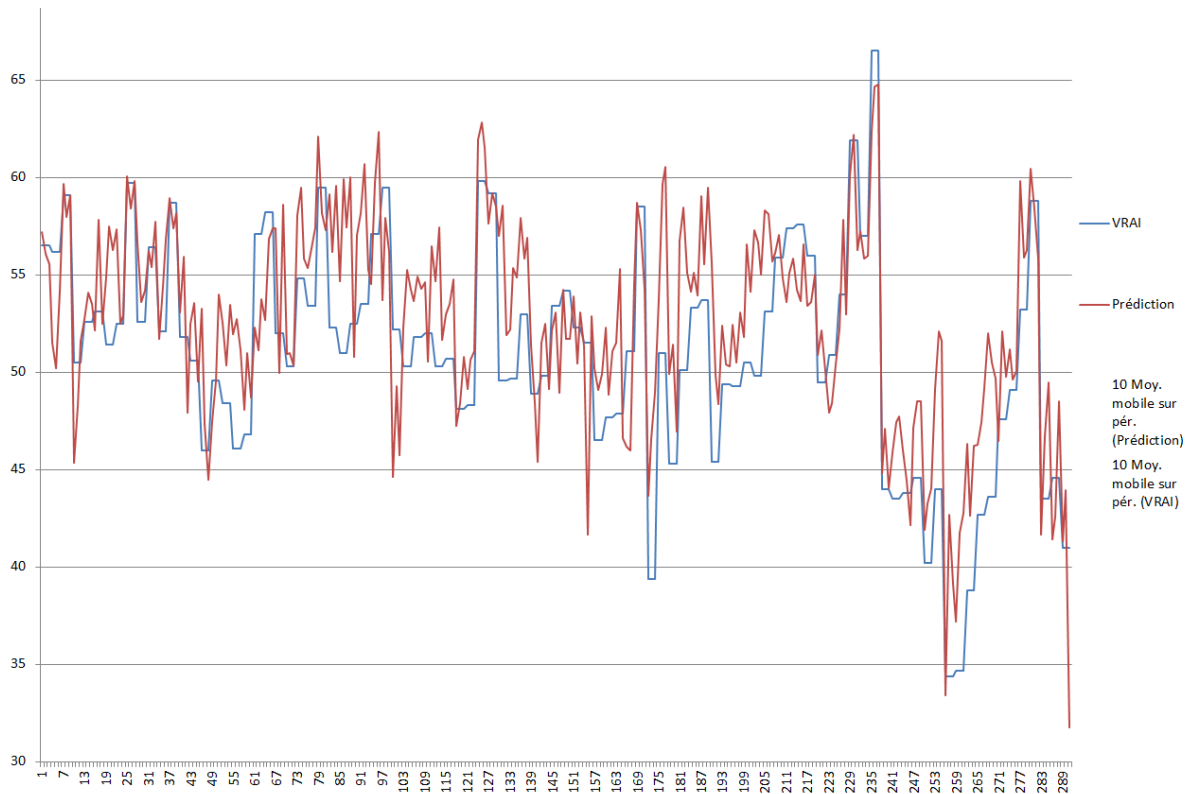


Figure 25 - Prédiction des valeurs individuelles avec le modèle 005.q2

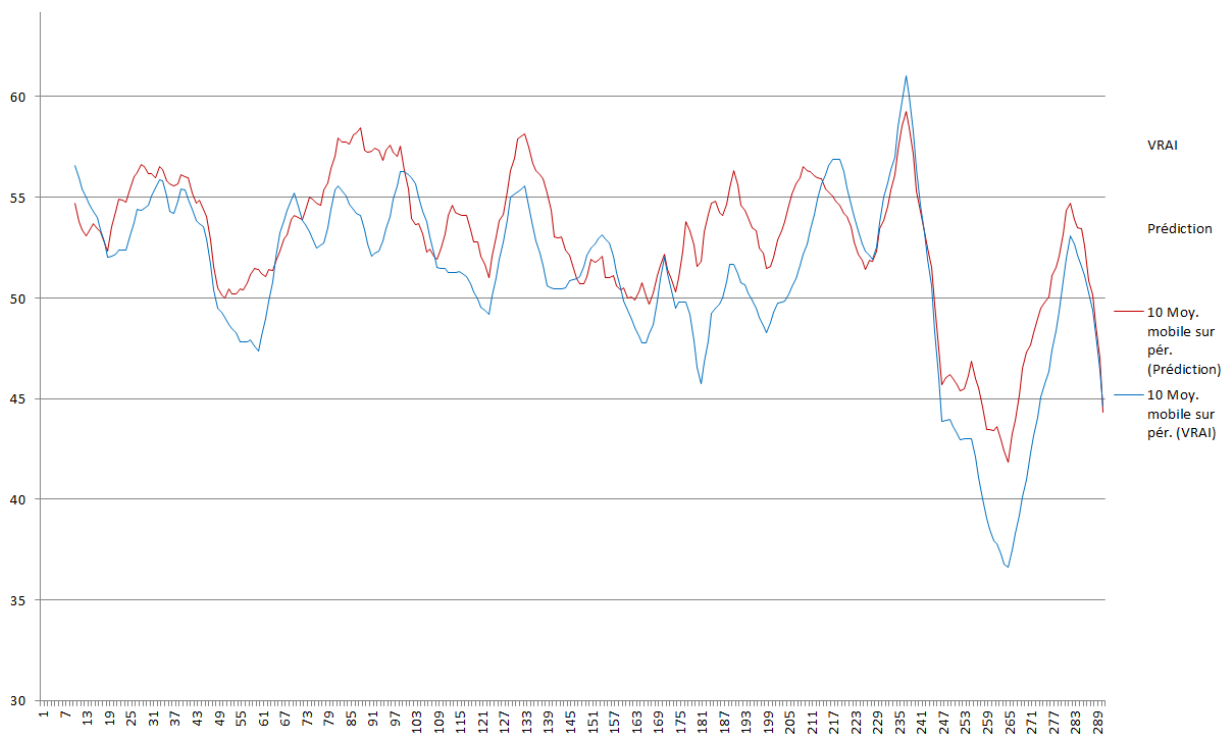


Figure 26 - Prédiction des moyennes avec le modèle 005.q2

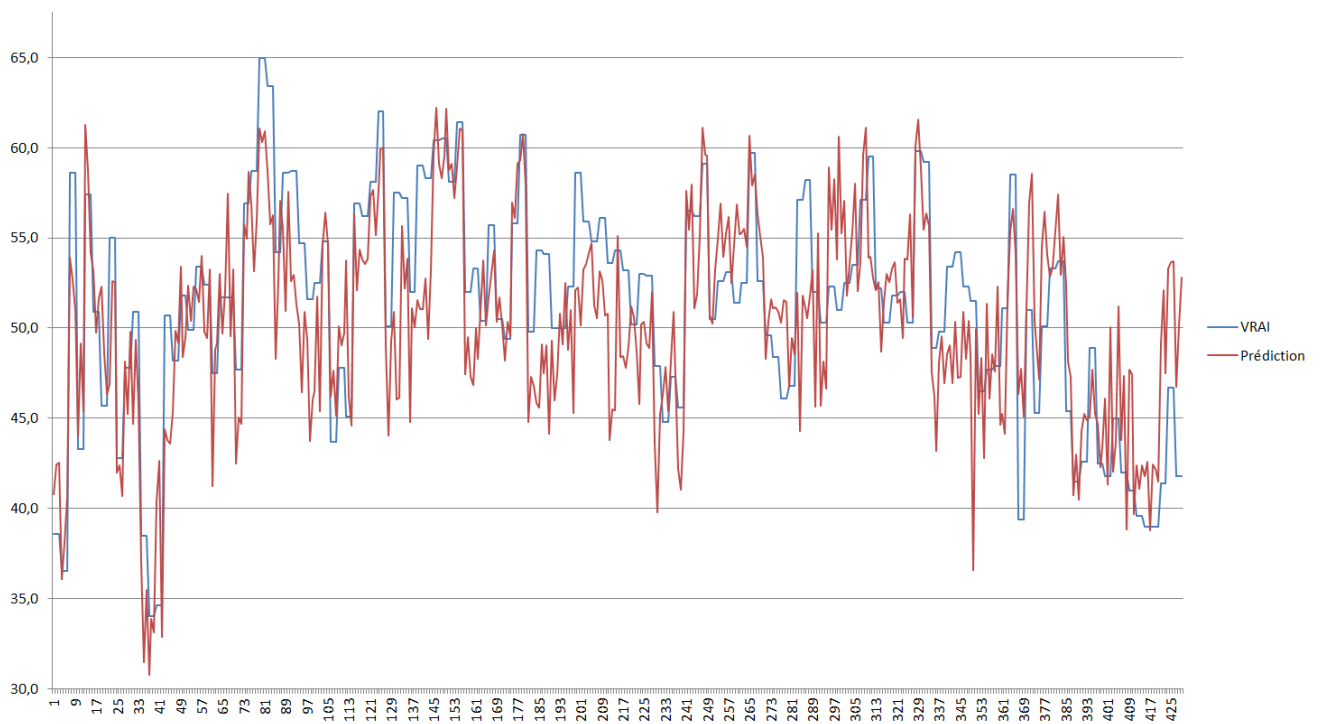


Figure 27 - Prédiction des valeurs individuelles avec le modèle 006.q2

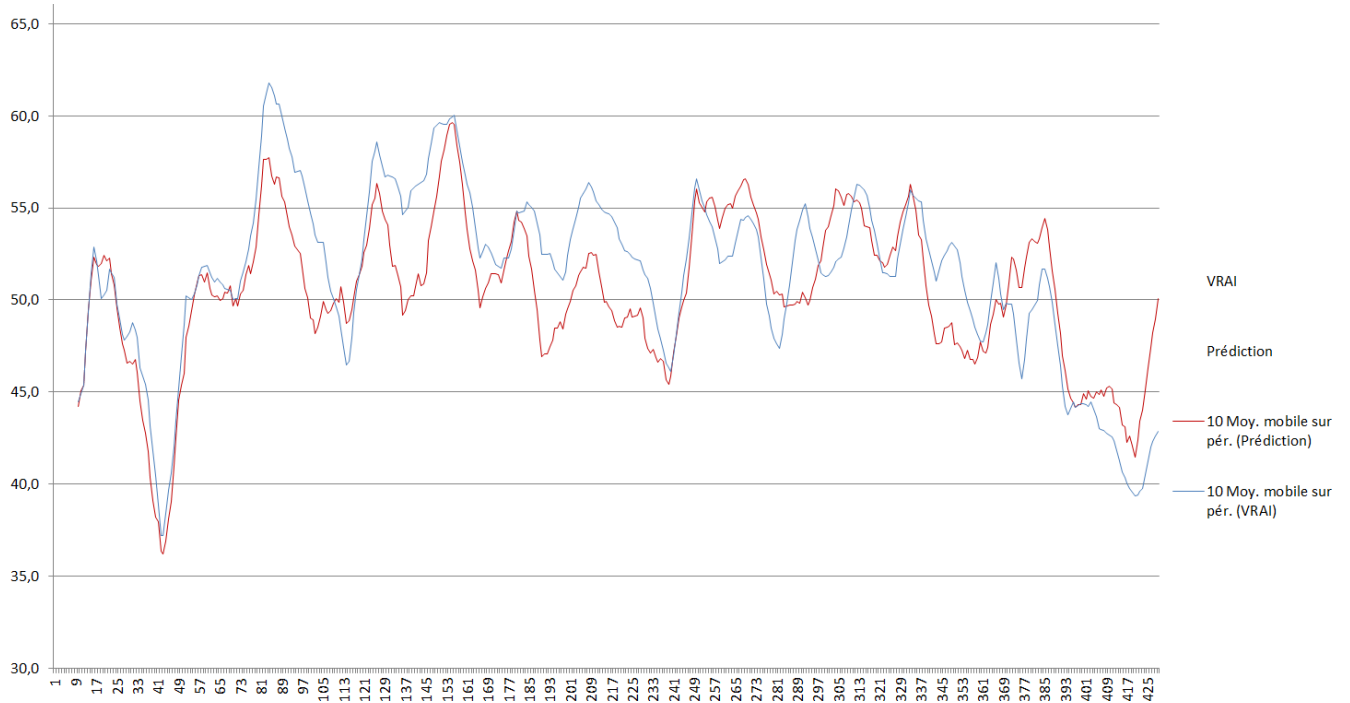


Figure 28 - Prédiction des moyennes avec le modèle 006.q2

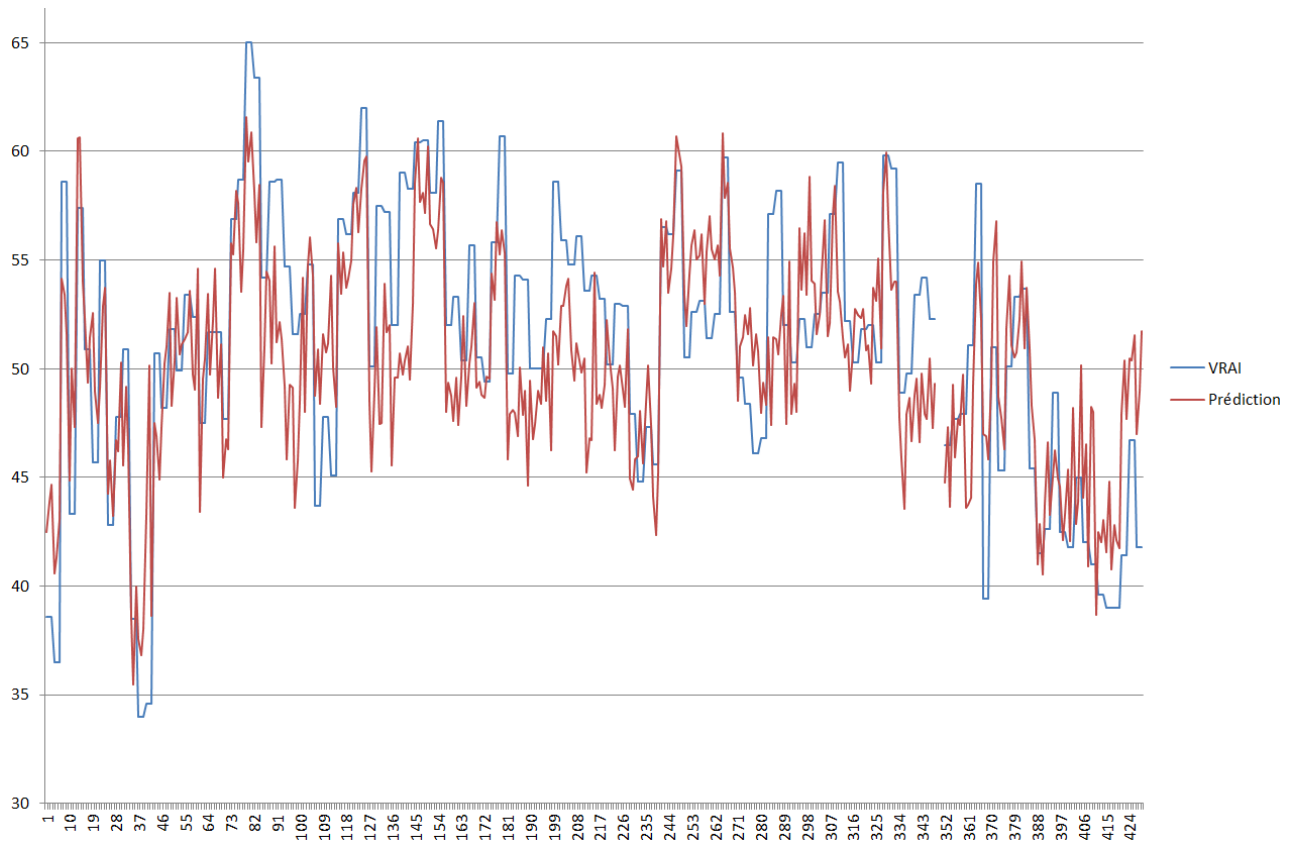


Figure 29 - Prédiction des valeurs individuelles avec le modèle 008.q2

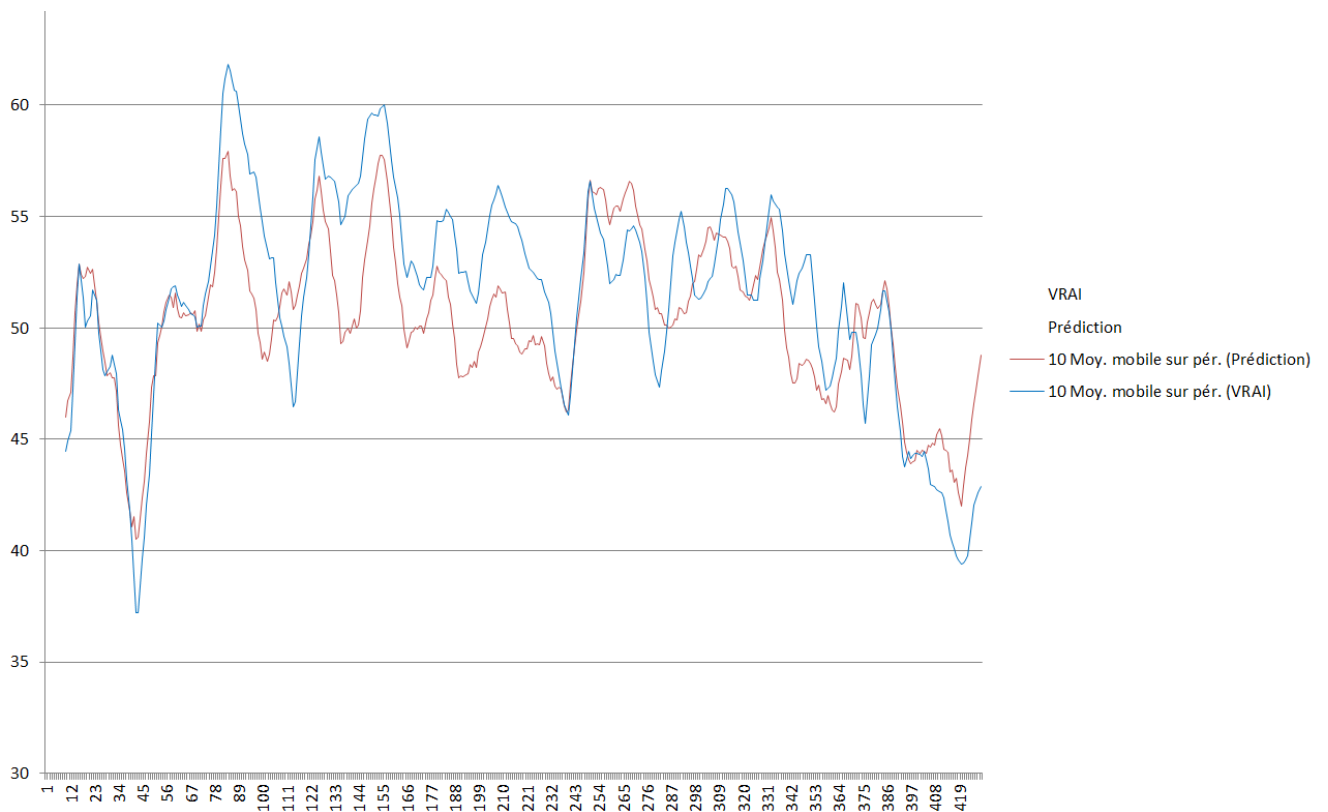


Figure 30 - Prédiction des moyennes avec le modèle 008.q2

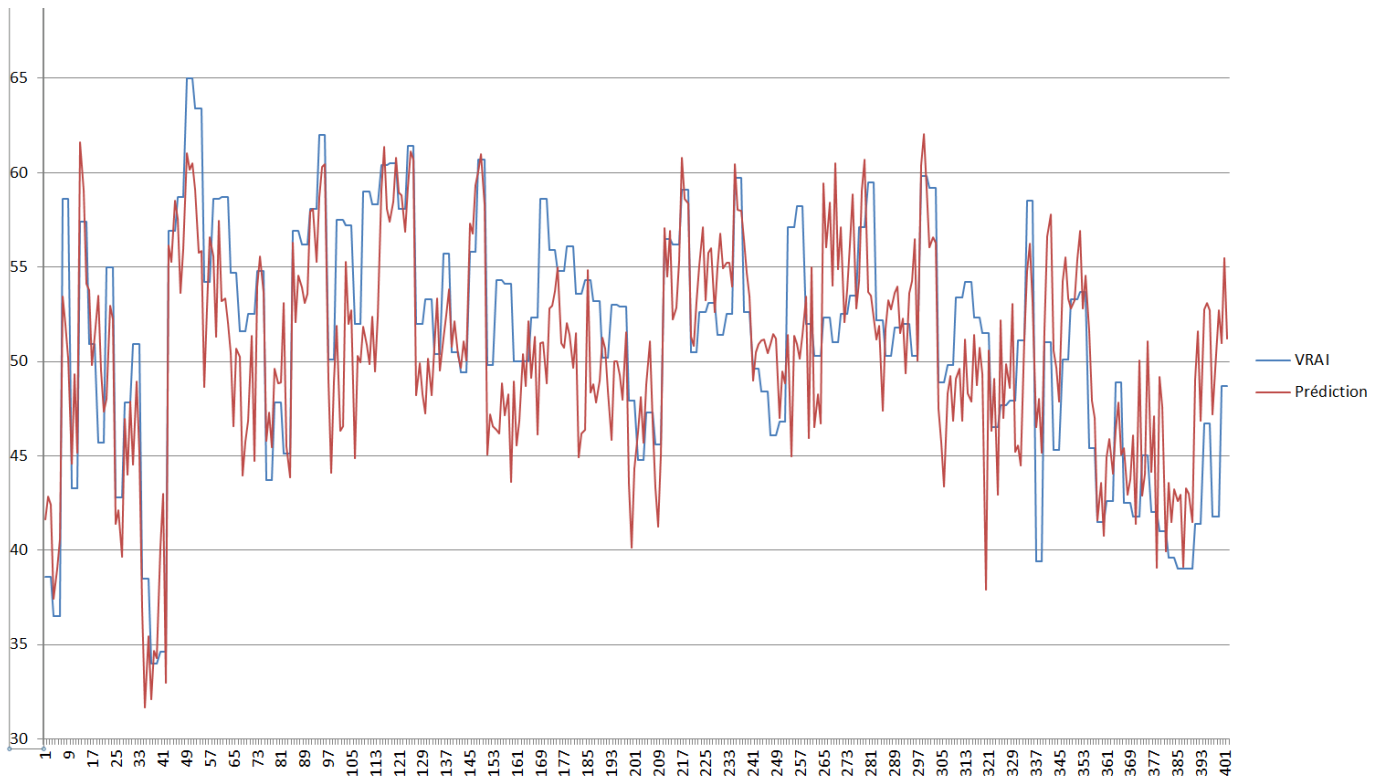


Figure 31 - Prédiction des valeurs individuelles avec le modèle 066.q2

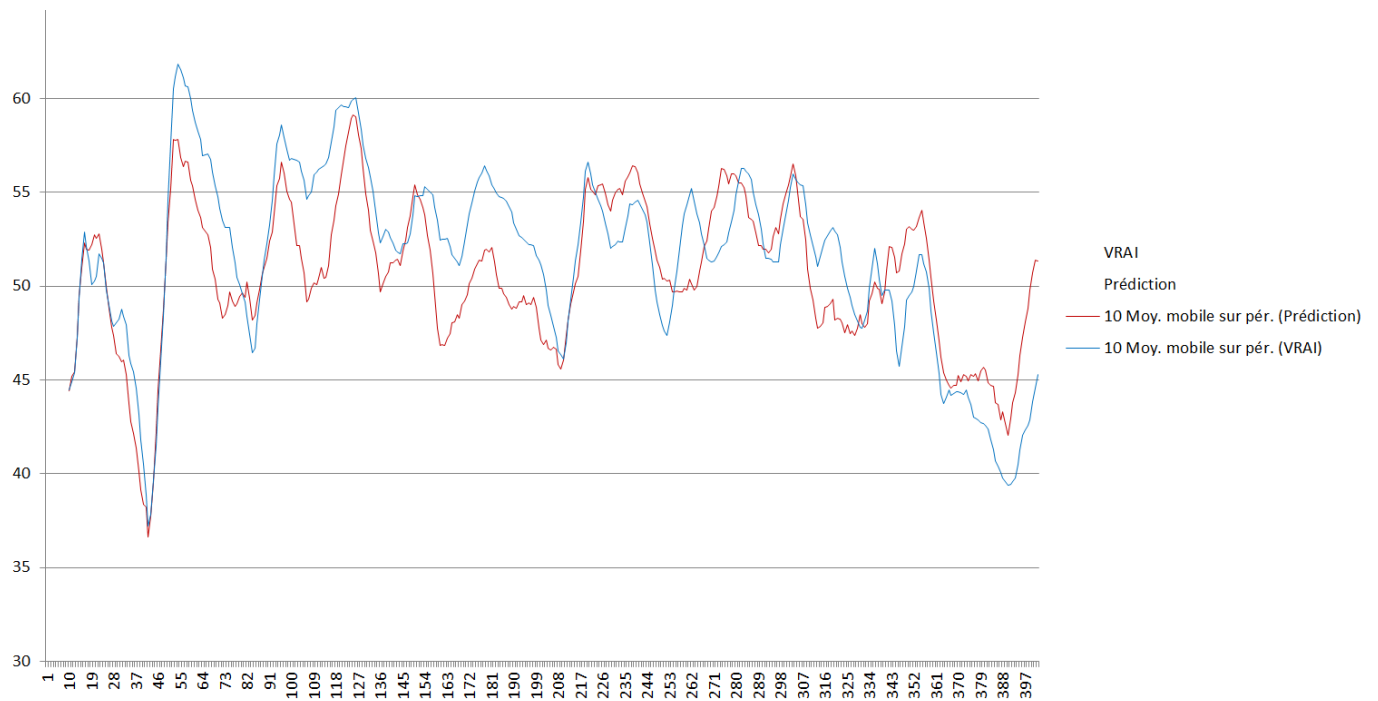


Figure 32 - Prédiction des moyennes avec le modèle 066.q2

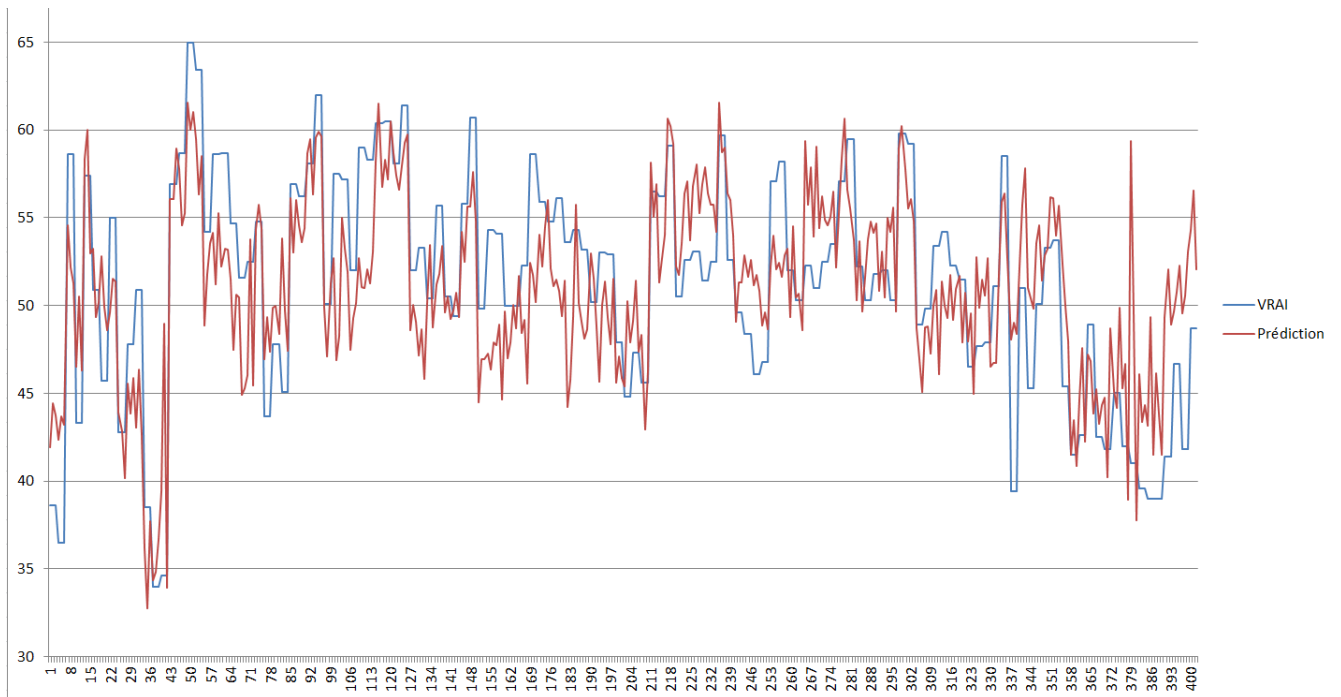


Figure 33 - Prédiction des valeurs individuelles avec le modèle 077.q2

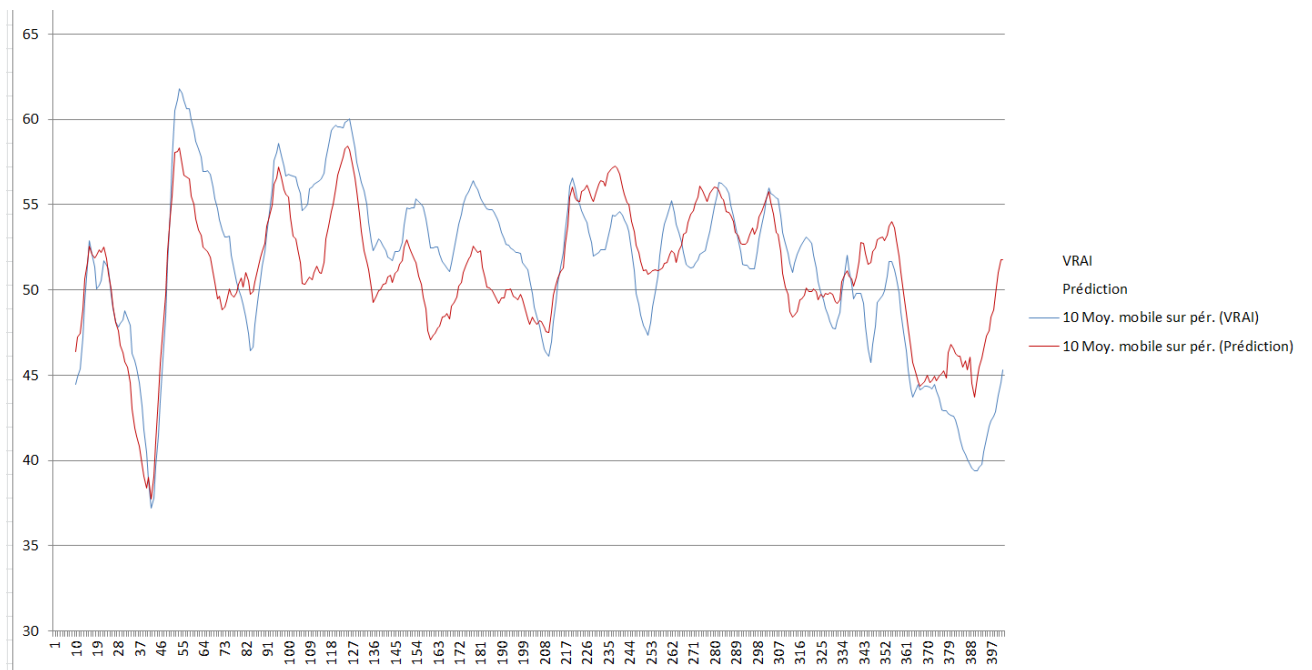


Figure 34 - Prédiction des moyennes avec le modèle 077.q2

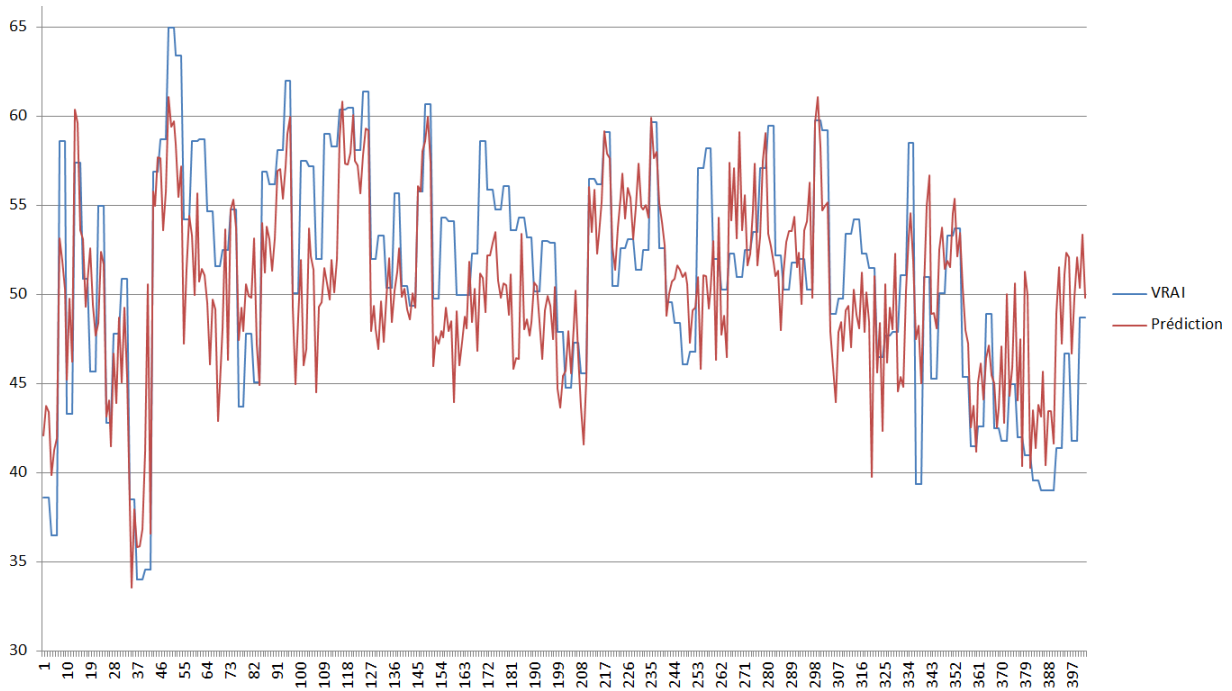


Figure 35 - Prédiction des valeurs individuelles avec le modèle 088.q2

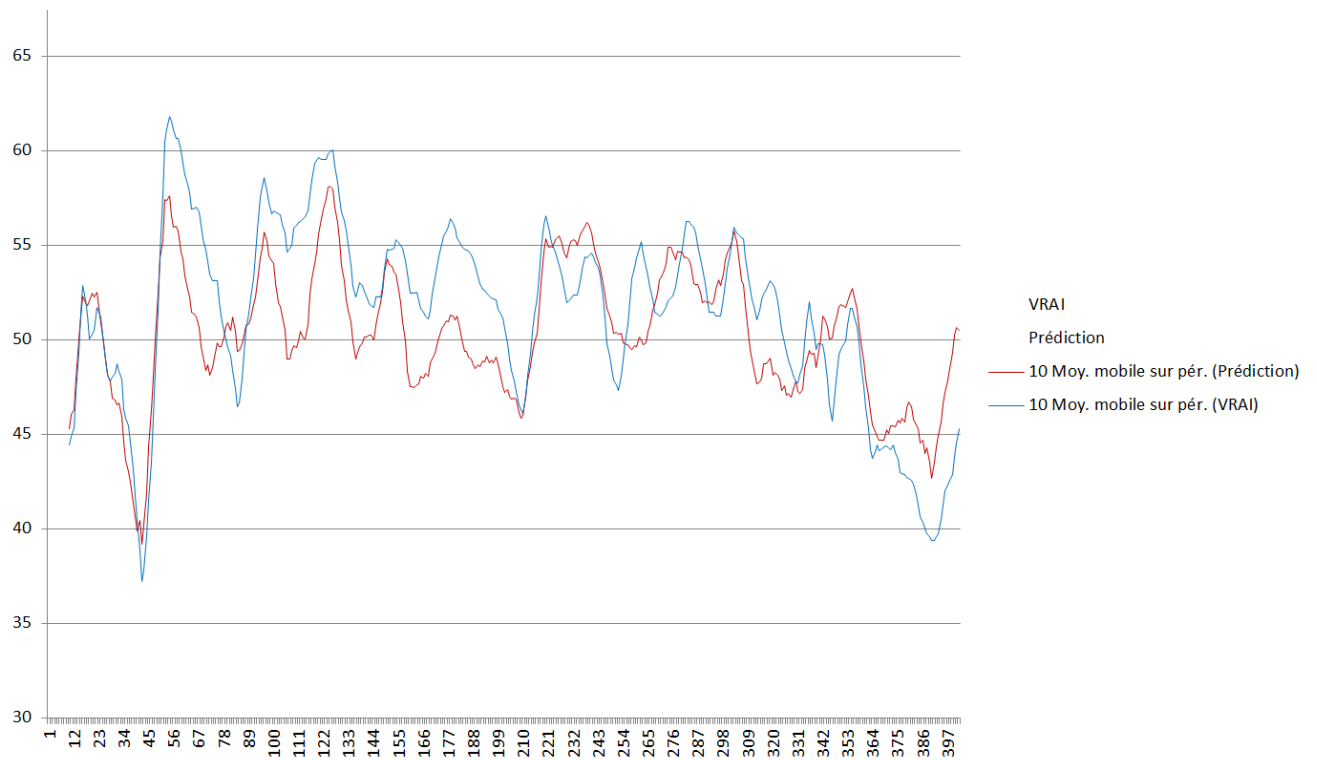


Figure 36 - Prédiction des moyennes avec le modèle 088.q2

La meilleure méthode est celle qui fait le moins d'erreurs. Chaque modèle révèle des données statistiques similaires et donne donc des prédictions de ratios L/P similaires. C'est plutôt bon signe. Des tendances peuvent être dégagées des données recueillies. Si une autre personne avait créé un modèle, il aurait été différent mais normalement avec des propriétés statistiques dans les mêmes ordres de grandeurs. En utilisant la totalité des échantillons pour créer un modèle, il n'y en avait plus afin de réaliser une validation externe mais les résultats statistiques sont concluants au vu des écarts entre les valeurs vraies et les valeurs prédites.

6. Modèle avec la totalité des échantillons

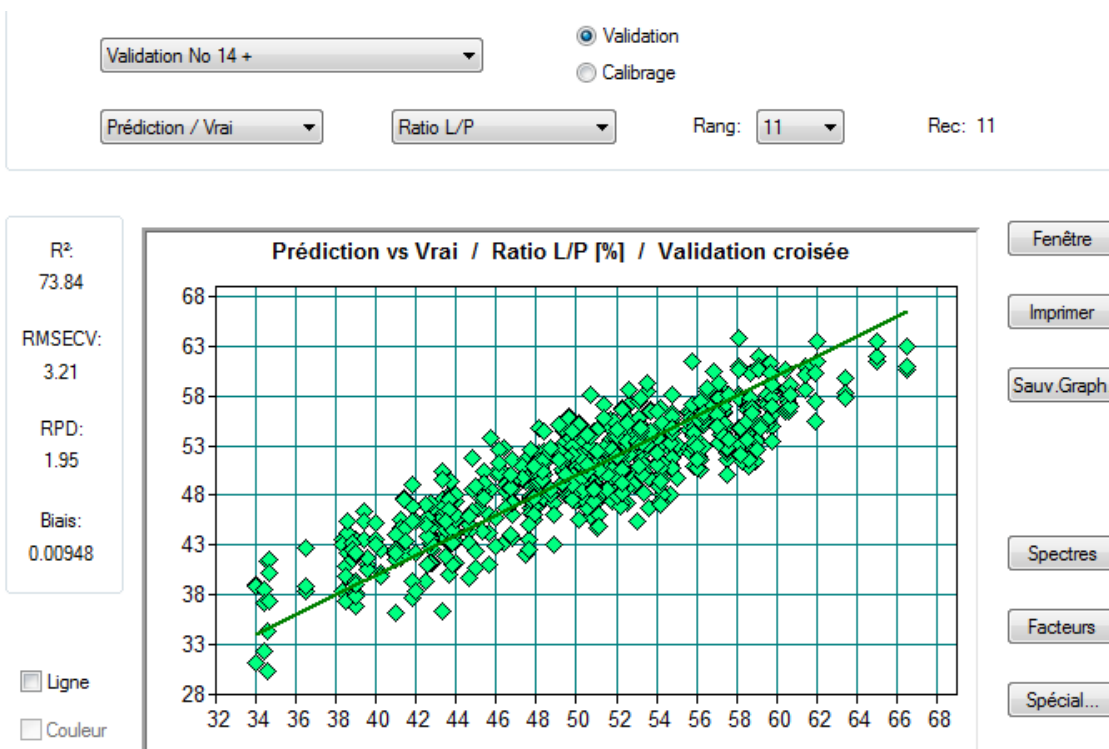


Figure 37 - Modèle incluant la totalité des spectres

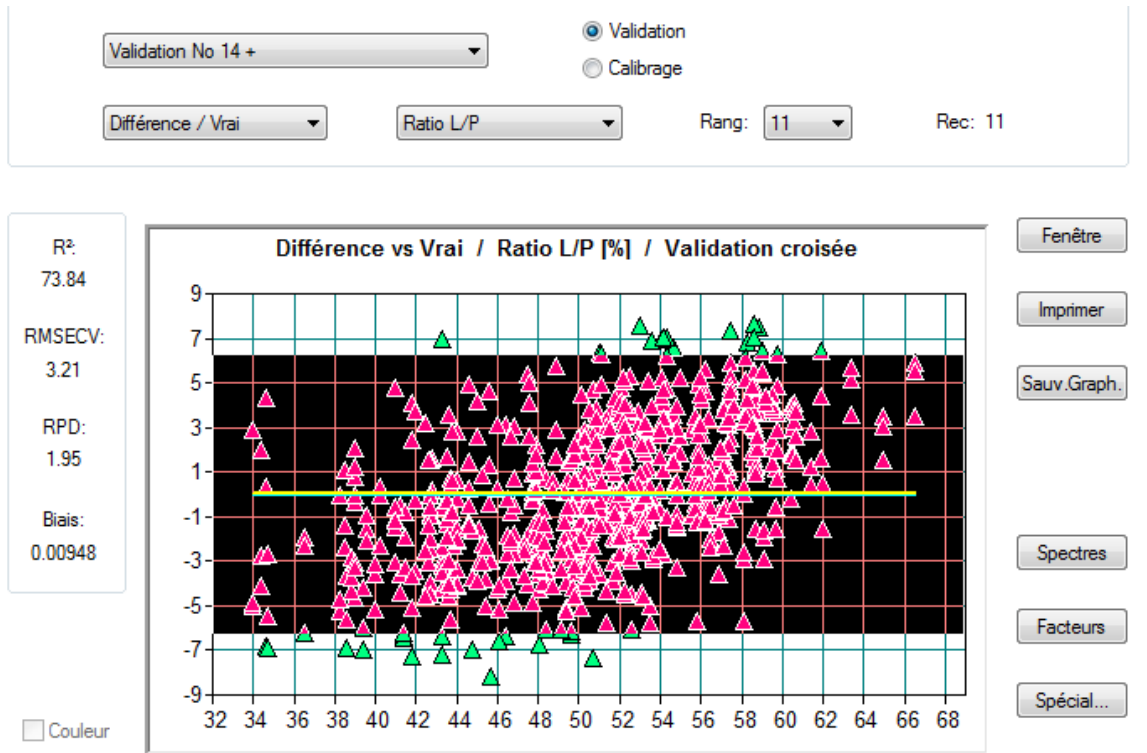


Figure 38 - Modèle incluant la totalité des spectres, différence par rapport aux vraies valeurs

Environ 95% des échantillons prédits se trouvent à + ou - 6% et 99,7% prédits se trouvent à +ou - 7%. (Loi normale, avec $\alpha = 0,05\%$).

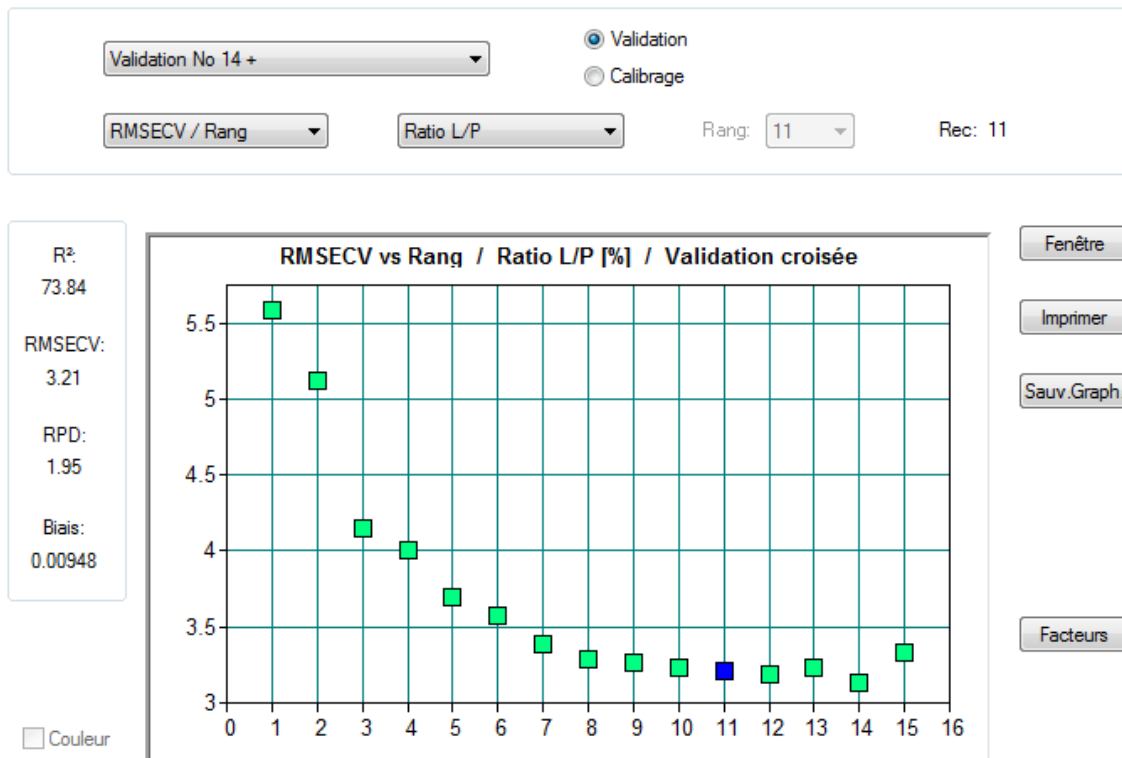


Figure 39 - Nombre de facteurs

11 facteurs semble expliquer les différents ratios feuilles/tiges. Le nombre de facteurs est élevé mais au vu de la courbe, ils représentent bien les différents ratios L/P de chaque échantillon.

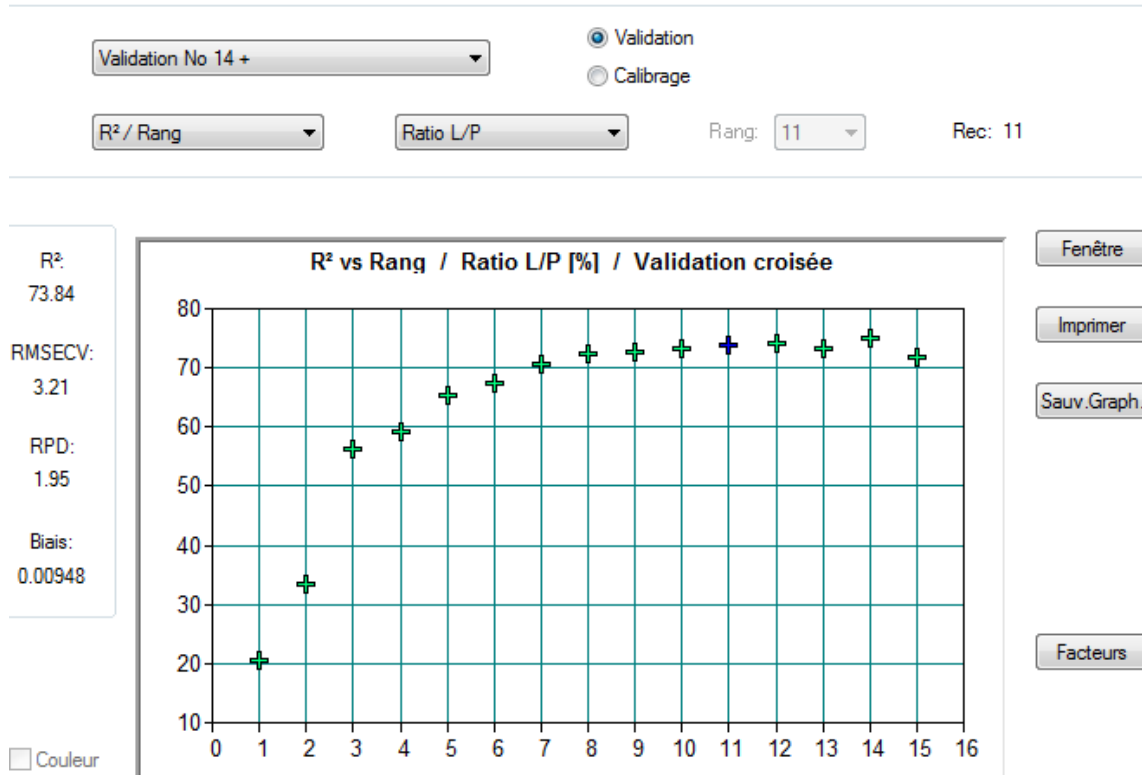


Figure 40 - Courbe des R²

On atteint un palier. R² pas aussi élevé qu'il le faudrait mais la courbe est fiable.

7. Discussion des résultats

7.1. Interprétation des résultats

En utilisant les moyennes de chaque lot analysé, pour la création des modèles, il n'y a pas une bonne représentation de chaque échantillon. En effet, l'écart-type inter-échantillon est en moyenne de 2,4. C'est-à-dire qu'il peut y avoir, en moyenne, un écart de + ou - 5% entre deux ratios L/P de deux échantillons provenant d'un même lot. Ceci peut être expliqué par l'hétérogénéité des parcelles rien qu'entre les côtés et le centre. Afin de mieux représenter la réalité nous prenons donc le ratio L/P manuelle trouvé pour chaque échantillon de 100g. Les spectres ont donc un ratio L/P identique 3 par 3.

Le RMSECV prend des valeurs aux alentours de 3. Ce qui reflète bien l'écart-type obtenu grâce à la méthode de référence (2,4 en moyenne). La prédiction du ratio L/P des échantillons est plutôt bonne mais avec une incertitude pouvant aller jusqu'à plus ou moins 6%, pour 95% des échantillons. Le modèle ne fera pas de prédictions meilleures que la méthode de référence. Graphiquement on peut retrouver des tendances entre les vraies valeurs et celles prédites. En moyennant les valeurs, le biais entre les vraies valeurs et celles prédites diminue. C'est-à-dire qu'avec la mesure d'un grand nombre d'échantillons, la moyenne peut être actualisée et être de plus en plus précise au fur et à mesure de la production. La moyenne sera lissée, les biais de + ou - 6% « s'annuleront ». L'écart-type entre chaque mesure est divisé par la racine du nombre de mesures.

Le RPD devrait être supérieur à 3 afin de valider directement une méthode. (RPD = RMSECV/SD). Ici il est aux environs de 2, mais les résultats sont encourageants. Le coefficient de détermination R^2 quant à lui est situé entre 73 et 85%. En regardant l'allure de la courbe il est tout à fait justifiable.

Une méthode avec un rang faible est plus fiable. Dans notre cas le rang semble élevée (souvent supérieur ou égal à 8). Néanmoins, en regardant l'allure de la courbe des composantes (qui expliquent les différences entre les variables) le rang obtenu est tout à fait justifiable. Beaucoup de facteurs différents expliquent le rapport feuilles/tiges. Nous pouvons imaginer que la couleur du persil, la taille des feuilles et des tiges, l'humidité ou encore la température peuvent influencer l'allure des spectres.

7.2. Limites

Lors de la création de modèles, la validation croisée est à privilégier au set de validation. En effet, en fonction de l'échantillonnage le nombre d'échantillons peut être trop faible pour appliquer le set de validation. L'échantillonnage appliqué pour les modèles semble influencer leur fiabilité. L'échantillonnage par échantillon (« 3 par 3 ») et non par lot (« par noms ») semble donner des modèles plus fiables. En effet, les prédictions sont plus précises et plus proches des valeurs vraies, trouvées manuellement. Il faut également repérer (graphiquement et statistiquement) les valeurs aberrantes pour les exclure du modèle.

Limites et risques pendant les mesures :

- Poids de l'échantillon
- Position/placement de l'échantillon
- Propriétés physiques (taille) et chimiques (composition) de l'échantillon
- Température de l'échantillon
- Humidité de l'échantillon

Limites dans le temps :

- Saisonnalité
- Variété

Tous ces facteurs ont pu influencer les spectres et donc jouer sur la fiabilité du modèle en plus de l'incertitude de la méthode de référence. Afin de rester fiable dans le temps, des données supplémentaires doivent être ajoutées au modèle utilisé. Les résultats obtenus via le modèle peuvent également être vérifiés de temps en temps grâce à des agrégés manuels.

7.3. Recommandations

Les prochaines étapes de l'expérimentation seraient l'amélioration de la méthode de référence et l'essai et la calibration du spectromètre en ligne. Le moyen fiable d'obtenir un ratio L/P précis d'un échantillon serait de mesurer des échantillons broyés/séchés mais il faut attendre 24 heures avant d'avoir les premiers résultats.

En ce qui concerne la SPIR, une sonde pourrait être placée en ligne. Un calibrage est nécessaire pour adapter le « modèle labo » à la ligne de production. Pour ce faire, les spectres mesurés en ligne doivent être corrélés avec de nouveaux ratios L/P.

L'évaluation du ratio feuilles/tiges peut aussi être faite par imagerie NIR hyperspectrale sur base d'images RGB. Après création d'un modèle, le traitement des images est réalisé sur base de règles de segmentation de couleurs.

Conclusion

Les résultats de cette étude sont encourageants. L'utilité de l'utilisation d'un spectromètre proche infrarouge pour la détermination du ratio L/P du persil plat frais a été démontrée. Il y aura toujours une incertitude sur la méthode de référence qui dépend de l'échantillon effeuillée et de l'opérateur. Au vu des premiers résultats, le nombre de facteurs à prendre en compte dans la méthode est d'au moins 8. Le rang à choisir est donc 8 au minimum.

Parmi les échantillons analysés, la prédiction du ratio L/P diffère de la valeur réelle à plus ou moins 6%. La prédiction du ratio L/P moyen du lot, quant à lui, diffère de 4 à 5%.

Pour aller plus loin, le persil plat pourrait être analysé par imagerie proche infrarouge hyper spectrale.

Pour la mise en ligne du spectromètre il y a nécessité d'adapter la méthode créée. Une calibration en ligne est nécessaire car les paramètres changent par rapport à l'expérimentation en laboratoire. Par ce calibrage on annule le biais qu'il y a entre les mesures faites en labo et celles faites en ligne.

1. Effectuer une série de mesures en ligne sur un même lot
2. Appliquer la méthode de référence sur ce même lot
3. Associer les données de la méthode de référence aux spectres obtenus avec les mesures en ligne
4. Ajouter à la méthode le biais des mesures en ligne
5. Utilisation en routine

Une mesure en ligne ne sera jamais plus précise que la méthode de référence mais permet surement d'augmenter la représentativité du lot. En effet, la moyenne sera calculée sur un grand nombre d'échantillons. Dans l'absolue, les écarts observées entre les vraies ratios et ceux prédits devraient diminuer significativement en moyennant toutes les mesures.

Liste des abréviations et symboles

CRAW : Centre Wallon de Recherches Agronomiques

INRA : Institut National de la Recherche Agronomique

L/P : Limbes/Pétioles

NIR : Near infrared

PCA : Principal component analysis

PIR : Proche infra-rouge

PLS: Partial Least Squares

RVB: Rouge vert bleu

SNIR: spectroscopy near infrared

SPIR: spectroscopie proche infrarouge

Liste des références bibliographiques

Norme Européenne, NF EN ISO 12099, Juillet 2010. Aliments des animaux, céréales et produits de mouture des céréales. Lignes directrices pour l'application de la spectrométrie dans le proche infrarouge.

BELLON-Maurel V., GRENIER P., CROCHON M., ROGER JM., 2003. *Spectroscopie optique pour caractériser l'état physiologique des plantes. Ingénieur N°spécial* – p.35 à 45.

CONZEN Jörg-Peter. 2003. *Multivariate Calibration: A practical guide for developing methods in the quantitative Analytical Chemistry.*

Davrieux F., Bastianelli D. et Thuriès L. (CIRAD), mars 2016. *La spectrométrie dans le proche infrarouge, Présentation générale, Théorie, principes et applications. Document pdf : 5.2.1-support-cours_LT.*

Bruker [en ligne]. Date de consultation : 10/06/16. Disponible sur : <https://www.bruker.com/fr/products/infrared-near-infrared-and-raman-spectroscopy/ft-nir/matrix-f/applications.html>

BIEMONT E., 2008. *Spectroscopie moléculaire: Structures moléculaires et analyse spectrale [en ligne]. Disponible sur : <https://books.google.fr/books?id=DoCvYRPKv-0C&pg=PA318&lpg=PA318&dq=spectroscopie+proche+infrarouge+finist%C3%A8re&source=bl&ots=zoq4y84yH7&sig=oA2pUOO WgtalRUf42h_T1VZ43Xo&hl=fr&sa=X&ved=0ahUKewjnsMiQyaTNAhWEbBoKHSmtCTEQ6AEITjAG#v=onepage&q=spectroscopie%20proche%20infrarouge%20finist%C3%A8re&f=false>*

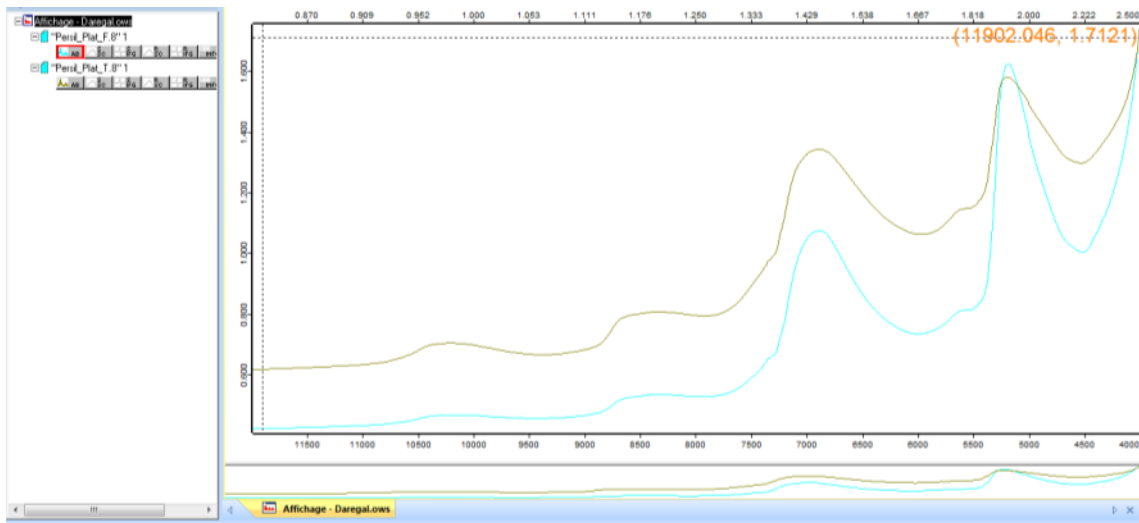
Genot V., Bock L., Dardenne P. et al. *Site de Biotechnologie, Agronomie, Société et Environnement [En ligne]. Date de consultation : 10/06/16. Accessible sur: <http://popups.ulg.ac.be/1780-4507/index.php?id=11096>*

Lachenal G. *La spectroscopie proche infrarouge. Site internet de l'Université de Lyon [En ligne]. Date de consultation : 09/06/16. Accessible sur: <http://nte-serveur.univ-lyon1.fr/spectroscopie/techniquepir.html>*

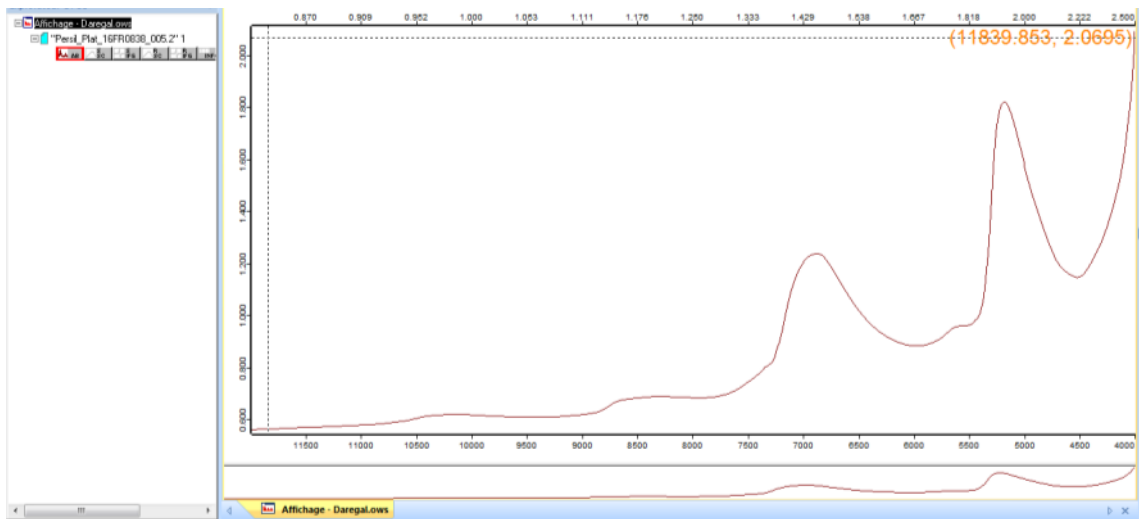
W. Herschel. *Experiments on the Refrangibility of the Invisible Rays of the Sun. [En ligne] Accessible sur: http://www.jstor.org/stable/107057?seq=6#page_scan_tab_contents*

Informations complémentaires et conseils : Alicia Ayerdi-Gotor (Institut Polytechnique Lasalle Beauvais), Philippe VERMEULEN (Walloon agricultural research center), Bernadette JULIER (INRA)

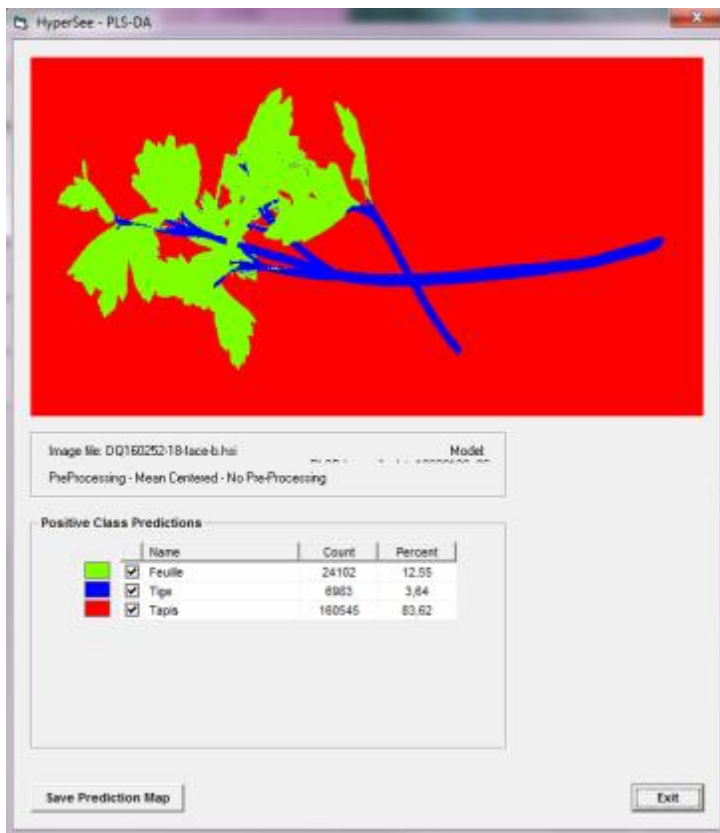
Annexes



Annexe 1 - Exemple d'un spectre de feuilles pures et d'un spectre de tiges pures de persil plat frais



Annexe 2 - Exemple de spectre d'un échantillon de 100g de persil plat frais



Annexe 3 - Persil plat frais entier analysé par imagerie proche infrarouge hyper spectrale



Image file: DQ160252-16-face-b.hsi Model: _____
 PreProcessing - Mean Centered - No Pre-Processing

Positive Class Predictions

	Name	Count	Percent
<input checked="" type="checkbox"/>	Feuille	49076	25,56
<input checked="" type="checkbox"/>	Tige	5656	2,95
<input checked="" type="checkbox"/>	Tapis	136723	71,21

Save Prediction Map Exit

Annexe 4 - Persil plat frais effeuillé analysé par imagerie proche infrarouge hyper spectrale