

## Mesure du % d'acides oléiques et linoléiques dans les graines de Tournesol par Spectroscopie Proche infrarouge AOTF.



## Résumé

Depuis les dernières années, il existe un intérêt croissant de produire des graines de tournesols contenant une quantité élevée d'acides oléiques.

Une fois les souches reconnues comme souches parentales, ont été sélectionnées et la première génération issue du croisement a été réalisée, l'objectif du sélectionneur est d'identifier dans les semences ainsi produites, quelles sont celles qui contiennent un fort taux d'acide oléique. Une fois ces dernières sélectionnées, il peut alors produire la deuxième génération en croisement inverse et une fois encore identifier celles qui contiennent un fort taux d'acides oléiques.

La méthode actuelle pour déterminer l'acide oléique, est de réaliser l'analyse en CPG sur une petite quantité d'embryon de la graine débarrassée de son enveloppe. Cette partie de l'embryon non détruite pour l'analyse en CPG peut toujours être plantée et, dans la plupart des cas, germera. Cependant, sans la protection de l'enveloppe charnue, l'embryon pourra subir rapidement des altérations chimiques, qui le rendront alors inutilisable.

Une méthode non destructive, avec laquelle l'acide oléique pourrait être déterminé rapidement pour une population importante de semences en tenant compte de l'intégrité de la protection de la coque, serait extrêmement utile.

L'étude présentée ici, montre qu'il est possible de déterminer le % d'acide oléique ainsi que d'autres constituants des graines de tournesol, sans avoir à détruire la coque. L'analyseur utilisé ici, est un Luminar 3076 SEED MEISTER de Brimrose.

Les semences sont placées sur un petit tapis roulant, qui les amène sous le faisceau PIR AOTF. Le spectre réalisé sur chaque échantillon, est analysé par un modèle en PLS1 et la quantité en acide oléique est ainsi déterminée simultanément.

Les semences sont automatiquement triées selon le taux d'acide oléique prédit.

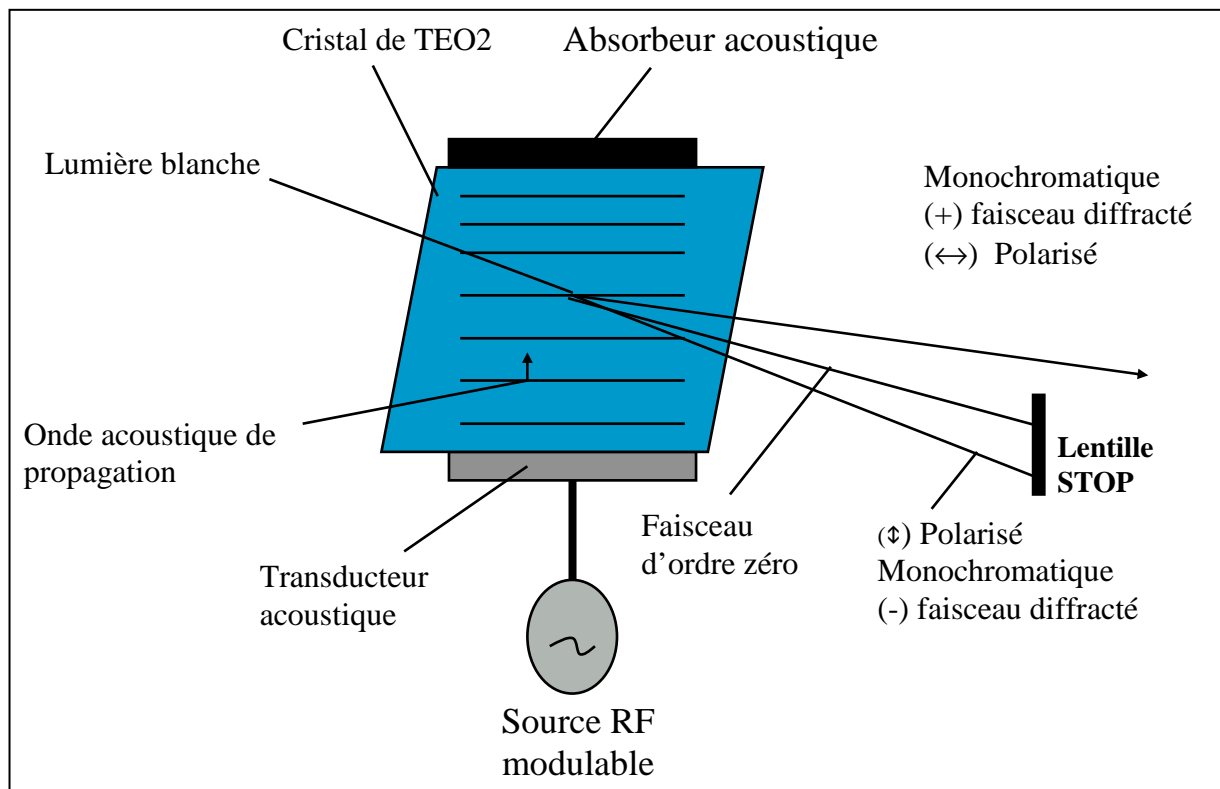
## 1. Introduction

Le principe de l'AOTF (Acousto-Optic Tunable Filter) est basé sur la diffraction acoustique de la lumière dans un milieu anisotropique. Le spectromètre est formé par le couplage d'un transducteur piézo-électrique à un cristal biréfringent.

Quand le transducteur est excité par un signal RF, des ondes acoustiques sont générées dans le cristal. Les ondes acoustiques de propagation produisent une modulation périodique de l'indice de réfraction. Cela produit un déplacement de la phase mobile qui dans les conditions adéquates, va diffracter une partie d'un faisceau lumineux incident.

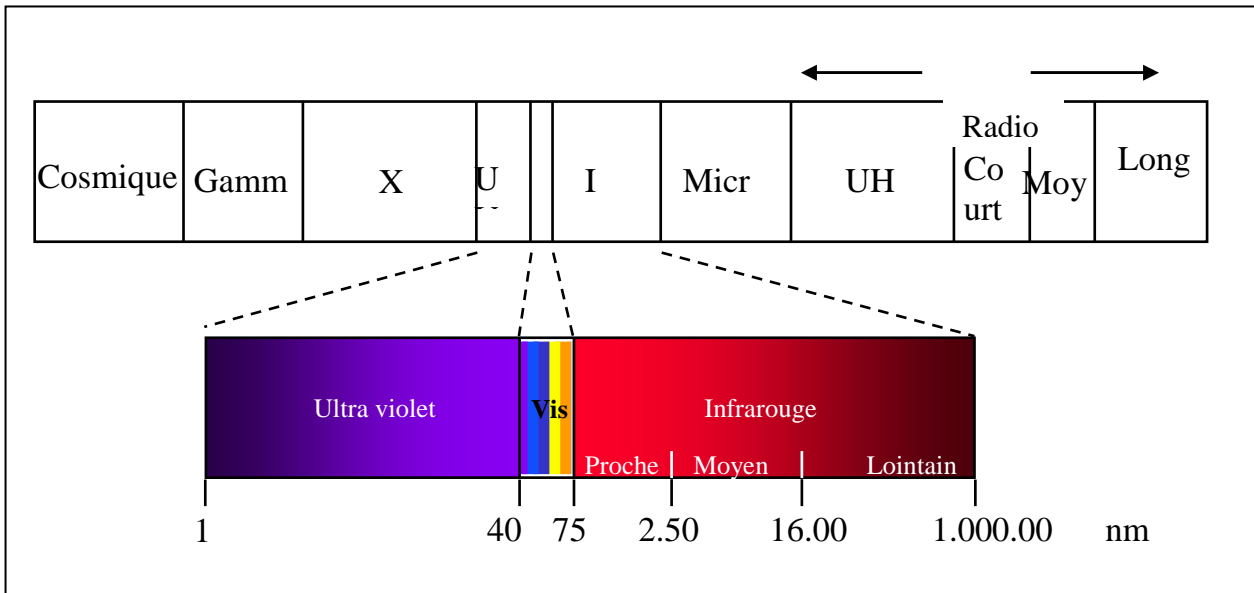
A une fréquence acoustique donnée, une bande de fréquence optique étroite correspond aux conditions sélectives de phase et peut être diffractée cumulativement.

Lorsque la fréquence RF change, le centre de la bande passante optique est modifié en respectant les conditions sélectives de phase.



**Figure 1. Schéma de principe de l'AOTF**

## Petit rappel sur la spectroscopie proche infrarouge



**Figure 2. Gamme spectrale**

La région du spectre proche- infrarouge s'étend de 800nm à 2500nm.

Les bandes d'absorptions les plus intenses sont dues aux harmoniques et aux bandes de combinaisons des vibrations fondamentales actives dans la région du moyen-infrarouge. Les transitions énergétiques discrètes se font de l'état vibratoire "au repos" vers les états vibratoires excités seconds et tertiaires. Plus l'énergie des transitions augmente, moins elles sont fréquentes ; ainsi les harmoniques qui en résultent sont de moins en moins intenses.

Comme l'énergie nécessaire pour atteindre le deuxième ou troisième état excité est deux à trois fois l'énergie nécessaire pour atteindre le premier état excité, la bande d'absorption résultante est égale à la moitié ou le tiers de la longueur d'onde de la bande fondamentale. En plus des harmoniques, des bandes de combinaisons peuvent également apparaître. Ces bandes correspondent le plus souvent à une bande d'étirement et à une ou plusieurs bandes de torsions. De nombreuses combinaisons sont possibles, et ainsi la région PIR est complexe, avec de nombreuses bandes inconnues.

La spectroscopie Proche Infrarouge est couramment utilisée comme outil quantitatif basé sur la chimiométrie, et permet de développer des étalonnages reliant des analyses de référence de laboratoire avec des spectres PIR.

Le traitement mathématique des données spectrales PIR se fait en MLR (Multi Linear Regression), en PCA (Principle Component Analysis), en PLS (Partial Least Squares) et en analyse discriminante. Tous ces algorithmes peuvent être utilisés séparément ou ensemble pour prédire quantitativement ou qualitativement les paramètres intéressants.

En résumé, pour le **PIR AOTF**:

- ✓ Petite portion du spectre électromagnétique (700-2500 nm)
- ✓ Pas de préparation de l'échantillon: Mesure directe sur échantillons multiples
- ✓ Mesures simultanées de différentes propriétés chimiques et/ou physiques
- ✓ Temps de réponse très court (quelques secondes)
- ✓ Non-destructif
- ✓ Applicable à tout type d'échantillon: solides, liquides, boues, pâtes, gels, gaz
- ✓ Fibres optiques en silice avec teneur OH faible, et multiplexage possible

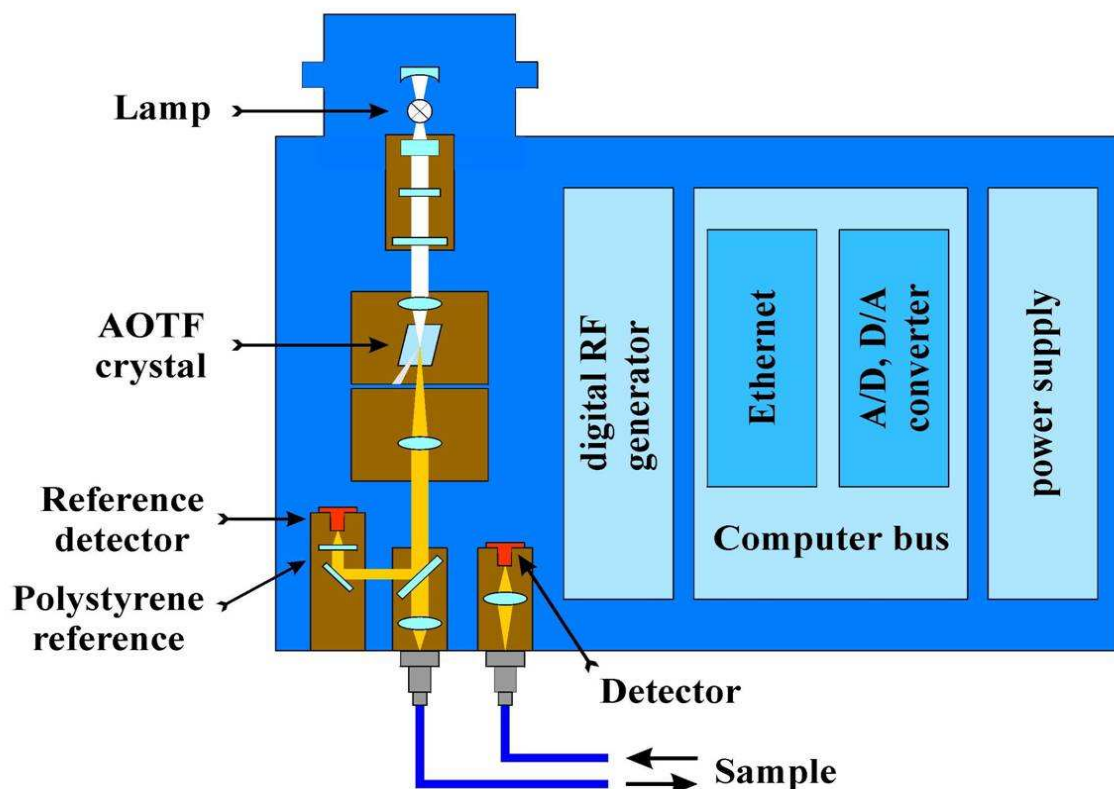


Figure 3. Schéma d'un analyseur AOTF

## 2. Méthodologie

Brimrose a reçu 200 graines de Tournesol, emballées chaque une individuellement et étiquetées avec un numéro identifiant unique.

Les 200 graines ont formé un jeu de calibration avec lequel le SEED MEISTER a été calibré. 4 sachets supplémentaires, de 20 graines chacun, ont été utilisés comme jeux de validation.

Nous avons utilisé la gamme de longueur d'ondes de 900nm à 1400nm ainsi qu'une résolution de 2nm. Avant d'acquérir les spectres des semences, le petit tapis roulant de l'analyseur sur lequel les graines sont placées, a été étalonné à une vitesse permettant d'acquérir 40 scans/graine.

Ces 40 spectres ont été moyennés en un spectre unique pour chaque échantillon par le programme. Les graines étaient placées manuellement sur le tapis roulant.

À l'exception de 2 graines, la population entière a été scannée deux fois.

Dans la première session de balayage spectral des échantillons, les graines ont été orientées de telle façon que la pointe de la graine était la première à rencontrer le faisceau PIR, puis dans la deuxième session, c'est la couronne de la graine qui était la première à rencontrer le faisceau.

Dans les 2 sessions, la face éclairée de la graine était présentée de façon aléatoire. Les graines ont été ensuite emballées et données à un laboratoire pour l'analyse en CPG, qui a rendu les résultats après 2 semaines.

Après réception des résultats d'analyse CPG, on a construit un modèle de régression, qui a été intégré dans l'application Seedme.exe.

Avant l'acquisition spectrale, les semences des 4 jeux de validation ont été d'abord séparées et emballées individuellement.

Les 4 populations ont été scannées 3 fois par le SEED MEISTER. En même temps que l'on a réalisé l'acquisition spectrale de chaque semence, le modèle a été appliqué et les résultats automatiquement sauvegardés dans un fichier .log.

Les 4 jeux de semences ont été renvoyés au labo pour analyse en CPG pour que les résultats obtenus en PIR AOTF soient comparés aux résultats du labo.

### 3. Résultats

#### 3.1. Spectres

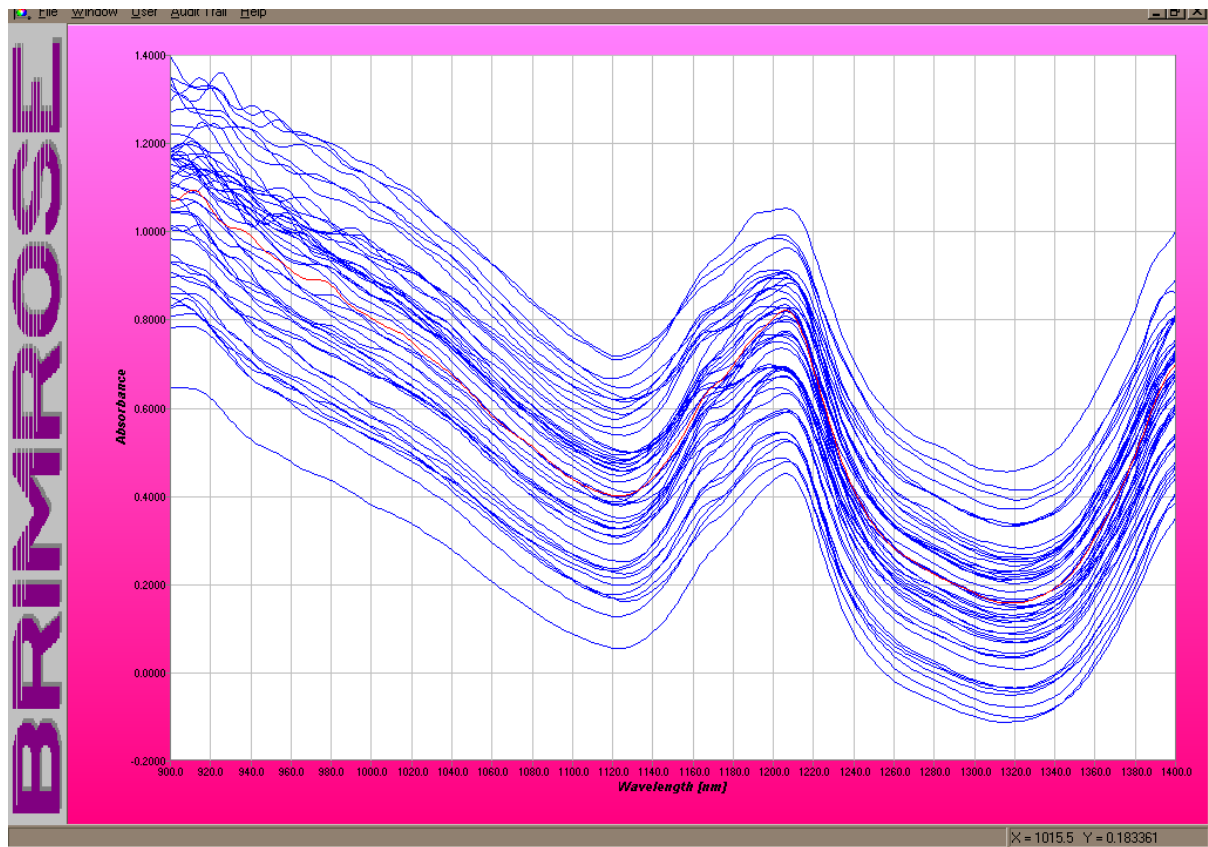


Figure 4. Spectres d'absorbance des graines de tournesol.

### 3.2. Modèles et Régressions linéaires

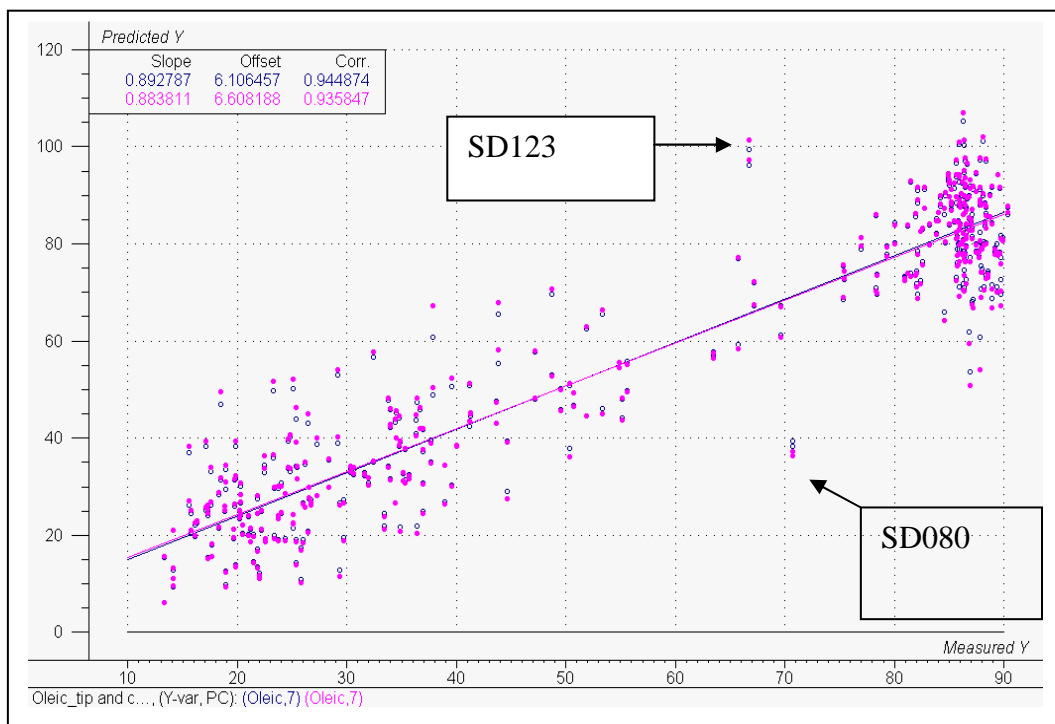
Les données spectrales ainsi que les résultats de l'analyse CPG ont été importés dans Unscrambler.

Les modèles d'étalonnages ont été construits en PLS1 (Partial Least squares). L'étalonnage initial, qui ne comprend que 7 composantes principales, paraît très bon.

Sans enlever aucun échantillon aberrant potentiel, la corrélation et la validation de l'étalonnage sont respectivement de 0.94 et 0.93 (voir Figure 5).

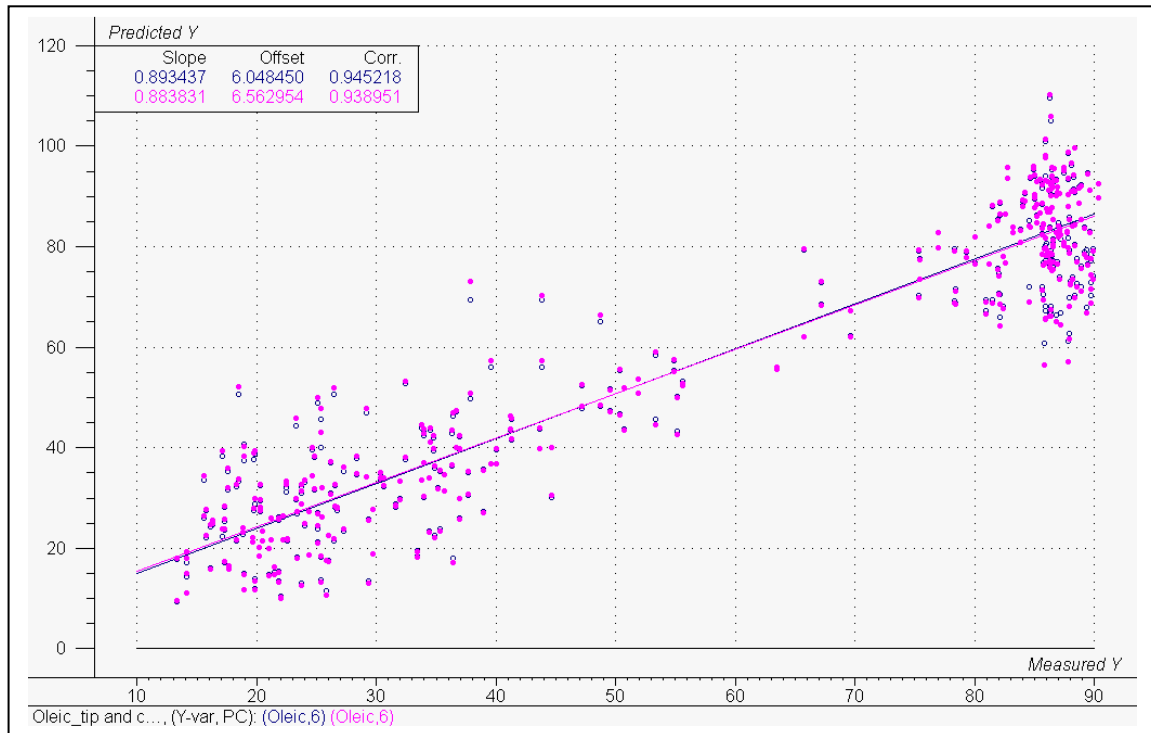
Les échantillons de semences SD080 et SD123 montrent tous les deux qu'ils ont une forte influence sur le modèle et qu'ils peuvent être considérés comme échantillons aberrants.

Enlever ces 2 échantillons, améliore légèrement le modèle en utilisant que 6 composantes principales et réduit la SEP de 10.1 à 9.9, (voir Figure 6).



**Figure 5. Régression en PLS1.**

Sans retirer les échantillons aberrants, la SEP est de 10.1



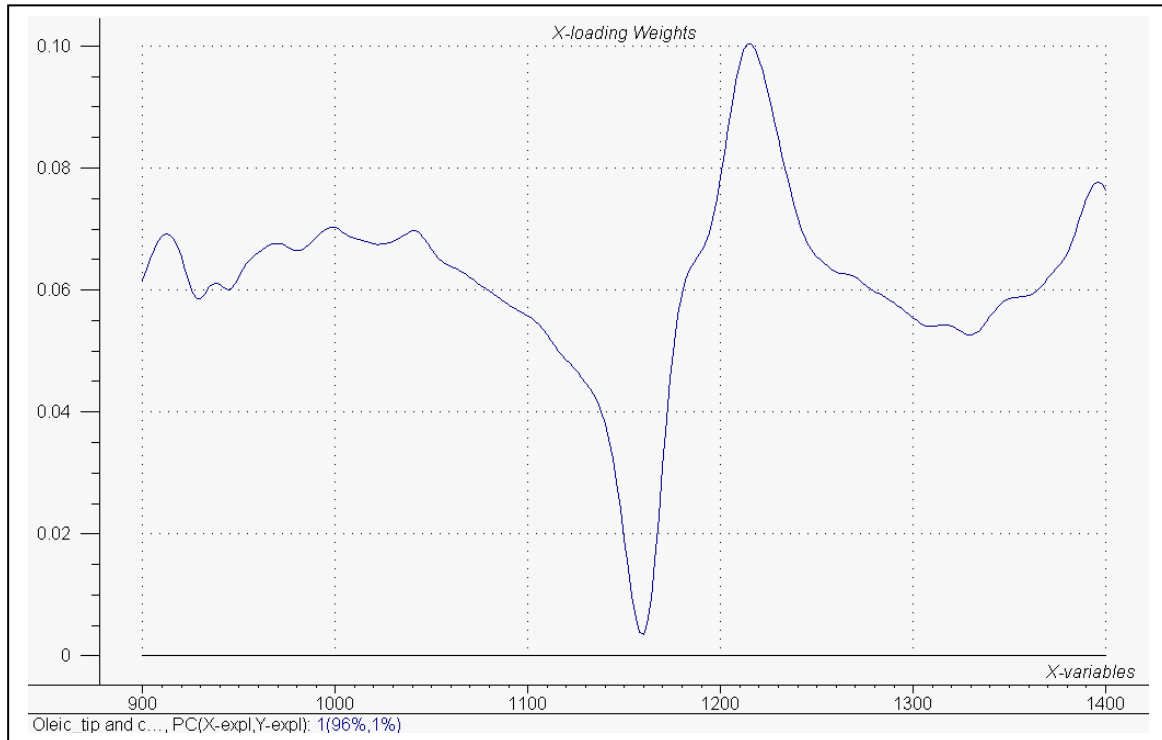
**Figure 6. Régression en PLS1.**

En retirant les échantillons aberrants, la SEP est de 9.9

Les X-loading weights indiquent que ces régions spectrales contribuent plus fortement au modèle de régression.

Selon les données de référence, consultables dans « l'Atlas des spectres Proche Infrarouge » de Sadler, l'acide oléique possède des pics d'absorbance entre 920nm et 930nm approximativement, ainsi qu'un épaulement entre environ 1150nm et 1160nm, et un pic large juste après 1200nm.

Comme visible dans la Figure 7, les X-loading pour la première composante principale, qui expliquent 93% du modèle, sont intenses plus précisément dans ces régions.

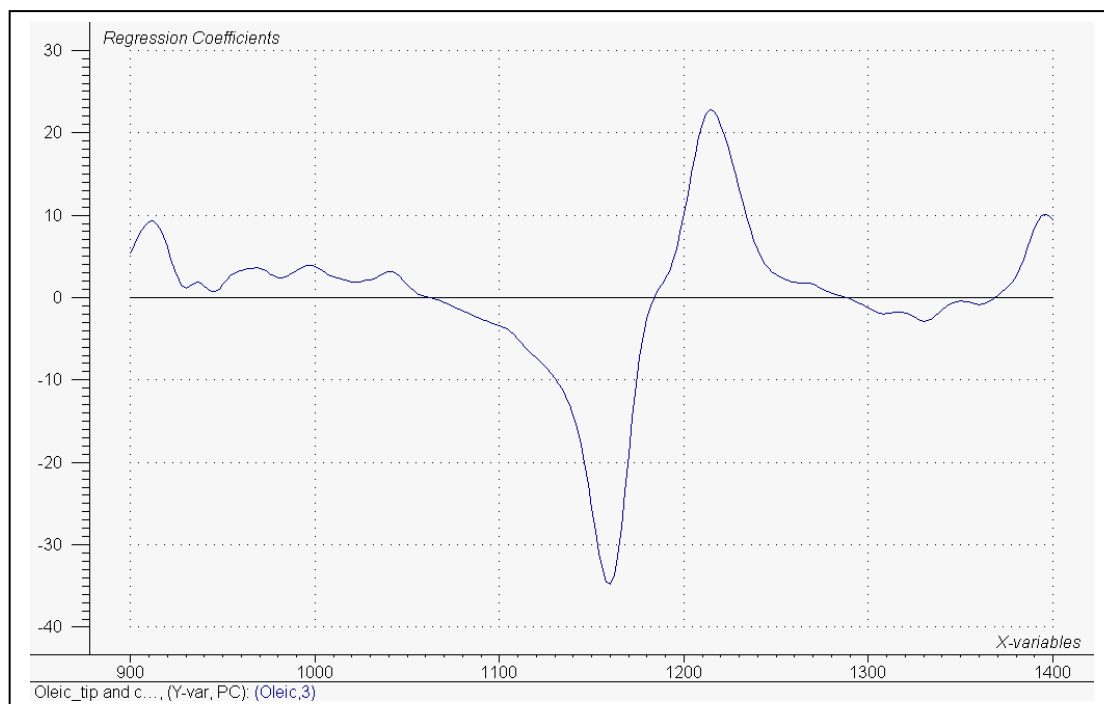


**Figure 7.** Les X-loading pour la première composante principale expliquent 93% du modèle.

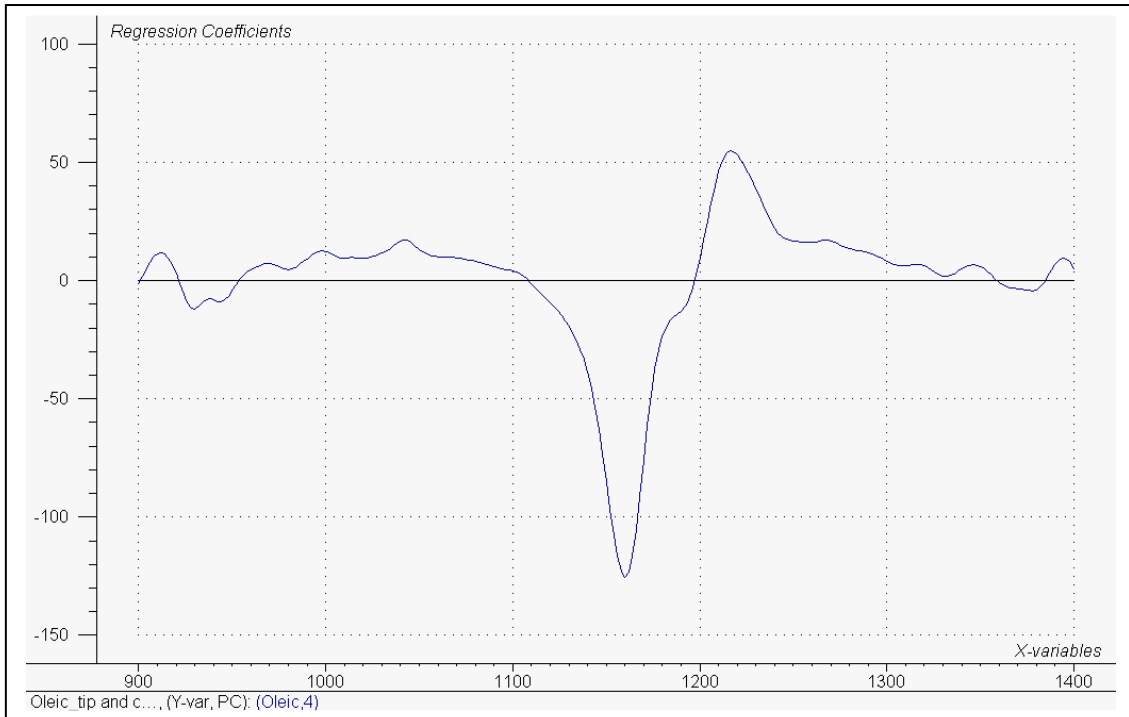
Les coefficients de régression qui résument les relations entre les facteurs prédictifs (données spectrales) et la réponse donnée, sont aussi élevés pour les régions spectrales correspondantes.

Les figures 8 et 9 ci-dessous, montrent les coefficients de régression pour les composantes principales 3 et 4 qui cumulativement décrivent 74% des données "Y", en montrant des pics clairement situés dans les zones à 1160nm (liaisons CH<sub>2</sub>) et à 1220nm (liaisons CH).

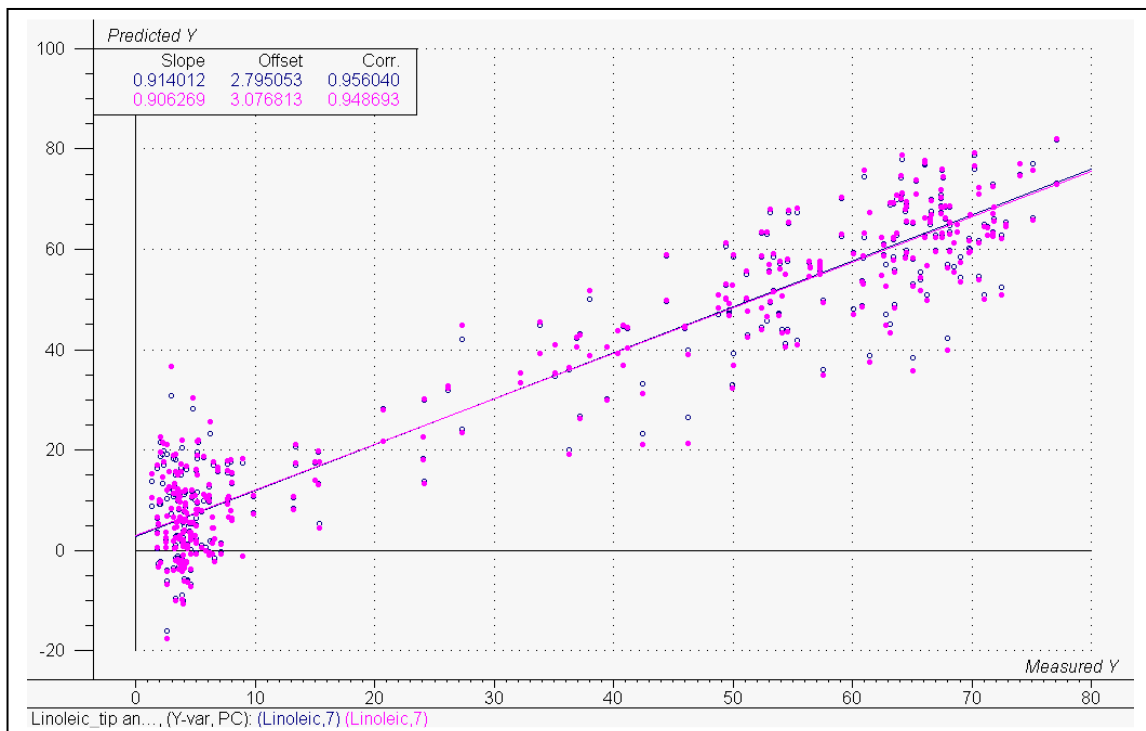
Moins prononcé mais tout autant significatif, est le faible pic situé dans la région de 920 nm. Le modèle pour l'acide linoléique est semblable à celui pour l'acide oléique (voir figure 10) et les résultats escomptés sont selon un scénario inverse au regard des X-loading et des coefficients de régression (voir figures 11 et 12).



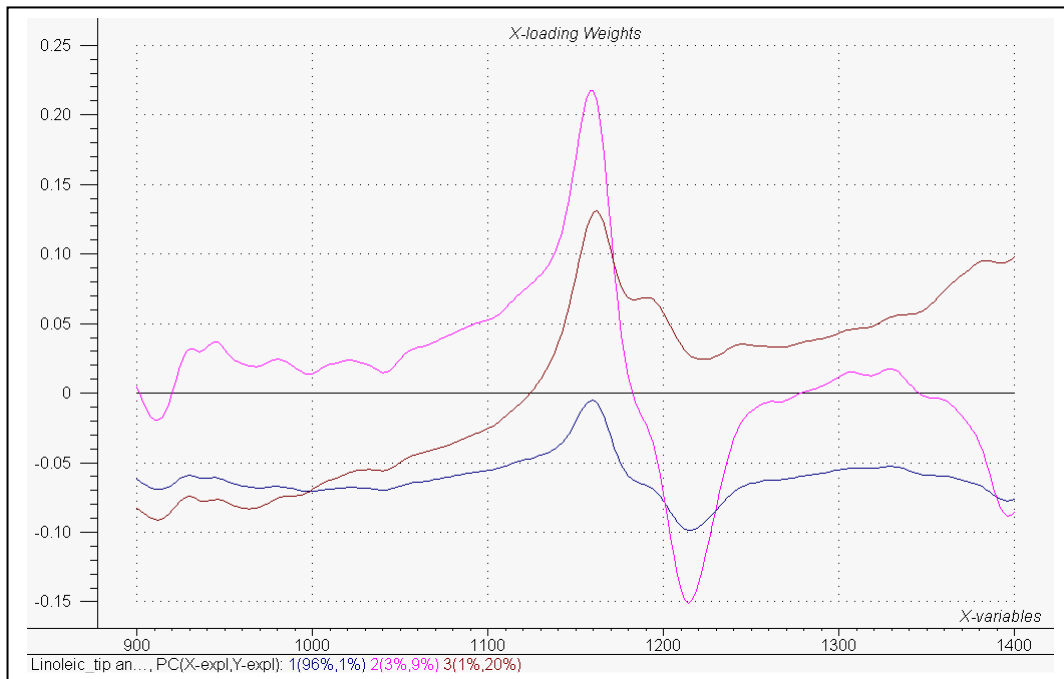
**Figure 8. Coefficients de régression pour la troisième composante principale, qui expliquent 21% de la variation de Y.**



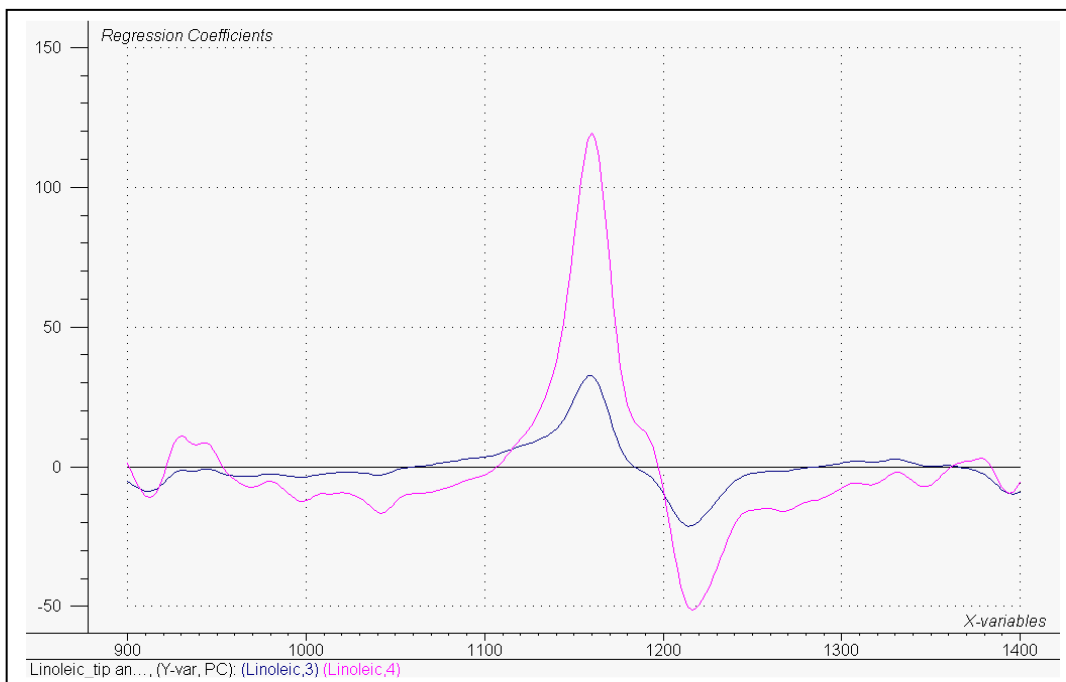
**Figure 9. Coefficients de régression pour la quatrième composante principale, qui expliquent 53% de la variation de Y.**



**Figure 10. Régression pour l'acide linoléique sans 2 aberrants.**  
La SEP est égale à 8.8.



**Figure 11. X-loading pour les PC 1, 2 et 3 pour la régression pour l'acide linoléique.**



**Figure 12. Coefficients de régression pour les PC 3 & 4 qui expliquent 73% de la variation de Y.**

### 3.3. Prédiction

Chacun des 4 paquets de semences, a été scanné 3 fois sans tenir compte de l'orientation des graines placées sur le tapis roulant.

Les résultats de prédiction pour les 3 répétitions de chaque paquet de semences sont affichés dans les tableaux 1, 2, 3 et 4.

Il est à noter que la valeur médiane pour chaque répétition est nettement plus élevée que la moyenne, ce qui semble montrer que les résultats sont biaisés.

Les figures 13, 14, 15 et 16 montrent la fréquence de distribution des résultats de prédictions pour les 3 répétitions de chacune des 4 populations.

Elles indiquent clairement que pour chaque population, il existe une discrimination entre les semences à haute ou basse quantité d'acide oléique.

Dans un effort de réduction de la quantité d'analyse au laboratoire en CPG, nous avons seulement choisi 9 graines sur les 20 disponibles pour chaque population.

Basé sur les prédictions en PIR, 6 graines choisies dans les extrêmes bas et haut et 3 graines choisies dans la moyenne, ont été sélectionnées pour l'analyse en CPG.

Les résultats sont dans les tableaux 1a, 2a, 3a et 4a.

	Population 104-2		
Echantillons	Rép 1 Prédiction	Rép 2 Prédiction	Rép 3 Prédiction
104-2_01	17.91	6.28	9.31
104-2_02	68.54	67.29	72.75
104-2_03	12.79	18.34	23.37
104-2_04	26.81	24.16	37.80
104-2_05	78.63	79.10	77.83
104-2_06	84.12	85.80	85.84
104-2_07	66.17	59.41	64.30
104-2_08	54.19	66.13	52.96
104-2_09	17.12	13.65	17.34
104-2_10	50.81	35.10	58.35
104-2_11	73.24	80.84	83.89
104-2_12	57.76	57.52	71.56
104-2_13	70.04	76.00	71.04
104-2_14	86.04	82.83	67.86
104-2_15	71.31	69.15	73.13
104-2_16	56.77	56.02	52.40
104-2_17	24.68	73.24	72.40
104-2_18	69.06	20.03	32.13
104-2_19	72.43	78.95	81.52
Moyenne	55.71	55.25	58.20
Médiane	66.17	66.13	67.86

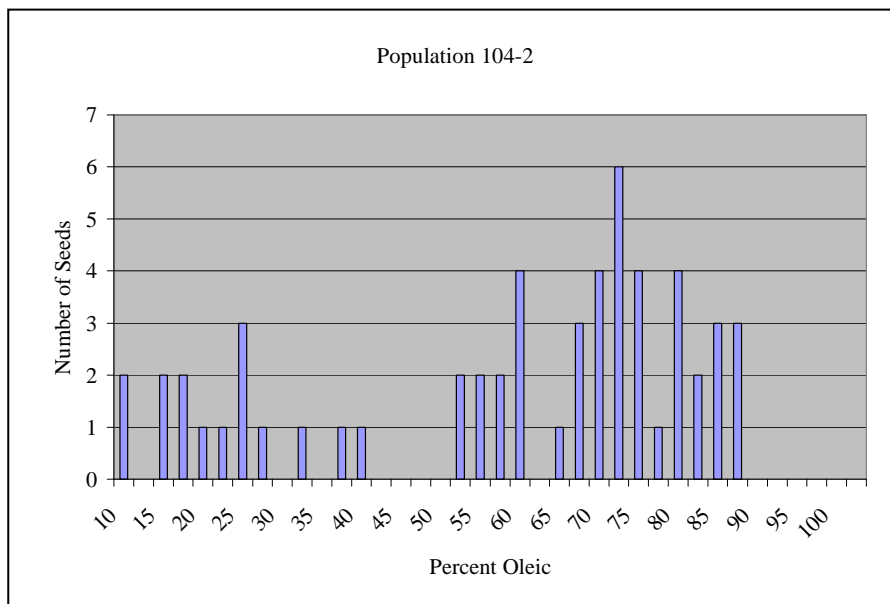
**Tableau 1.** Prédictions et statistiques de la population 104-2.

	Population 121-4		
Echantillons	Rép 1 Prédiction	Rép 2 Prédiction	Rép 3 Prédiction
121-4_01	79.72	78.05	72.98
121-4_02	76.68	76.86	80.61
121-4_03	76.59	76.62	83.94
121-4_04	82.01	84.21	80.92
121-4_05	90.42	78.17	84.39
121-4_06	88.97	92.83	
121-4_07	90.47	77.44	76.22
121-4_08	85.61	83.86	80.21
121-4_09	78.83	80.01	80.95
121-4_10	17.82	13.25	22.70
121-4_11	78.67	76.74	81.37
121-4_12	89.36	84.04	80.73
121-4_13	84.32	79.70	81.40
121-4_14	89.38	92.13	89.97
121-4_15	64.68	67.83	65.27
121-4_16	20.52	14.49	20.53
121-4_17	18.39	32.42	35.55
121-4_18	84.18	80.78	85.02
121-4_19	16.27	19.10	16.45
Moyenne	69.10	67.82	67.73
Médiane	79.72	78.05	80.67

**Tableau 2.** Prédictions et statistiques de la population 121-4.

	Population 104-2	
Echantillons	Résultats en CPG	Moyenne PIR des 3 répétitions
104-2_01	23.65	11.17
104-2_03	19.96	18.17
104-2_05	85.61	78.52
104-2_06	89.12	85.25
104-2_07	72.25	63.29
104-2_09	18.57	16.04
104-2_14	86.76	78.91
104-2_16	64.30	55.07
104-2_19	88.65	77.64
Moyenne	60.99	53.78
R <sup>2</sup>	0.986	
SEP	3.940	

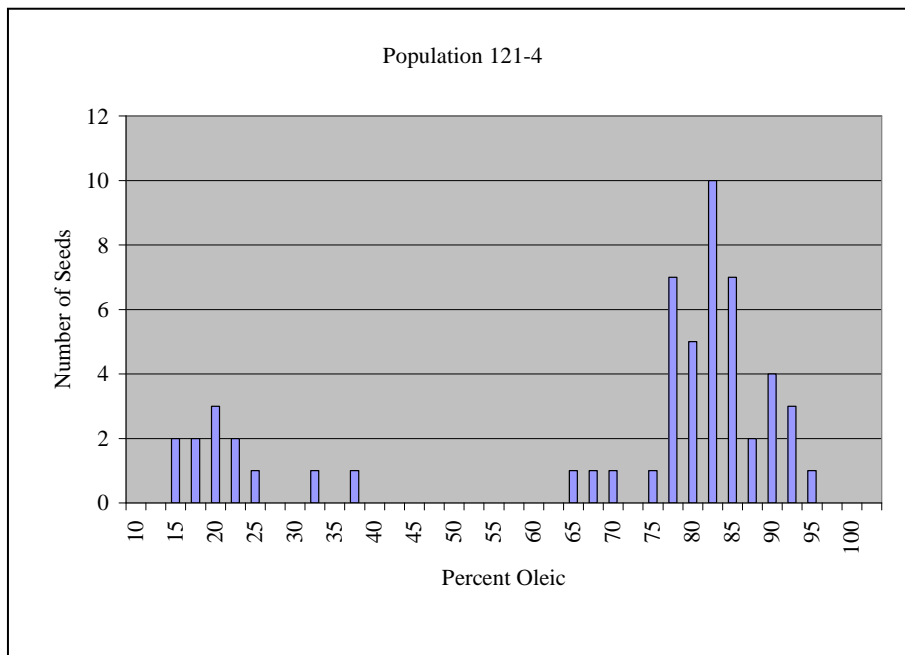
**Tableau 1a.** Analyse en CPG et moyenne des 3 répétitions en PIR pour les échantillons sélectionnés de la population 104-2.



**Figure 13.** Fréquence de distribution des résultats de prédiction pour les 3 répétitions de la population 104-2

	Population 121-4	
Echantillons	Résultats en CPG	Moyenne PIR des 3 répétitions
121-4_02	82.42	78.05
121-4_04	89.55	82.38
121-4_05	89.28	84.32
121-4_06	88.06	90.90
121-4_10	21.16	17.92
121-4_14	88.90	90.49
121-4_15	80.08	65.93
121-4_17	25.36	28.78
121-4_19	17.40	17.27
Moyenne	64.69	61.78
R <sup>2</sup>	0.971	
SEP	5.924	

**Tableau 2a.** Analyse en CPG et moyenne des 3 répétitions en PIR pour les échantillons sélectionnés de la population 121-4.



**Figure 14.** Fréquence de distribution des résultats de prédiction pour les 3 répétitions de la population 121-4.

	Population 174-4		
Echantillons	Rép 1 Prédiction	Rép 2 Prédiction	Rép 3 Prédiction
174-4_01	91.62	87.71	92.05
174-4_02	25.95	25.36	21.47
174-4_03	88.62	86.69	87.97
174-4_04	91.87	81.77	90.18
174-4_05	83.30	81.31	79.81
174-4_06	76.04	78.62	74.14
174-4_07	30.72	28.54	27.95
174-4_08	80.73	82.96	80.09
174-4_09	85.07	78.72	81.28
174-4_10	24.08	35.83	23.70
174-4_11	84.71	81.12	93.90
174-4_12	38.63	34.99	32.06
174-4_13	76.85	84.40	75.21
174-4_14	29.36	13.79	28.97
174-4_15	76.64	69.63	80.84
174-4_16	90.13	77.00	85.20
174-4_17	78.85	91.03	90.17
174-4_18	85.53	91.42	89.19
174-4_19	27.20	22.97	25.27
174-4_20	83.02	74.12	79.75
Moyenne	67.44	65.40	66.96
Max	91.87	91.42	93.90
Min	24.08	13.79	21.47
Médiane	79.79	78.67	79.95

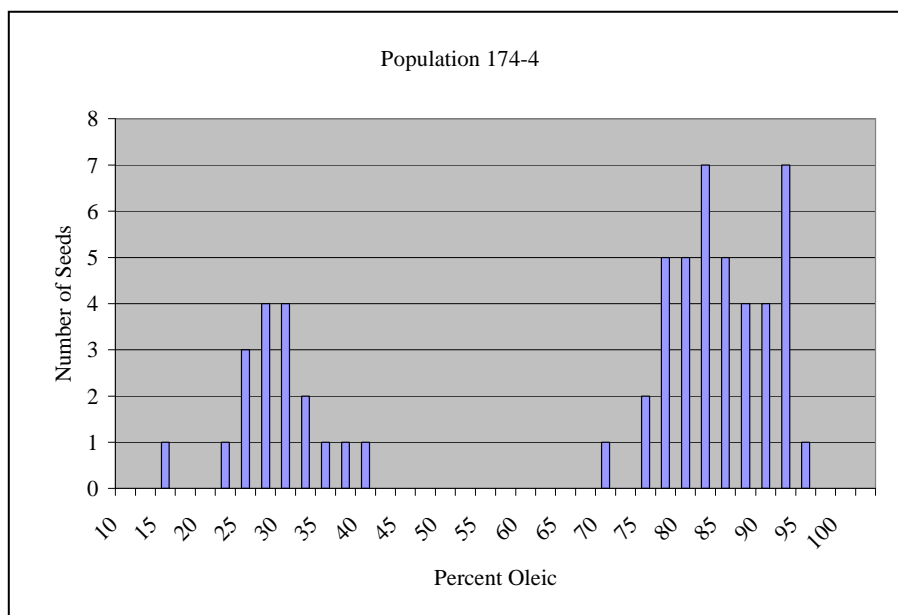
**Tableau 3.** Prédictions et statistiques de la population 174-4.

	Population 187-2		
Echantillons	Rép 1 Prédiction	Rép 2 Prédiction	Rép 3 Prédiction
187-2_01	47.17	58.95	51.53
187-2_02	75.86	85.12	83.50
187-2_03	76.21	75.61	71.26
187-2_04	73.14	73.47	74.24
187-2_05	71.57	75.40	73.62
187-2_06	64.20	83.47	76.50
187-2_07	72.67	73.71	69.57
187-2_08	74.58	77.89	86.13
187-2_09	13.37	13.39	13.33
187-2_10	80.20	73.07	75.85
187-2_11	19.87	15.86	17.99
187-2_12	7.74	17.85	10.36
187-2_13	60.42	54.44	59.83
187-2_14	73.53	67.76	65.56
187-2_15	77.34	71.54	82.83
187-2_16	68.54	84.06	70.92
187-2_17	56.18	61.41	60.52
Moyenne	59.56	62.53	61.38
Max	80.20	85.12	86.13
Min	7.74	13.39	10.36
Médiane	71.57	73.07	70.92

**Tableau 4.** Prédictions et statistiques de la population 187-2.

Population 174-4		
Echantillons	Résultats en CPG	Moyenne PIR des 3 répétitions
174-4_01	84.14	90.46
174-4_02	20.63	24.26
174-4_04	85.03	87.94
174-4_07	18.62	29.07
174-4_09	83.77	81.69
174-4_10	17.72	27.87
174-4_16	86.21	84.11
174-4_17	85.90	86.68
174-4_19	17.21	25.15
Moyenne	55.47	59.69
R <sup>2</sup>	0.990	
SEP	3.702	

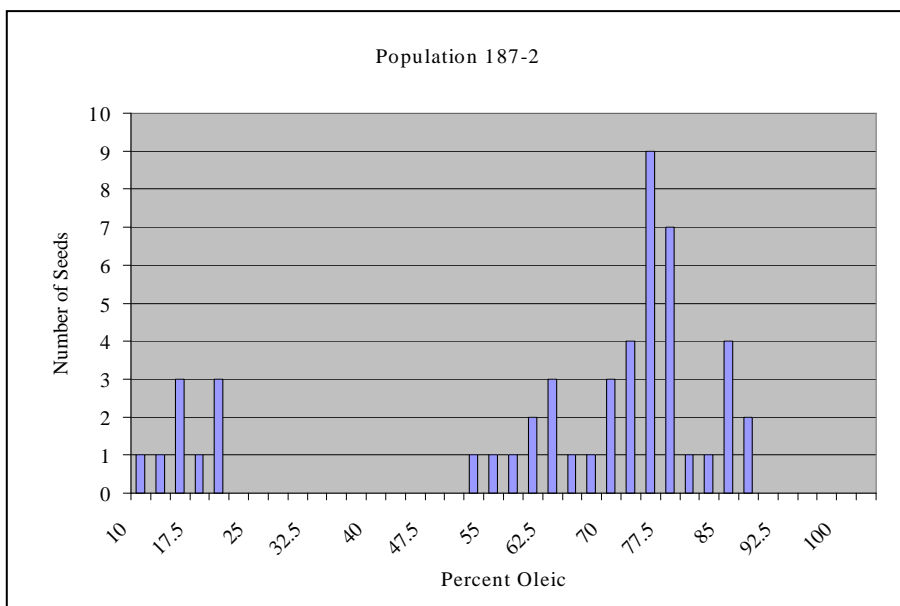
**Tableau 3a.** Analyse en CPG et moyenne des 3 répétitions en PIR pour les échantillons sélectionnés de la population 174-4.



**Figure 15.** Fréquence de distribution des résultats de prédiction pour les 3 répétitions de la population 174-4.

Population 187-2		
Echantillons	Résultats en CPG	Moyenne PIR des 3 répétitions
187-2_01	67.31	52.55
187-2_03	81.35	74.36
187-2_05	78.35	73.53
187-2_08	84.61	79.53
187-2_09	23.95	13.36
187-2_10	82.45	76.38
187-2_11	21.29	17.91
187-2_12	18.57	11.99
187-2_15	85.77	77.24
Moyenne	60.41	52.98
R <sup>2</sup>	0.987	
SEP	3.693	

**Tableau 4a.** Analyse en CPG et moyenne des 3 répétitions en PIR pour les échantillons sélectionnés de la population 187-2.



**Figure 16.** Fréquence de distribution des résultats de prédiction pour les 3 répétitions de la population 187.2.

## 4. Conclusions

Nous avons obtenu une très bonne corrélation de 0.94 et 0.93 pour respectivement la calibration et la validation en PLS1 sur la population de 198 graines.

La SEP avec seulement 2 échantillons retirés, est de 9.9 en unités actuelles d'acide oléique dans la gamme qui va de 13.33 à 90.39 %. La valeur élevée de la SEP est principalement due à l'enveloppe des graines, qui induit un bruit de fond dans les données spectrales.

Toutefois, le gain positif est bien plus important que le négatif, car nous pouvons obtenir des mesures nettement plus précises sans détruire l'enveloppe ou modifier la graine, qui peut maintenant être conservée dans des conditions normales jusqu'à sa plantation.

En se basant sur la forte corrélation entre les valeurs mesurées et prédites pour chacune des 4 populations de validation, il serait sain de procéder à l'analyse en CPG pour tous les échantillons analysés en PIR.

Malgré la forte corrélation pour chaque population, il existe un léger biais.

La population 121-4 a le plus petit offset de 1.1, tandis que les populations 104-2 et 187-2 sont semblables avec respectivement un offset de 6.0 et 8.0. La population 174-4 a un offset négatif de -10.5.

A ce stade, la valeur de l'offset n'est pas complètement comprise, elle peut être due au fait que les échantillons de validation ont été scannés plus tard et à un moment différent de la journée : Les spectres des jeux de calibrations ont été collectés tard dans l'après-midi durant deux semaines, tandis que les spectres pour les jeux de validations ont été collectés tôt le matin.

Il peut y avoir de considérables différences, si on prend en compte la taille et la forme des graines : comme les enveloppes ne sont pas retirées, ce sont des variables à inclure dans la calibration.

Aujourd'hui, les modèles de calibration sont robustes et permettent aux sélectionneurs d'utiliser le SEED MEISTER comme outil de tri non-destructif, efficace, économique et rapide.