

La sélection phénomique : Quand les généticiens s'essaient à la NIRS

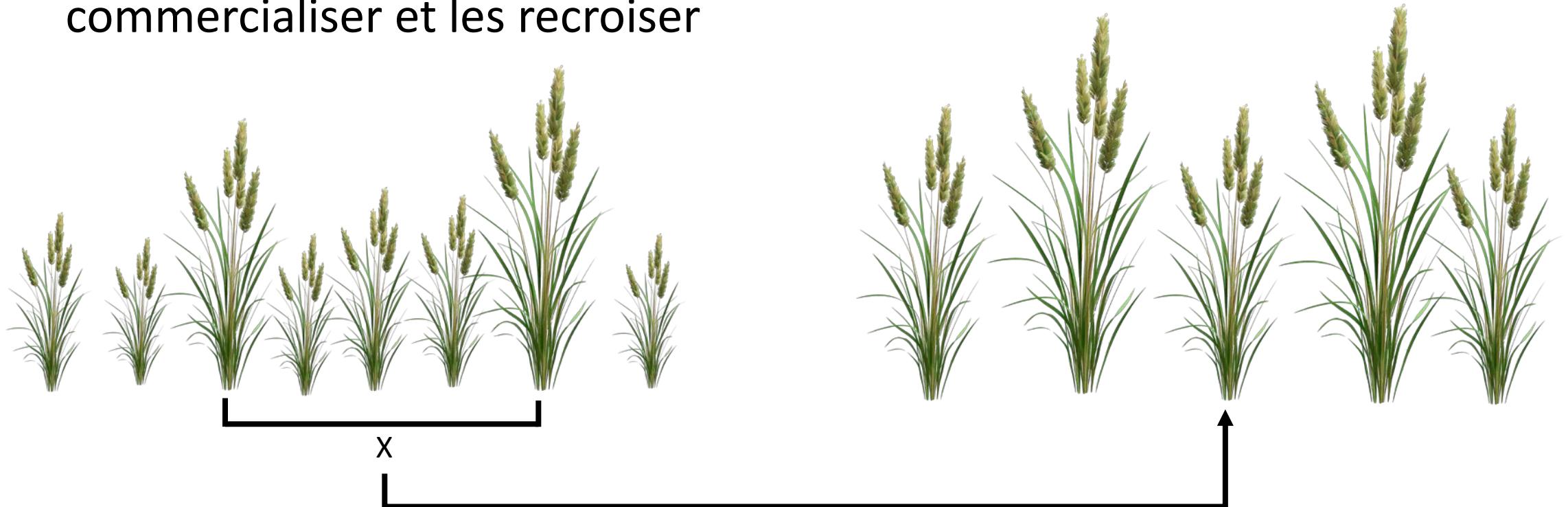


Clément Bienvenu 25/06/2025



Contexte : la sélection variétale

- Croiser des "bonnes" plantes entre elles pour obtenir une descendance améliorée
- Choisir les "meilleurs" plantes de la descendance pour les commercialiser et les recroiser



Problématique : comment identifier les "meilleures" plantes ?



Phénotypage :

Observation et mesures manuelles dans des essais au champ



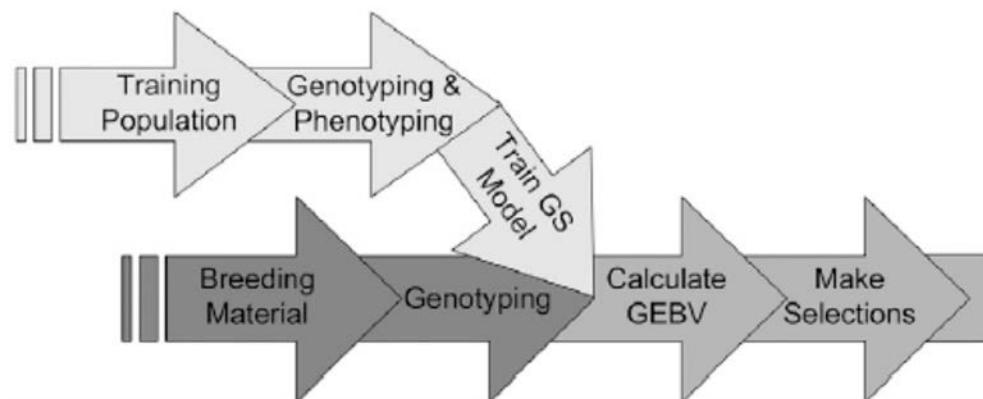
Prédiction génomique :

- Séquençage de l'ADN des plantes
- Utilisation d'un modèle prédictif



Prédiction Phénomique :

- Acquisition NIRS sur plantes
- Utilisation d'un modèle prédictif

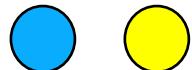


Quelle différence avec une calibration NIRS ?

Calibration NIRS vs sélection phénotomique

- Etablissement du phénotype :

Plante 1



Pousse dans
environnement 1



Plante 2



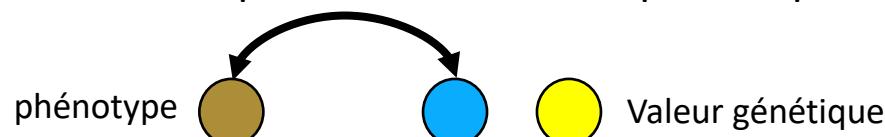
Pousse dans
environnement 2



$$P_1 = a_1 + a_2 + e_1 = \text{brown circle}$$

$$P_2 = a_1 + a_3 + e_2 = \text{purple circle}$$

Héritabilité = précision maximale que l'on peut obtenir



- Comparaison :

➤ Calibration NIRS

"Je veux prédire plante 1 = brun et
plante 2 = gris/violet"

➤ Phénomique

"Je veux prédire plante 1 = bleu + jaune
et plante 2 = bleu + rouge"

Le modèle de prédiction : GBLUP

$$\underline{y} = X\beta + Z\underline{u} + \varepsilon$$

Matrice d'incidence des génotypes
↓
Effets fixes non génétiques Vecteur aléatoire des valeurs génétiques

Avec $\underline{u} \sim N(0, K\sigma_g^2)$ et $\varepsilon \sim N(0, I\sigma_\varepsilon^2)$

$$\begin{bmatrix} \hat{\beta} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X'X & X'Z \\ Z'X & ZZ' + \frac{\sigma_\varepsilon^2}{\sigma_g^2} K^{-1} \end{bmatrix}^{-1} \begin{bmatrix} X'y \\ Z'y \end{bmatrix}$$

K = Matrice d'apparentement calculée à partir des matrices génotypiques ou des spectres

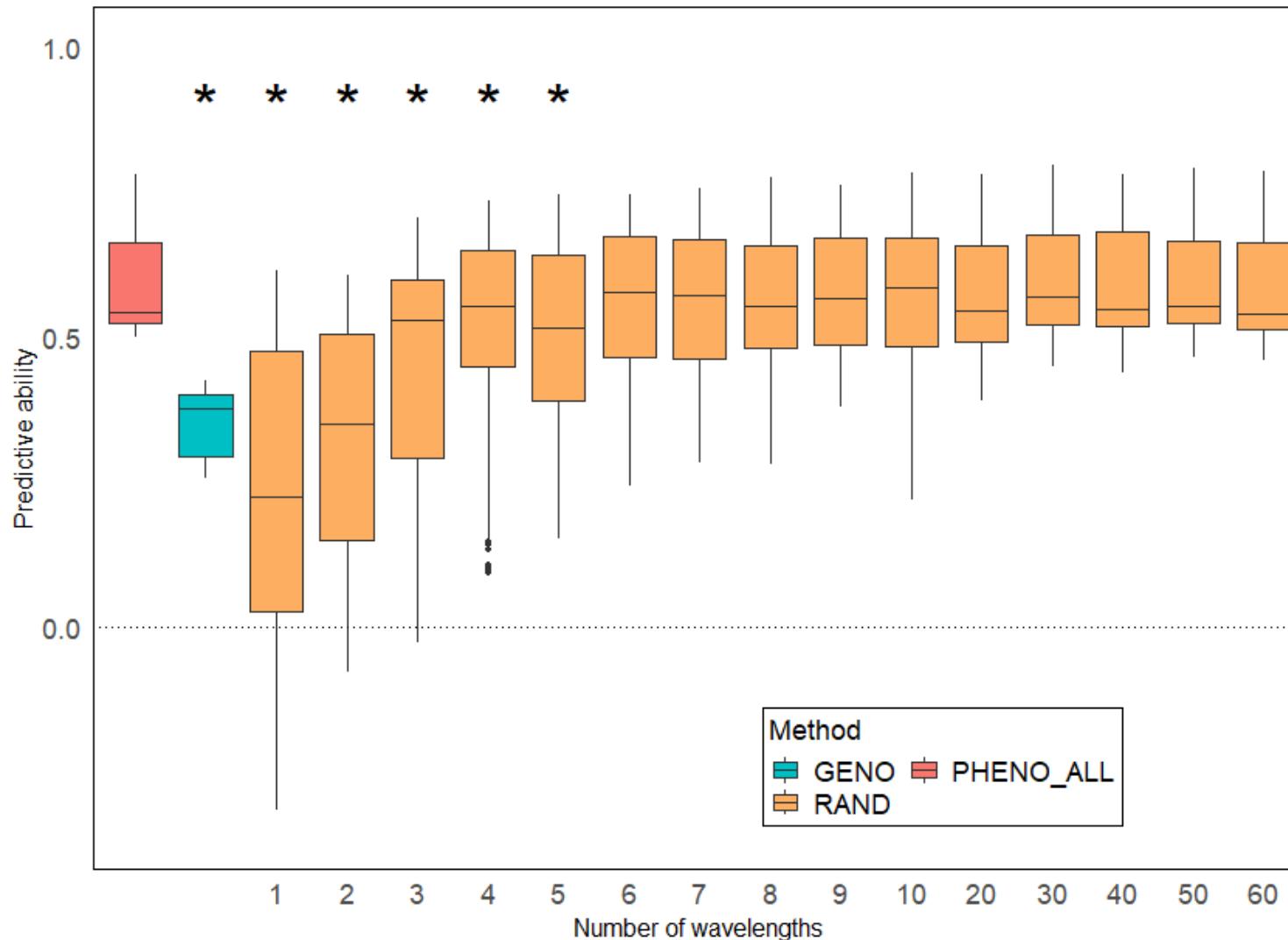
Calcul de K avec des spectres :

$$X \xrightarrow{\text{Centrage et réduction longueur d'onde par longueur d'onde}} S \xrightarrow{\text{Comparaison des individus longueur d'onde par longueur d'onde}} K = \frac{SS^T}{p}$$

Matrice spectrale de dimension (n,p) Matrice spectrale centrée réduite

- Pas de prise en compte des corrélations/dépendances entre les longueurs d'ondes
- Pas d'analyse sur les composantes du signal

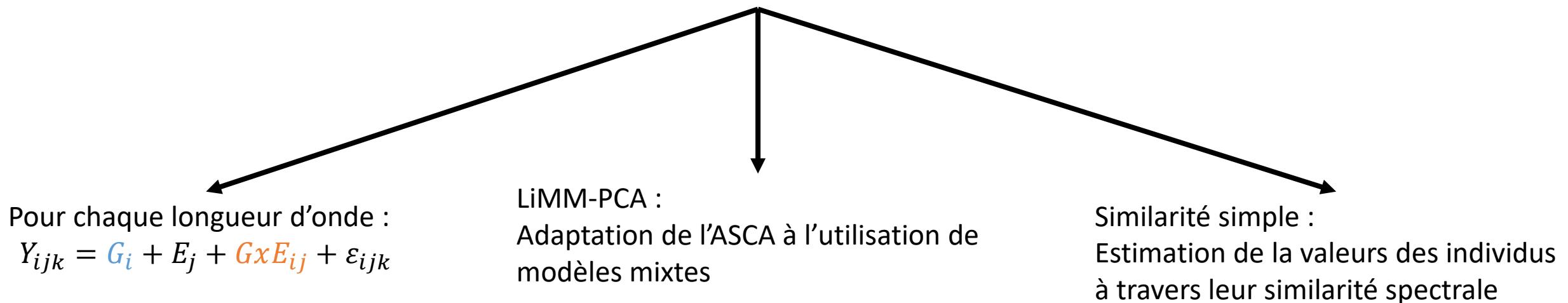
Comment mieux utiliser les spectres ? (1/2)



- Peu d'information dans le spectre et 5 longueurs d'ondes suffisent ?
 - Mauvaise utilisation de l'information disponible dans les 1154 longueurs d'onde ?
- Résultats liés à la dépendance/correlations entre longueurs d'ondes qui n'est pas exploitée ?

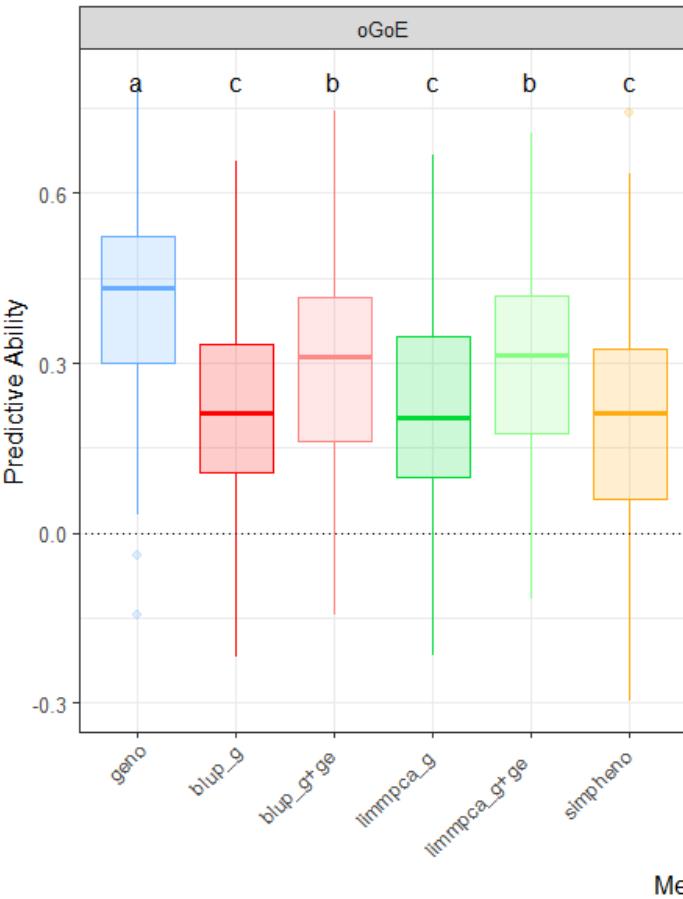
Comment mieux utiliser les spectres ? (2/2)

- Démélanger les parties génétique et environnementale pour mieux les utiliser dans les modèles de prédiction
- Mieux prédire la réaction spécifique d'un génotype à un environnement

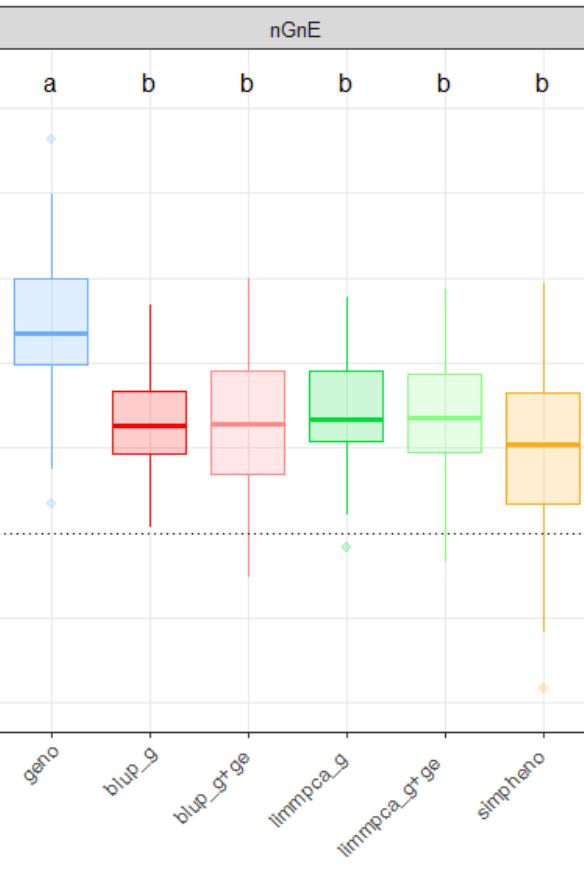


Prédictions après "démélange"

Génotypes et environnements communs entre jeux d'entraînement et validation



Génotypes et environnements différents entre jeux d'entraînement et validation



- Prédiction génomique meilleure dans tous les scénarios
- Scénarios favorables pour prédire le GxE :
 - Pas de différence entre décomposition G seule et sans décomposition
 - Meilleurs prédictions en décomposant G et GxE
- Scénarios défavorables :
 - Pas de différence entre toutes les méthodes de prédiction phénotypique
- Pas de différences entre modèle de décomposition simple et LiMM-PCA

Quelles limites à cette approche ?

- Héritabilité des spectres limite la capacité à démélanger
- Démélanger des signaux non chimiquement purs

Perspectives

- Test sur d'autres jeux de données
- Orthogonalisation pour retirer les composantes non désirées
- MCR-ALS pour démélanger plus efficacement ?

Merci !