

Estimation des proportions variétales de grains de blé par SPIR et imagerie hyper- spectrale

HelioSpir, 29/06/2021

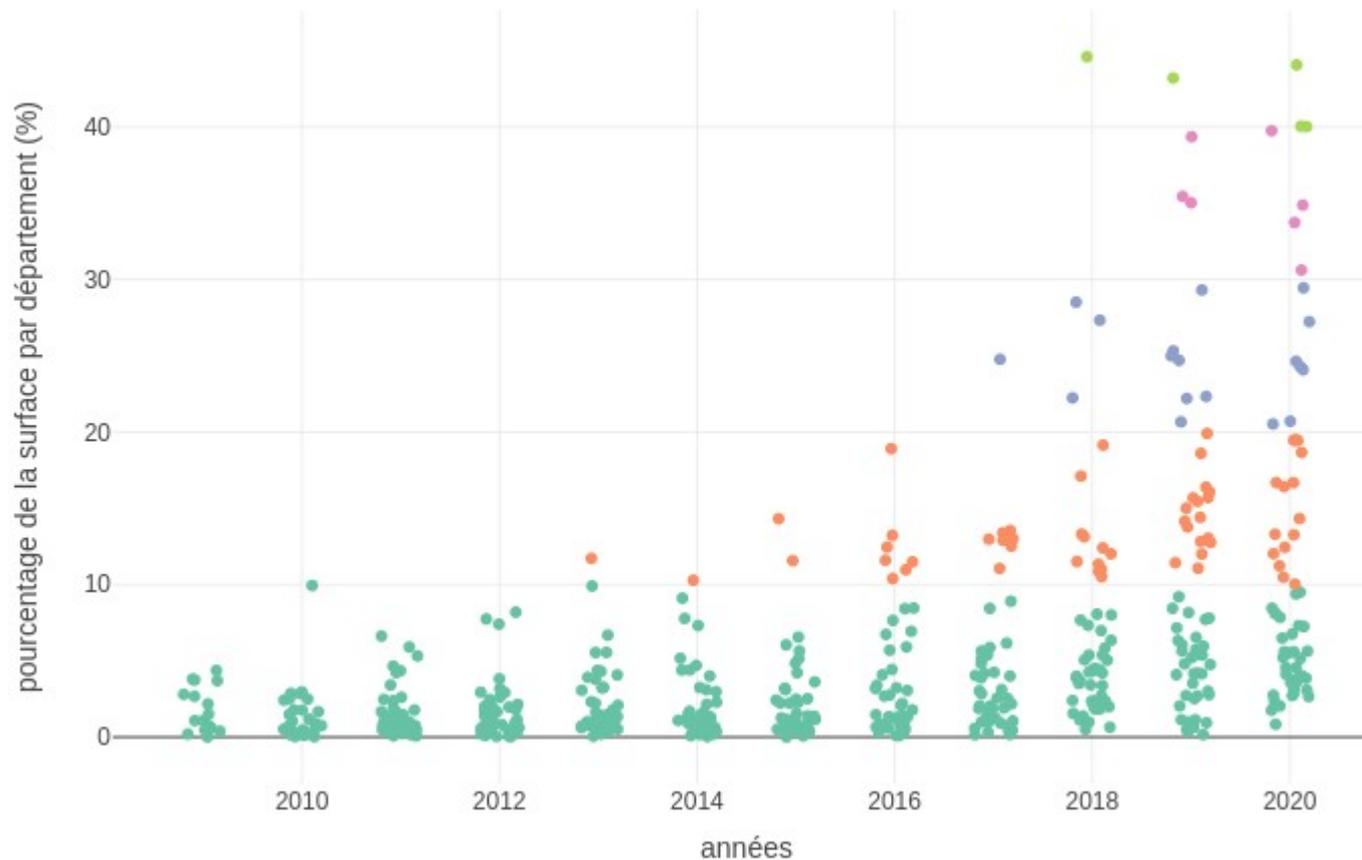
Adrien Belny¹, Martin Ecartot², Timothée Flutre¹

1) **GQE**, Université Paris-Saclay - **INRAE** - CNRS - AgroParisTech, Gif-sur-Yvette

2) **AGAP**, Université de Montpellier, CIRAD, **INRAE**, L'institut Agro, Montpellier

The logo for INRAE, featuring the acronym in a bold, teal, sans-serif font.The logo for AgroParisTech, featuring a stylized green triangle above the text 'AgroParisTech' in a bold, black, sans-serif font. Below the main text, there is a smaller line of text: 'INSTITUT DES SCIENCES ET INDUSTRIES DU VIVANT ET DE L'ENVIRONNEMENT' and 'PARIS INSTITUTE OF TECHNOLOGY FOR LIFE, FOOD AND ENVIRONMENTAL SCIENCES'.

Introduction

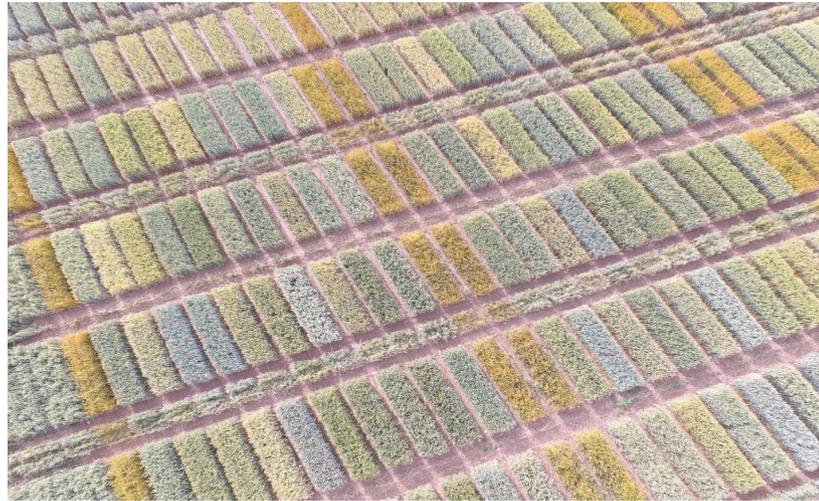


Source des données : FranceAgriMer - Enquête répartition variétale des céréales / Agreste - Statistique agricole annuelle.
Traitement : R. Perronne
<http://moulon.inra.fr/optimix/>

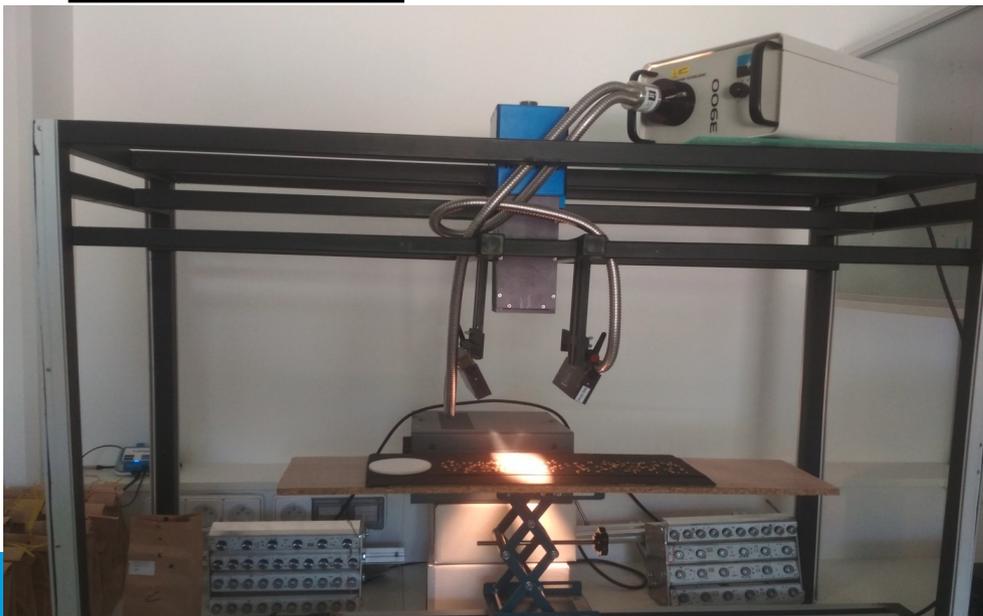
- de plus en plus de blé semé en mélange
- intérêt agroécologique
- **problème** : on sait ce que l'on sème, pas ce que l'on récolte !
- problème des **débouchés**
 - problème en **recherche** (sélection sur l'aptitude au mélange)
- enjeu d'estimer les proportions des grains récoltés

Grains : du champ au spectromètre

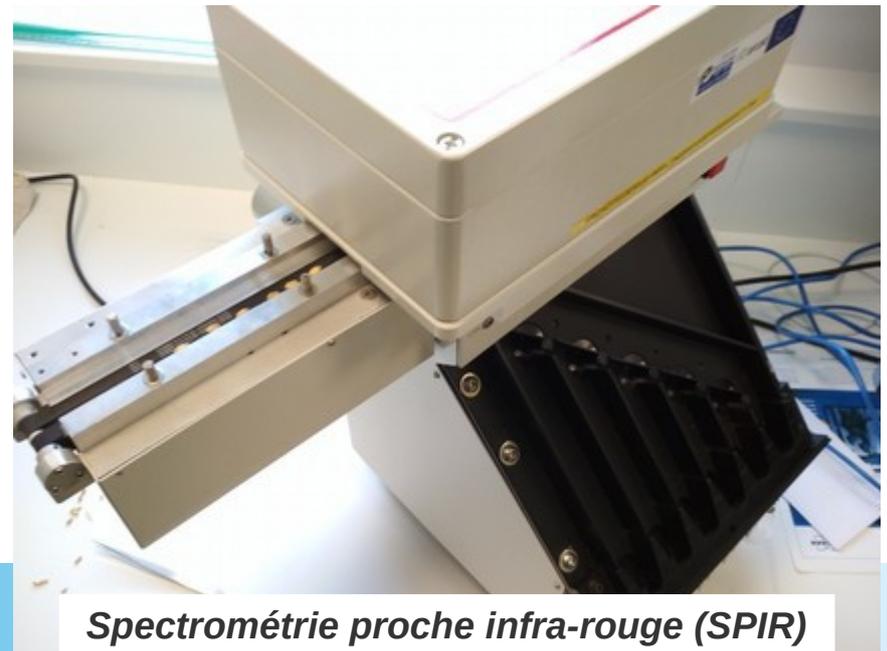
- **Projet Perfomix :**
 - **8 variétés** commerciales
 - 2 micro-parcelles par variété
- **Spectres sur grains individuels :**



Micro-parcelles, Le Moulon (© J. Enjalbert, 2020)

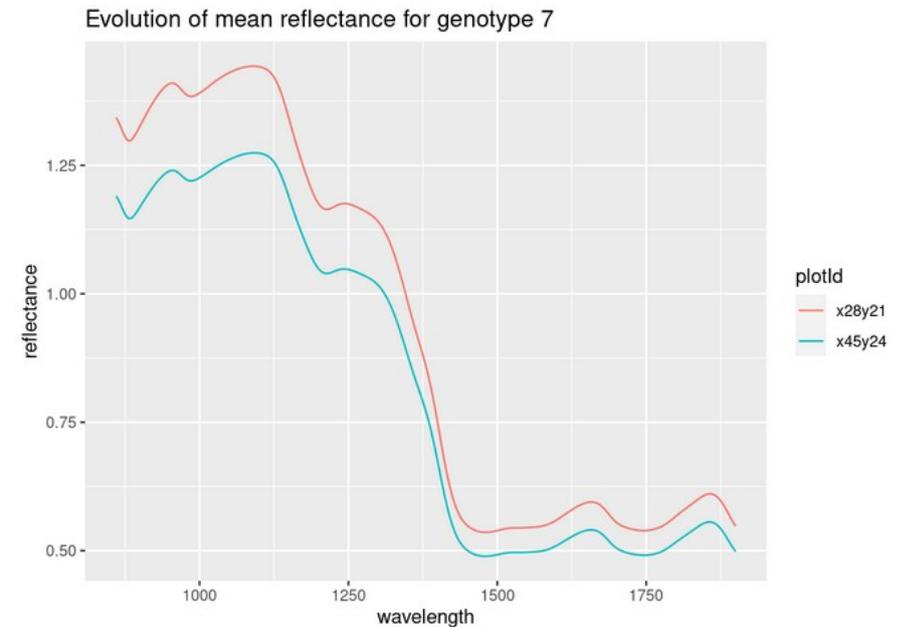
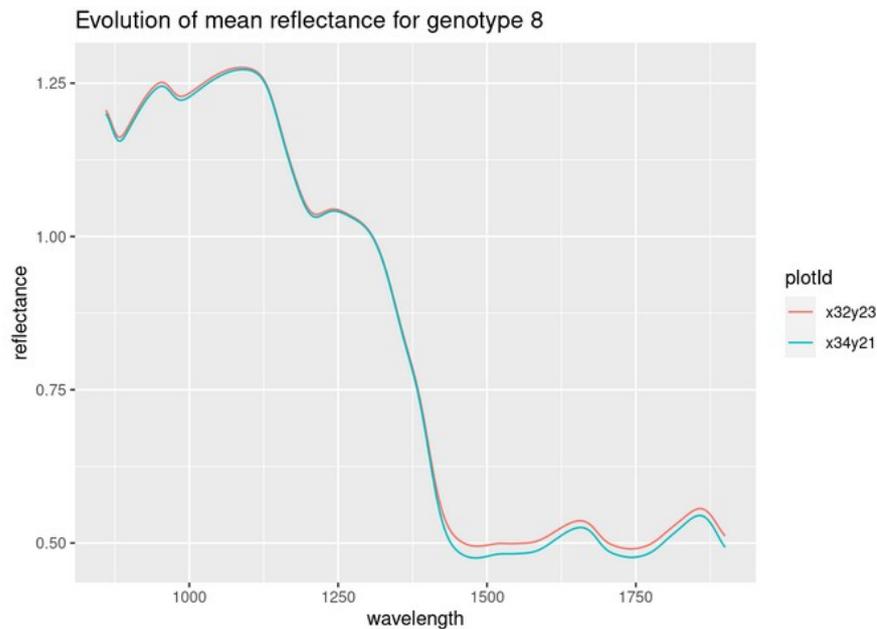


Imagerie hyper-spectrale (IHS)
(400-1000 nm), UMR AGAP (© M. Ecartot, 2021)



Spectrométrie proche infra-rouge (SPIR)
(900-1900 nm), UMR AGAP (© M. Ecartot, 2021)

Variabilité environnementale (ex. SPIR)



Une ANOVA fonctionnelle a été effectuée sur les grains de 2 génotypes différents, chacun ayant été cultivé sur 2 micro-parcelles différentes

- **sensibilité environnementale au sein d'un même génotype possible**
- **l'ANOVA fonctionnelle révèle aussi qu'il n'y a pas toujours de différence inter génotypes**

Evaluation des prédictions

Dataset ex : genos 3-7, prop 50-50



Méthode classique de cross-validation

```
$`3`  
[1] "x44y21" "x43y22"
```

```
$`7`  
[1] "x28y21" "x45y24"
```

Fold 1

Fold 2

Méthode de cross-validation par micro-parcelle

Validation croisée externe :

- 5 folds (tirage aléatoire) → 5 reps
- 2 folds (micro-parcelles) → 4 reps pour faire toutes les combinaisons possibles de plots

Validation croisée interne :

- **estimation des paramètres** du modèle (nombre de composantes pour la PLS, nombre de voisins du KNN etc), 5 folds dans tous les cas

Indicateur d'évaluation des prédictions:

- proportions de grains bien classés
- coefficient de Matthews

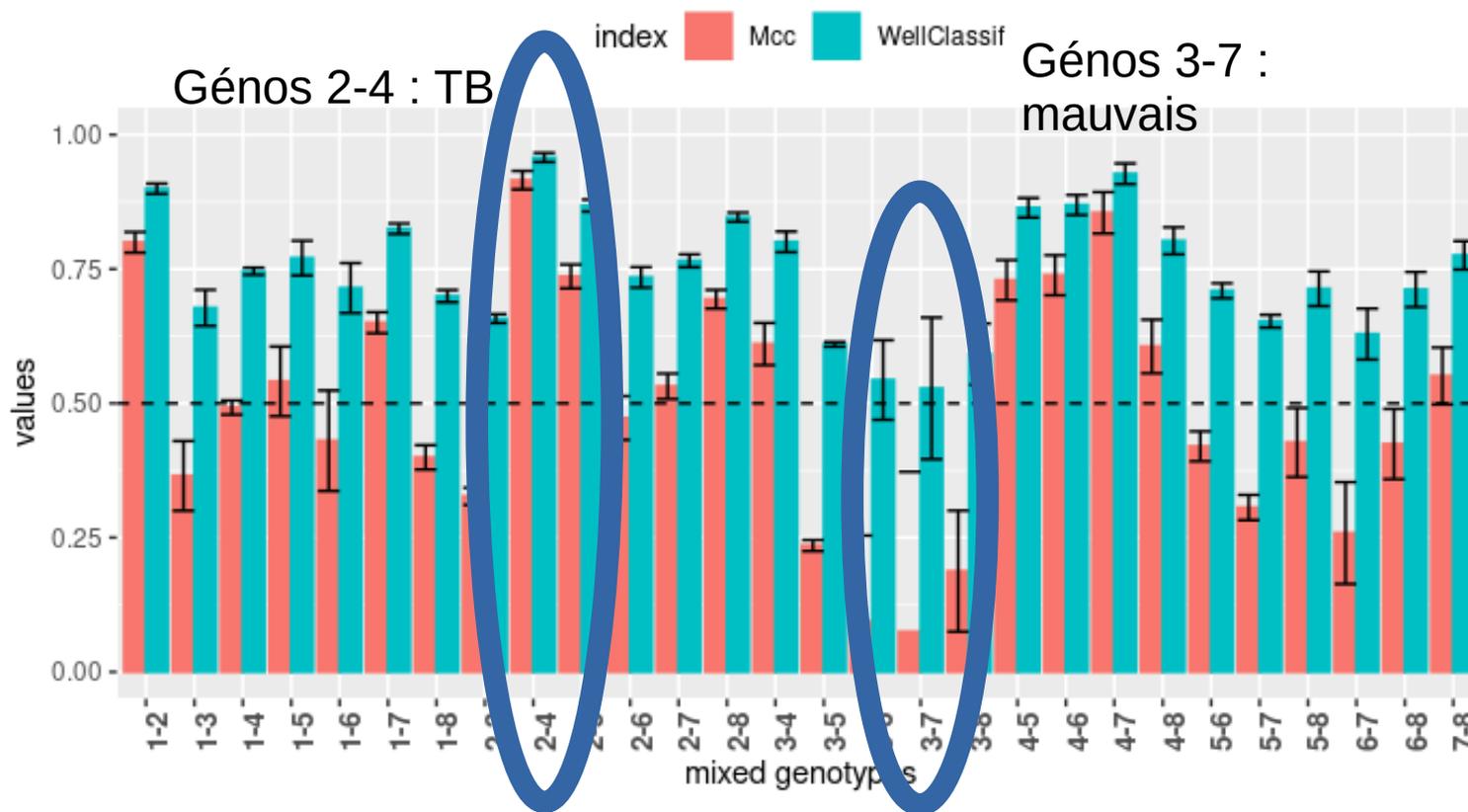
Présentation des modèles utilisés

- **PLS (partial least squares)** → `rnirs`
 - PLS-DA avec `dis` (distance)
 - PLS-DA avec `lm` (modèle linéaire)
- **SVM (séparateurs à vaste marge)** → `rnirs` + `kernlab`
 - noyau gaussien
 - noyau polynomial
- **KNN (k plus proches voisins)** → `rnirs`
- **Réseaux de neurones** → `nnet`
 - 1 couche
- **Régression fonctionnelle multinomiale (mflm)** → `glmnet` + `wavethresh`

Résultats avec SPIR et PLS

Model = plsda1m and valid = microplot

with overall Matthews coef = 0.5 with overall well-classified = 0.75 and nbReps = 4

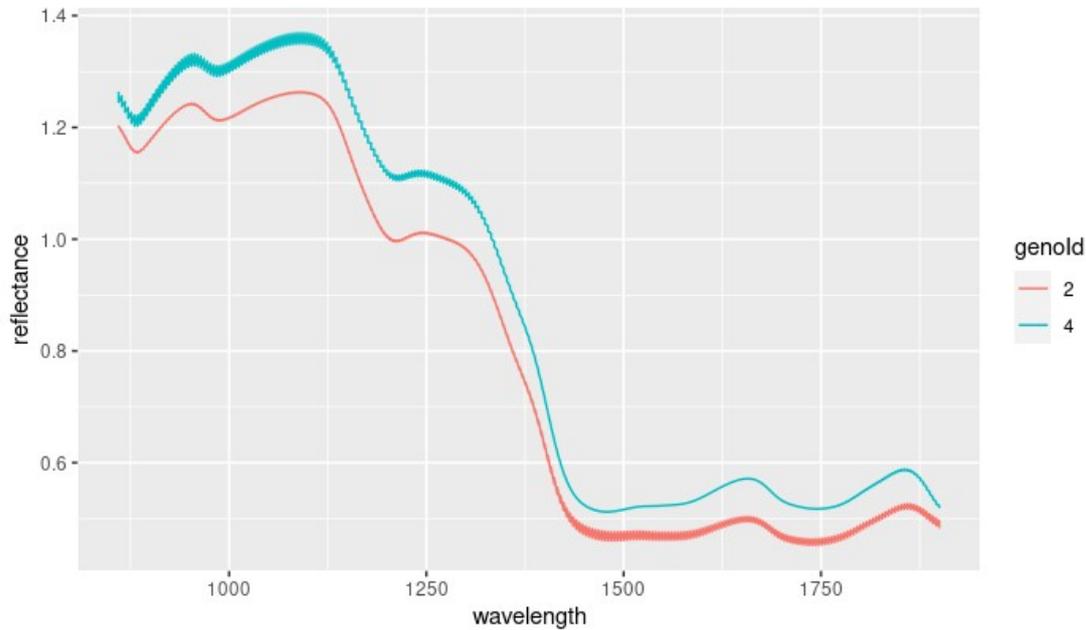


Prédiction des génotypes avec le modèle PLSDALM, coefficient de Matthews général = 0,5, pourcentage général de grains bien classés = 0,75, méthode de validation : microplot

→ PLS-DA 1m : prédictions satisfaisantes mais dégradation des résultats avec +2 variétés mélangées

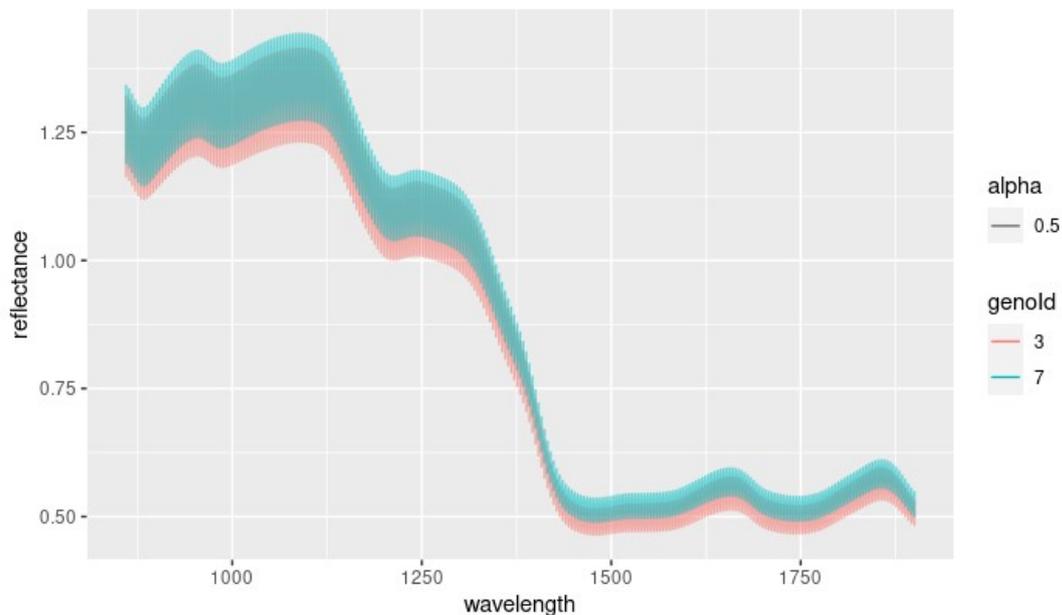
Ex. de spectres (2-4 et 3-7)

Evolution of mean reflectance for genotype 2 & 4



Les courbes ne se chevauchent pas → facile de discriminer les génotypes

Evolution of mean reflectance for genotype 3 & 7



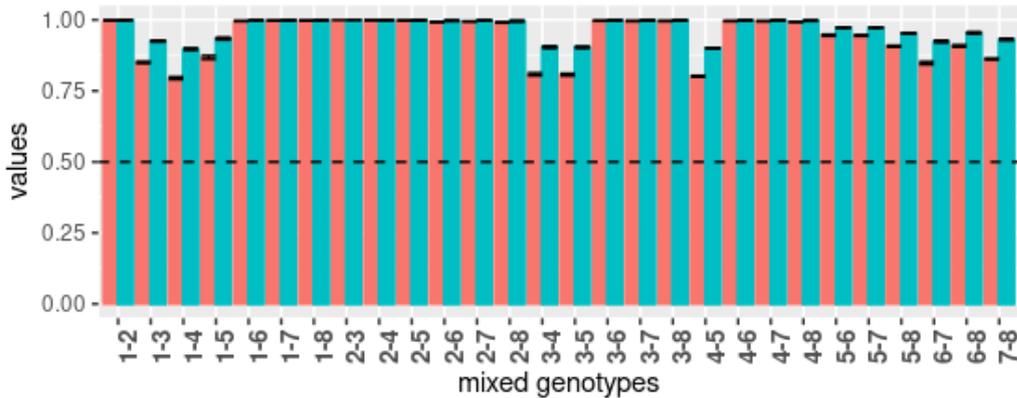
Les courbes se chevauchent → difficile de discriminer les génotypes

Résultats avec IHS et PLS

Model = plsda1m and valid = cv

with overall Matthews coef = 0.94 with overall well-classified = 0.97

index Mcc WellClassif



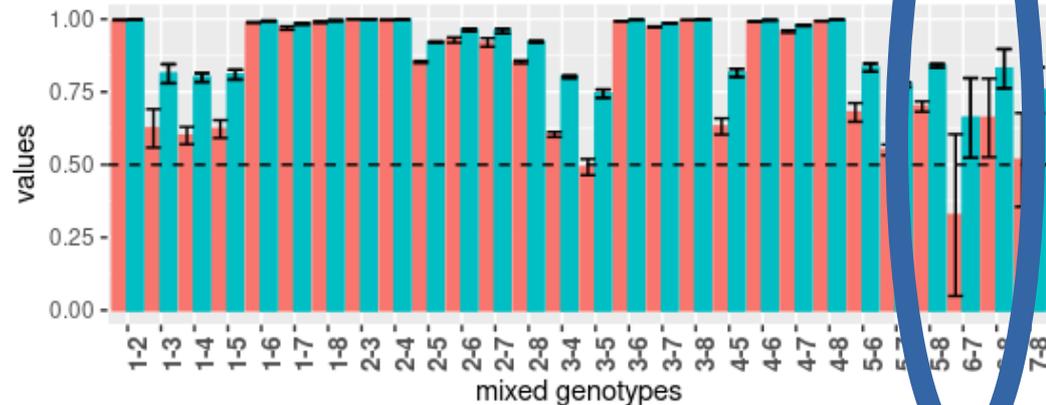
Prédiction des génotypes avec le modèle PLSDALM, coefficient de Matthews général = 0,94, pourcentage général de grains bien classés = 0,97, méthode de validation : cv classique

Tb à part quelques génos (ex 6-7)

Model = plsda1m and valid = microplot

with overall Matthews coef = 0.8 with overall well-classified = 0.9

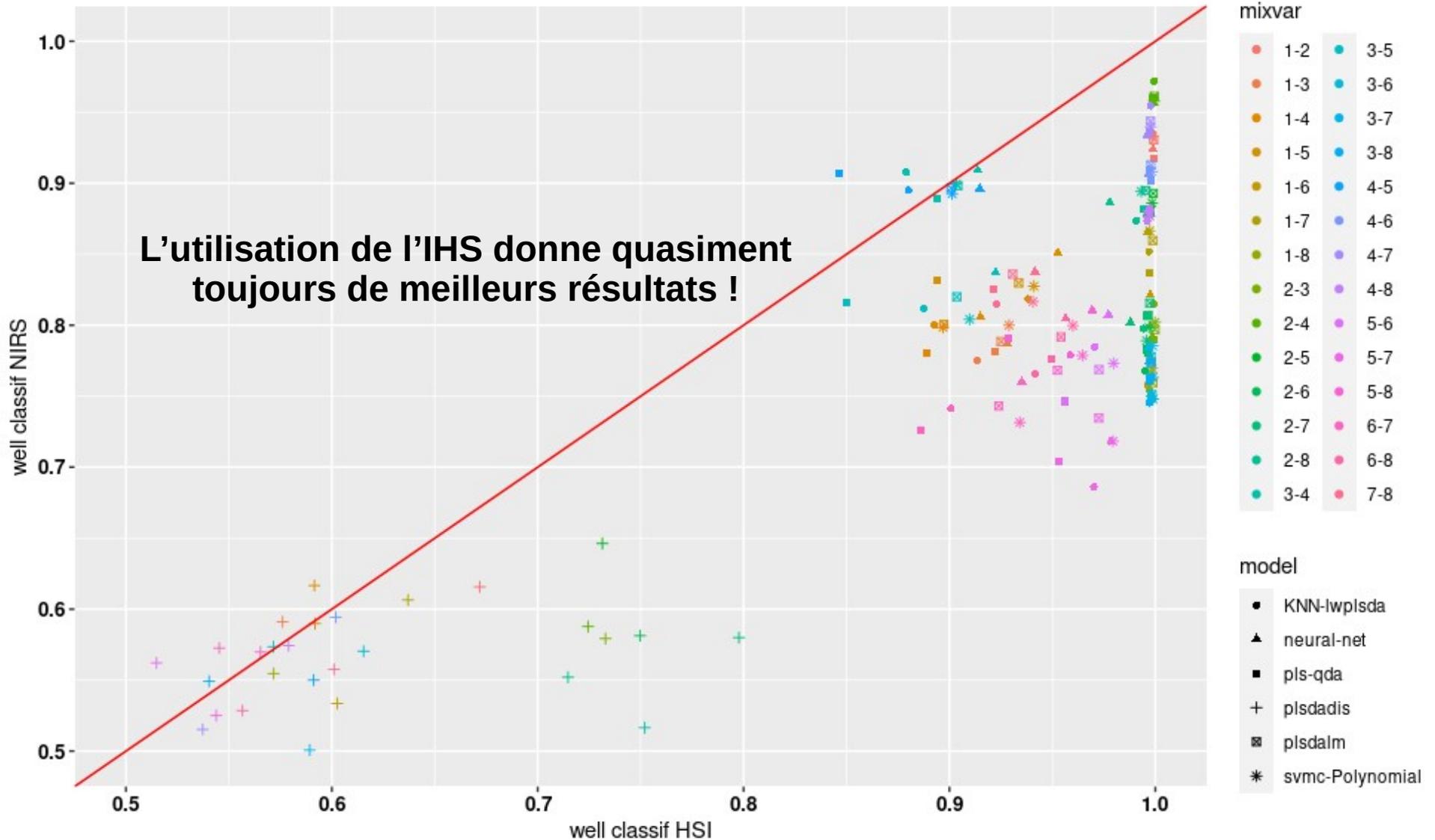
index Mcc WellClassif



Prédiction des génotypes avec le modèle PLSDALM, coefficient de Matthews général = 0,8, pourcentage général de grains bien classés = 0,9, méthode de validation : microplot

IHS vs SPIR

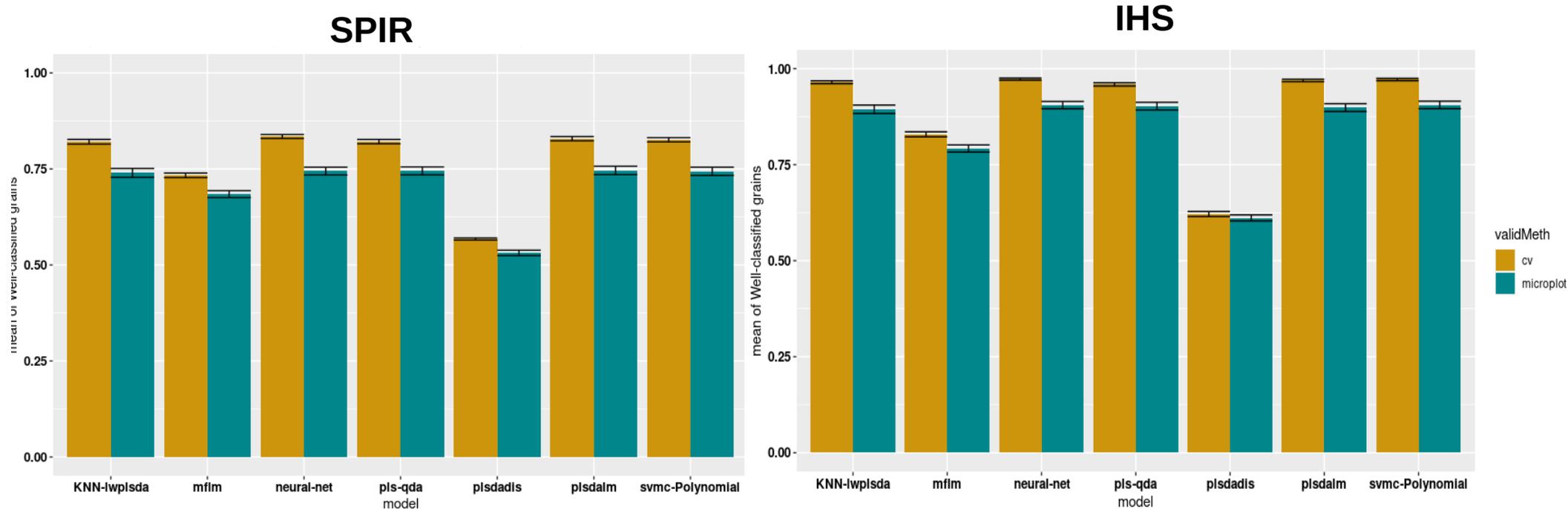
well classific NIRS vs well classific HSI for each mix



1 point par rep pour chaque modèle pour chaque mélange (ici CV classique, 2 variétés)

Comparaison de modèles

(mix de 2 variétés, proportions 50/50)

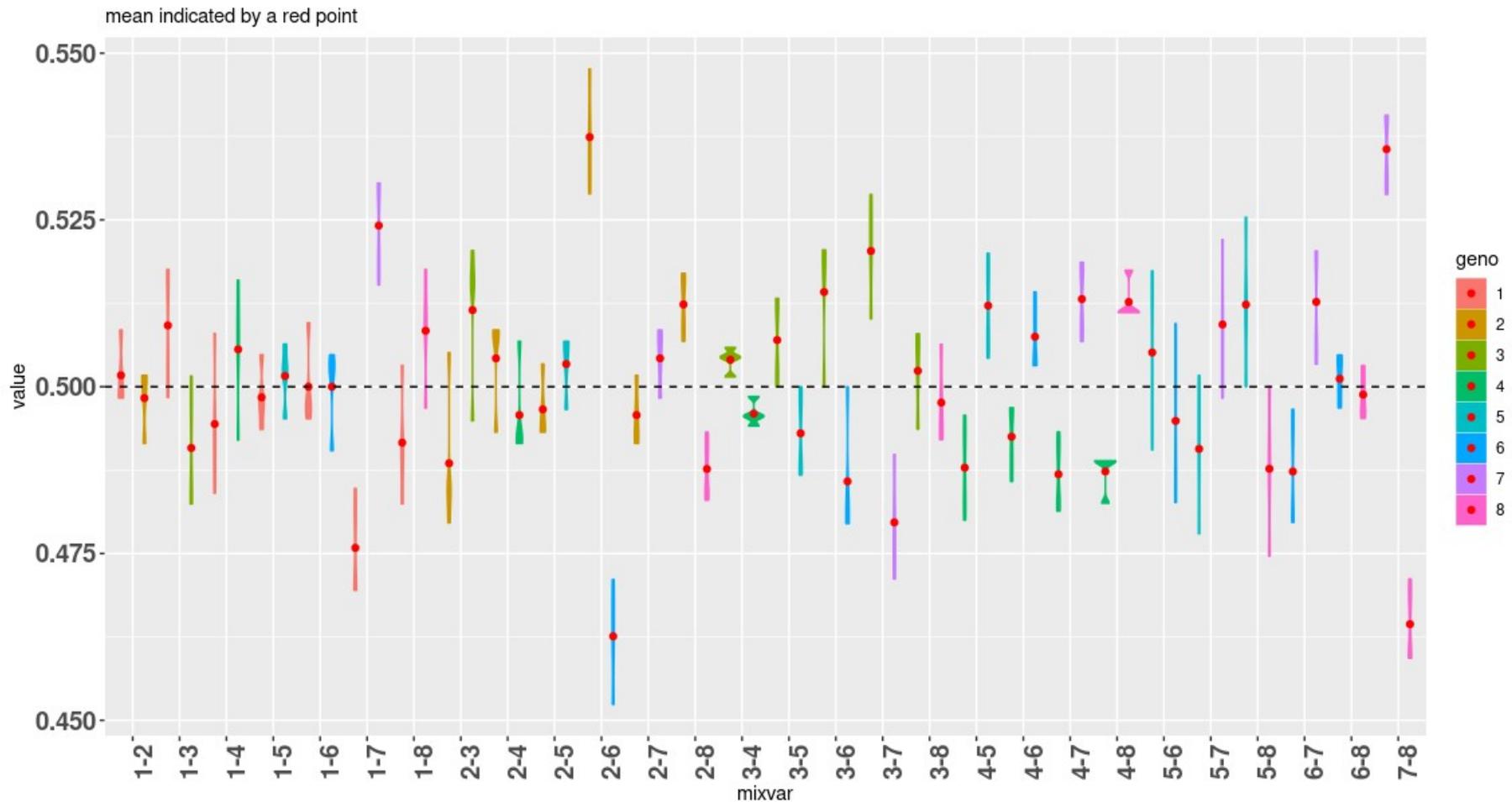


La hiérarchie des modèles est conservée entre SPIR et IHS.

KNN, réseau de neurones, PLS-DA lm & qda et SVM avec noyau polynomial sont des modèles **robustes**.

Temps de calcul : les méthodes **PLS** sont les **plus rapides** !

Estimation des proportions

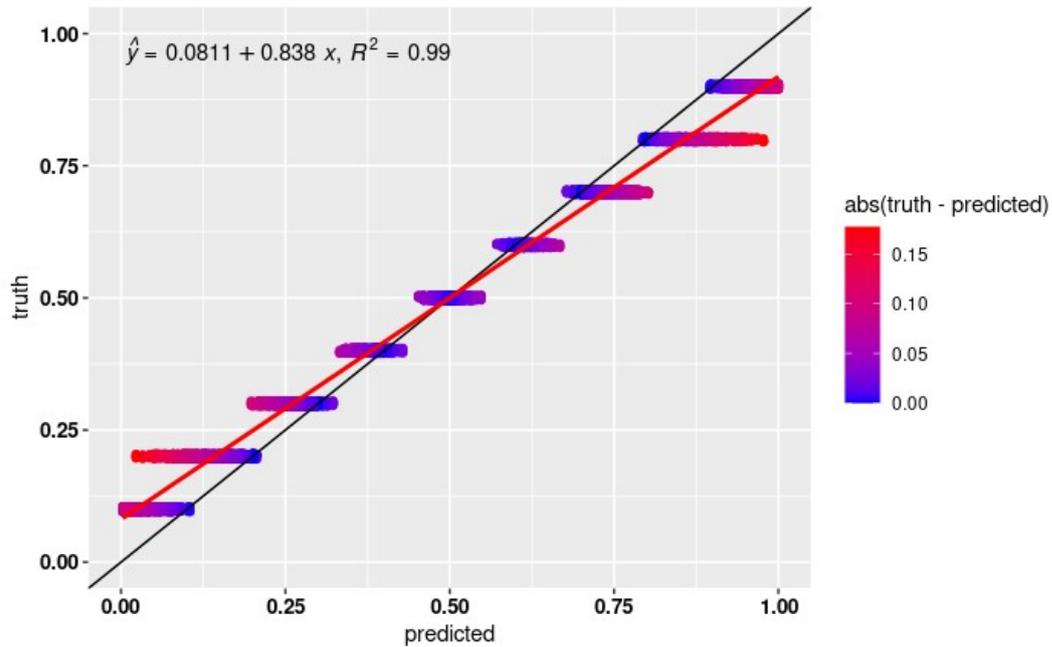


Estimation de proportions pour PLS-DA lm, cv classique, SPIR, 5 reps

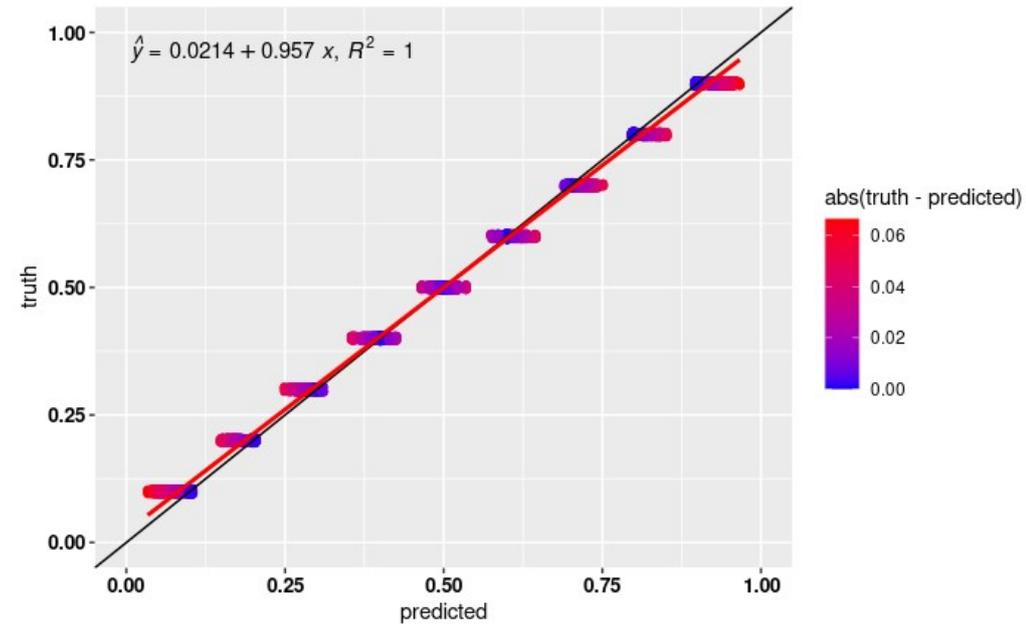
$$\hat{p}_i = \frac{\text{nb of grains predicted to be from genotype } i}{\text{total number of grains}}$$

Proportions non-équiprobables

PLS-DA Im, SPIR, 2 var, cv classique



PLS-DA Im, IHS, 2 var, cv classique



5 reps pour chacune des 28 combinaisons possibles de mélange (5*28 points)

L'utilisation de l'**IHS** permet de **bonnes estimations**.
Quand les **proportions** deviennent très **déséquilibrées**, le **biais augmente** !

Autres résultats et perspectives

- mélanges ternaires (annexe) et quaternaires
- caractéristiques morphologiques (IHS)
- corrélation spatiale avec la teneur en azote des grains (annexe)
- Calcul de l'intervalle de confiance autour de proportions
- prédiction sur des grains issus de parcelles en mélange, et effet de l'année
- Rédaction d'un article pour publication

Remerciements

- Collègues de GQE : Jérôme Enjalbert, Meije Gawinowski, Gaël Vieceli, Didier Tropée, Tatiana Randrianarisoa, A. Hospital et M. Colas
- Collègues d'AGAP : Frédéric Compan
- Financement : INRAE (projet PerfoMix)

Merci de votre attention !

**Je recherche actuellement une thèse/un emploi
autour de la SPIR et de la génétique quantitative**

adrien.belny@agroparistech.fr

adrien.belny@inrae.fr