Exploring NIRS Application:

Fundamentals of Phenomic Prediction and Linear Mixed Models with R

Jemay SALOMON

PhD candidate
jemay.salomon@inrae.fr
Université Paris-Saclay

Linear Mixed Models

Extension of linear models

Allows incorporation of both random and fixed effects

Used to assess genetic values

…

The linear mixed model

……

- Expériemental design :

250 sorghum génotypes
X2 blocks

Scalar form :

$$y_{ikr}^{(s)} = \mu^{(s)} + \alpha_k + a_{G,i} + \epsilon_{ikr}^{(s)}$$

Yikr : Yield of genotype i (sorghum) in block k and repetition r

μ: intercept (mean yield of the population)

αkαk: block effects modeled as fixed

aG,i: genotype effects modeled as random

ϵikr: residuals

The linear mixed model

……

Matrix form :

$$\mathbf{y} = X\mathbf{b} + Z\mathbf{a}_G + \epsilon$$

y: vector of yield

X: Incidence matrix for fixed effects (intercept µµ, blocks)

b: vector of fixed effects

Z: Incidence matrix for random effects

aG: vector of random effects (genetic values)

e: vector of residuals

With :

$$\epsilon \overset{iid}{\sim} N_N\left(\mathbf{0}, \sigma^2 I\right)$$

$$\boldsymbol{a}G \sim N_N\left(\mathbf{0}, \sigma_G^2 I\right)$$

- Tests :

H0 : $\sigma_G^2 = 0$

H1 : $\sigma_G^2 > 0$

What interest ?????

.....

Artificial selection (Timothée Flutre/INRAE)

A breeder's job to reach her target:

- choose which genotypes to keep
- choose which genotypes to cross

and repeat → population "improvement"

Goal: obtain genetic gain → genetic "progress"

- efficiently
- in the short term but also in the long term

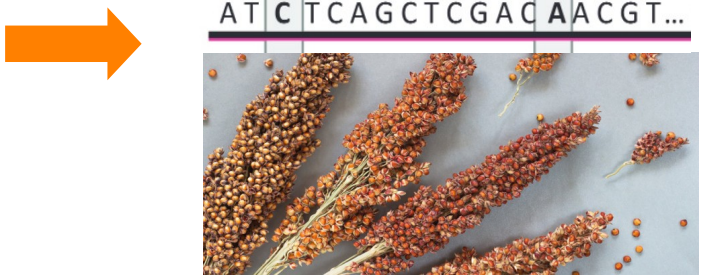
Key points:

- value genotypes for their expected progeny
- maintain genetic diversity

Used of mixed models in Genomic Prediction

- Principle of genomic prediction or selection

Reference population



Genotyping

candidate population

Genotyping

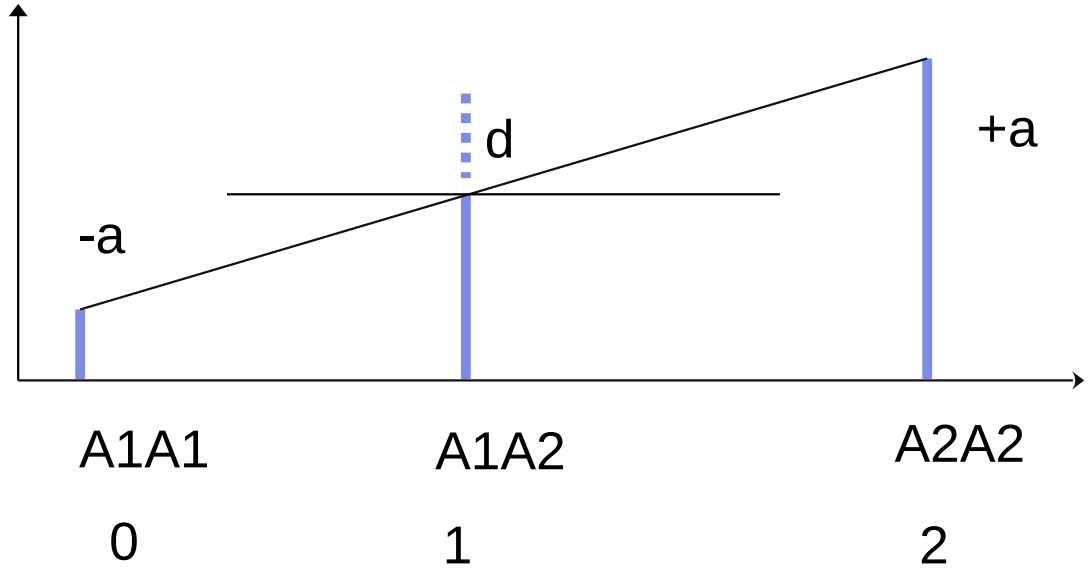
Statistical models for genomic prediction

Used of mixed models in Genomic Prediction

- Principle of genomic prediction or selection

Statistical models for genomic prediction

Genotyping



Used of mixed models in Genomic Prediction

- Principle of genomic prediction or selection
  - Model : $y_i = \mu + \sum_{p=1}^{P} x_{p,i}\beta_p + \epsilon_i$ (mod1)

$$y_i = \mu + \sum_{p=1}^{P} x_{p,i}\beta_p + \boldsymbol{u_i} + \epsilon_i \quad \text{(mod2)}$$

$$\boldsymbol{y} = 1\mu + X\beta + \boldsymbol{\epsilon} \quad \text{(mod3)}$$

$$\boldsymbol{y} = 1\mu + \boldsymbol{a} + \boldsymbol{\epsilon} \quad \text{(mod4)}$$

y : vector of genotype yield

1µ : fixed parameter representing the global mean yield of génotypes

a : breeding values

e :residuals

$$\epsilon \overset{iid}{\sim} N_N(\mathbf{0}, \sigma^2 I)$$

$$\boldsymbol{a} \sim N_N(\mathbf{0}, \sigma^2 A_{mark})$$

With :

$$\hat{a} = X\hat{\beta}$$

Used of mixed models in Genomic Prediction

- Principle of genomic prediction or selection

Then :

Suppose cov(Xβ, ϵ) = 0 $\Rightarrow$ $var(\boldsymbol{y}) = \sigma^2_\beta XX^T + \sigma^2 I$

$$\sigma^2_\beta XX^T = \sigma^2 A_{mark}$$

A_{mark} Variance-covariance matrix incorporating the relatedness between individuals

- **genetically related individuals are more likely to share alleles at causal loci, and thus have similar phenotypes.**

Used of mixed models in Genomic Prediction

- Principle of genomic prediction or selection

$$\boldsymbol{y} = 1\mu + \boldsymbol{a} + \boldsymbol{\epsilon}$$

Keys point of this model (GBLUP) :

Primarily applied in scenarios with a polygenic architecture.

Assumes all markers have non-null effects with equal variance.

Facilitates accurate estimation of marker effects and variance (var(β)).

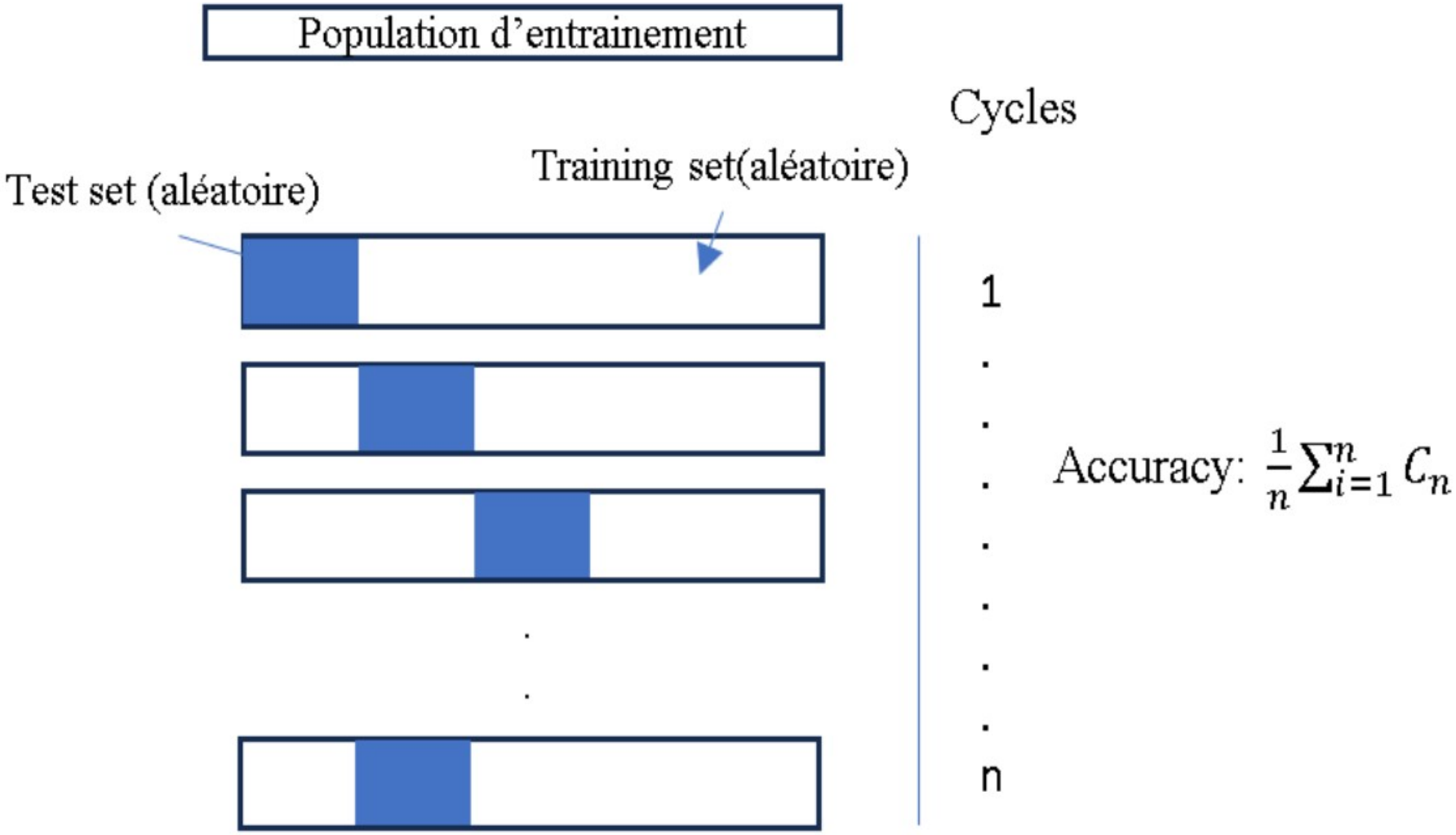
Predicts additive genotypic values (â).

Estimates the additive genetic component of variance (var(a)).

Estimates narrow-sense heritability. : $\hat{h}^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma^2}$

Used of mixed models in Genomic Prediction

- Principle of genomic prediction or selection : Training models



What it looks like in real-world applications :

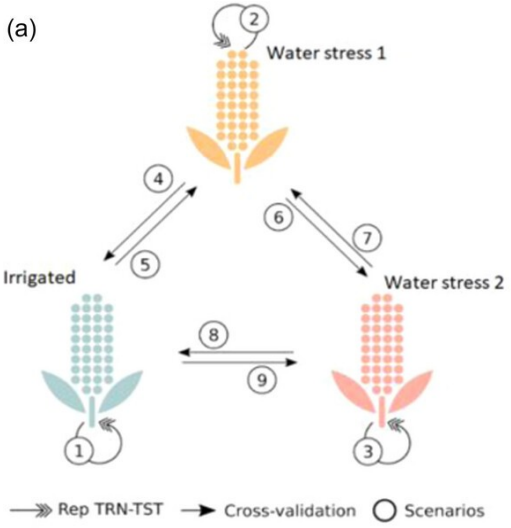
Crop Science

CSSA
Crop Science Society of America

ORIGINAL ARTICLE | Open Access

Genomic prediction of sweet sorghum agronomic performance under drought and irrigated environments in Haiti

Jean Rigaud Charles, Marie Darline Dorval, Jean Bernard Durone, Luis Felipe Ventorim Ferrão, Rodrigo Rampazo Amadeu, Patricio Ricardo Munoz, Geoffrey Morris, Geoffrey Meru, Gael Pressoir



(b)

	Stratified prediction			Cross prediction					
Plant Height	0.56	0.45	0.60	0.48	0.48	0.14	0.16	0.22	0.26
Stem Diameter	0.62	0.52	0.48	0.60	0.50	0.35	0.36	0.46	0.39
Heading	0.57	0.53	0.64	0.56	0.54	0.32	0.29	0.31	0.30
Maturity	0.58	0.53	0.30	0.57	0.53	0.18	0.23	0.23	0.18
Leaf Number	0.57	0.54	0.55	0.53	0.49	0.55	0.50	0.58	0.56
Green Leaf	0.58	0.58	0.53	0.54	0.51	0.54	0.53	0.58	0.55
Stem Number	0.38	0.41	0.49	0.32	0.35	0.37	0.33	0.35	0.44
Total Soluble Solids	0.71	0.64	0.67	0.65	0.58	0.59	0.59	0.68	0.64
Juice Weight	0.58	0.59	0.41	0.53	0.54	0.30	0.38	0.35	0.30
Stem Weight	0.60	0.59	0.57	0.56	0.54	0.36	0.34	0.43	0.45
Leaf Weight	0.56	0.53	0.63	0.54	0.49	0.47	0.37	0.50	0.57
Grain Yield	0.55	0.43	0.37	0.32	0.34	0.08	0.08	0.13	0.13
	1	2	3	4	5	6	7	8	9

Overall, stratified prediction demonstrated higher predictive performance than cross-prediction.

What it looks like in real-world applications :

Optimisation des modèles de prédiction génomique de la performance du sorgho (Sorghum bicolor) pour de nouveaux environnements : analyse destabilité et facteurs physiologiques influents, prédiction inter et intra environnementale, intégration de la composante GXE, et approche phénomique. (Jemay SALOMON – Memoire- M2)

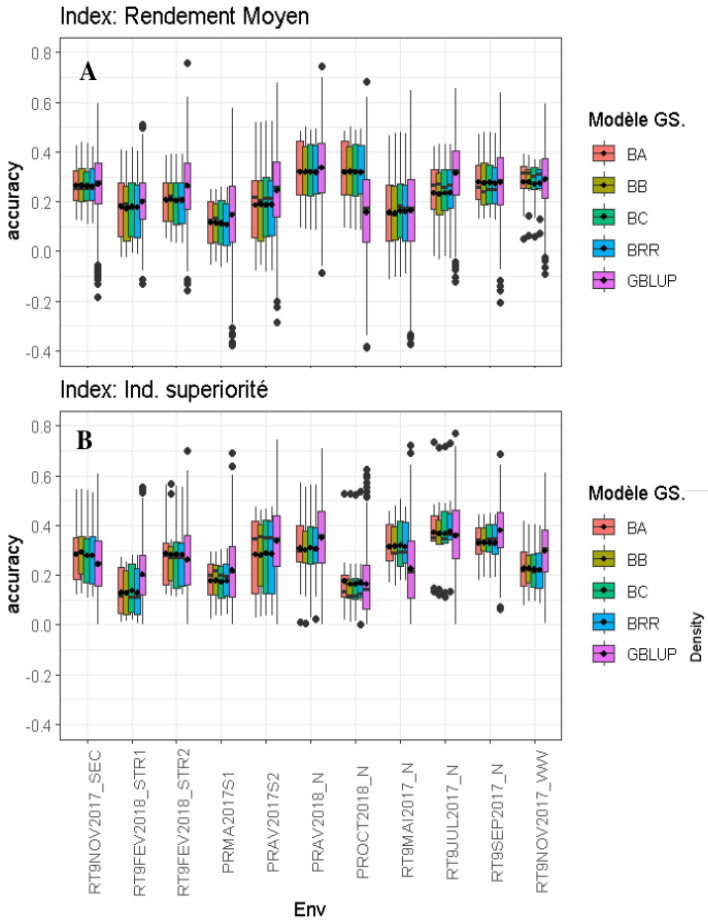


Figure 19: Comparaison de la prédiction inter-environnementale pour les index de sélection : rendement moyen (A) et indice de supériorité (B)

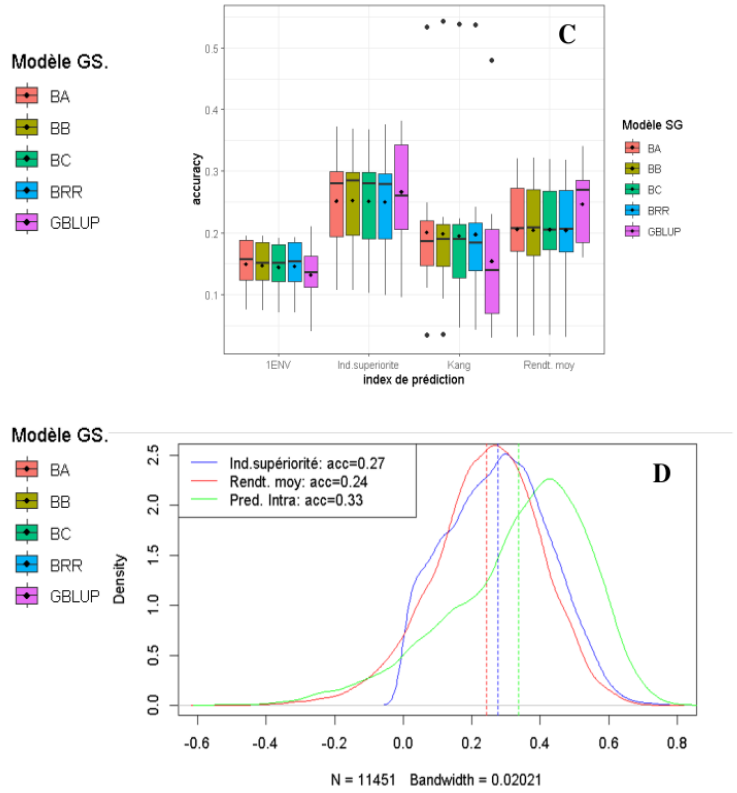
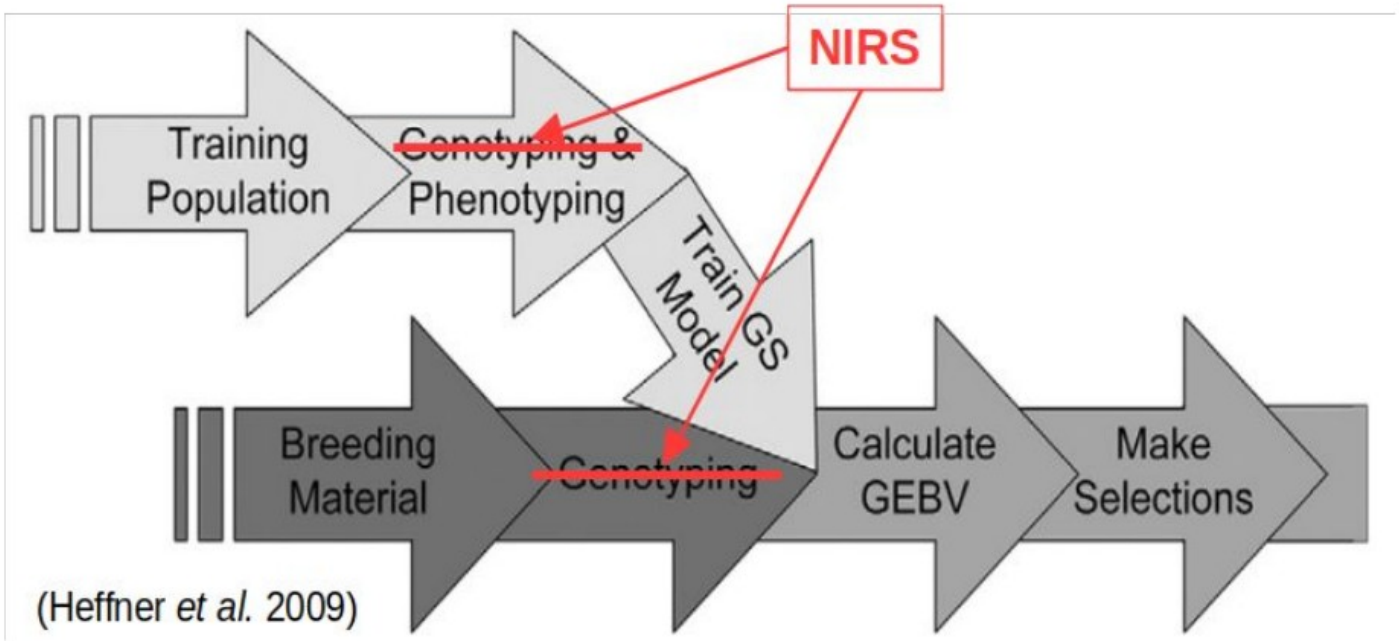


Figure 18: Comparaison de la capacité prédictive des index de sélection selon les modèles de prédiction (C) ; distribution des capacités prédictives pour les index : rendement moyen, indice de supériorité et prédiction intra environnementale. (D)

Phenomic Selection :

Phenomic Selection Is a Low-Cost and High-Throughput Method Based on Indirect Predictions: Proof of Concept on Wheat and Poplar

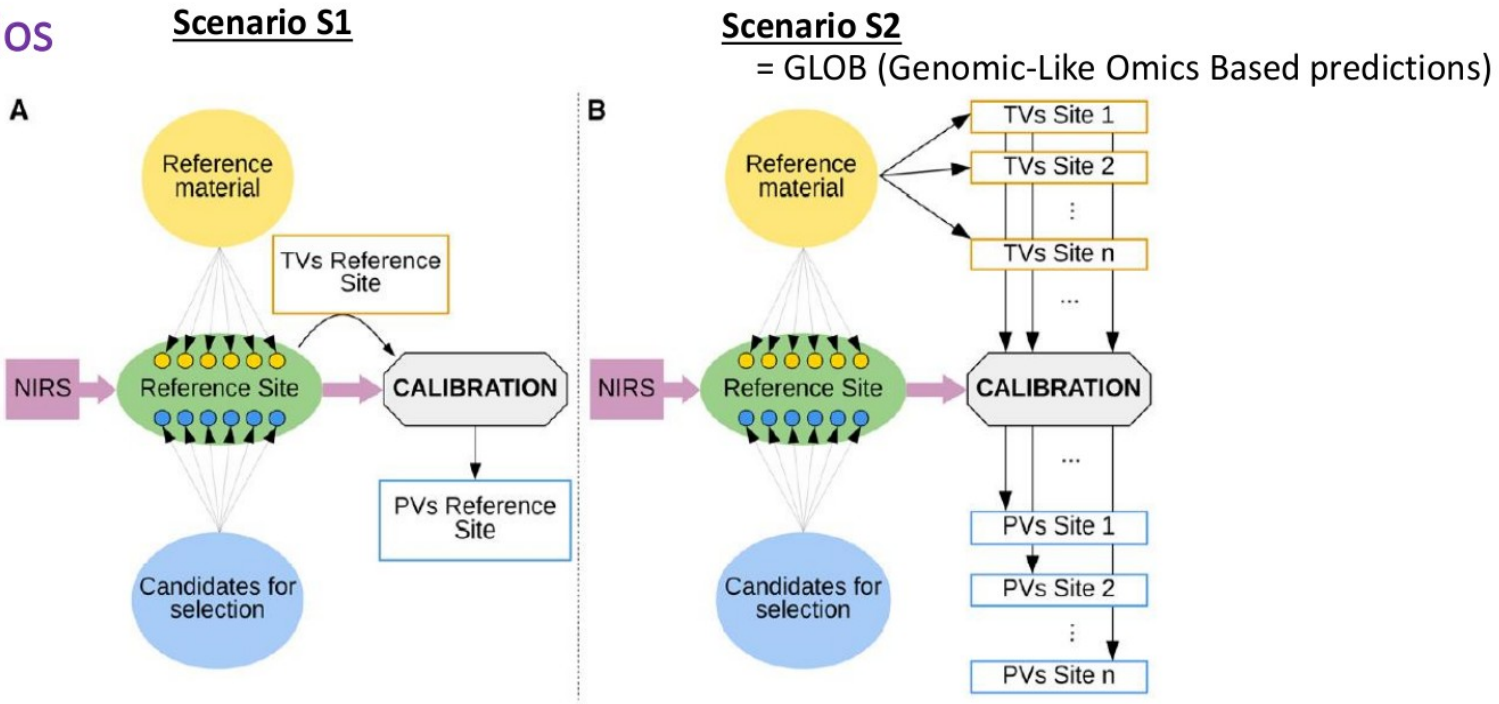
Renaud Rincent,* Jean-Paul Charpentier,[†,‡] Patricia Faivre-Rampant,[§] Etienne Paux,*
Jacques Le Gouis,* Catherine Bastien,[†] and Vincent Segura[†,1]



Diapos from : atelier – prédiction-génomique
https://github.com/timflutre/atelier-prediction-genomique

Phenomic Selection :

Prediction scenarios

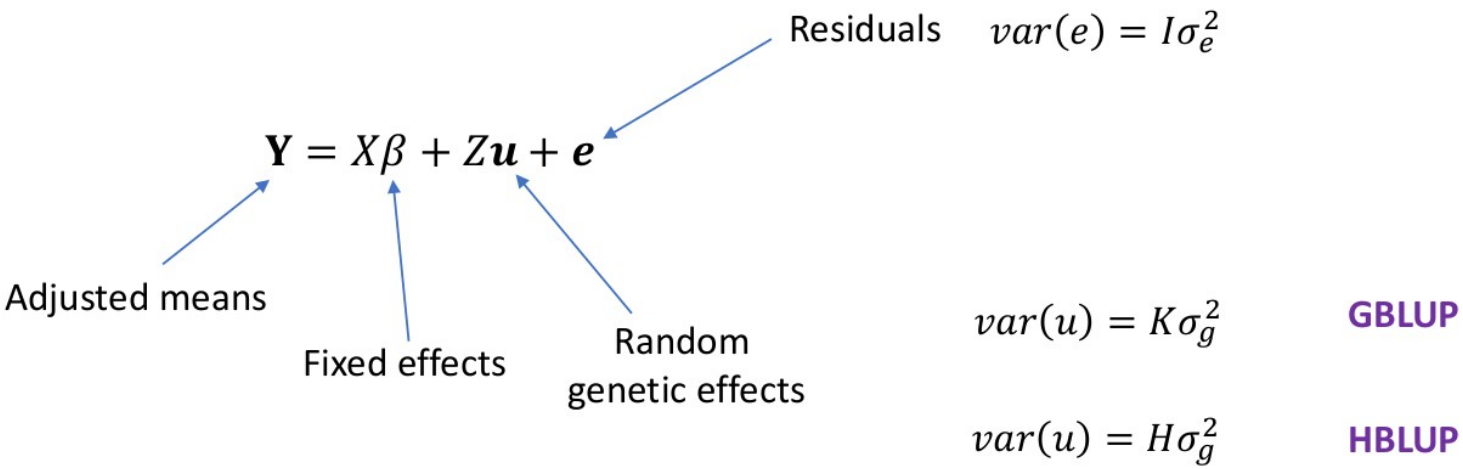


- **In Scenario S1**, all varieties (calibration + test set) are NIRSed in the environment in which the calibration set is phenotyped. The test set is predicted in the reference environment.
- **In Scenario S2**, all varieties (calibration + test set) are NIRSed in the reference environment, and the calibration set is phenotyped in an independent multi-environment trial (MET). The phenomic selection model is used to predict the test set in the MET. The test set is completely absent from the MET. (This scenario could also include situations in which all individuals are NIRSed in nursery, and the calibration set is phenotyped the next year in a MET).

Diapos from : atelier – prédiction-génomique
https://github.com/timflutre/atelier-prediction-genomique

Phenomic Selection :

Prediction models : GBLUP and HBLUP



K is the kinship matrix estimated with the genetic markers.
H is the hyperspectral similarity matrix estimated with the (pre-treated) spectra.

Remark: most GS models can be used for phenomic predictions (ridge, lasso, Bayesian alphabet, machine learning…)

Diapos from : atelier – prédiction-génomique
https://github.com/timflutre/atelier-prediction-genomique

Phenomic Selection :

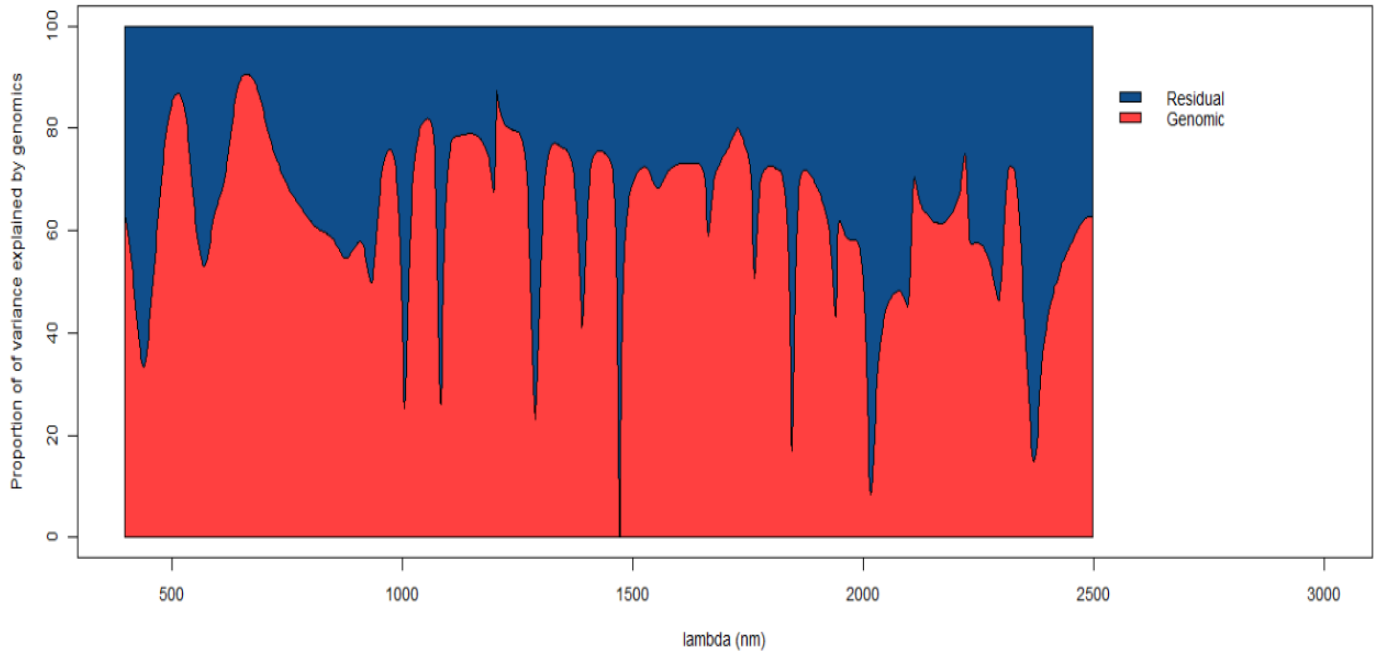
Proof of concept - material

- **Wheat:** A panel of 228 French elite varieties of winter wheat evaluated in Clermont-Ferrand under two hydric treatments and a subset of 161 varieties in 6 independent trials (other years/locations/treatments)
  - ***Genotyping***: TaBW280K genotyping array → 84,259 SNPs (Rimbert et al. 2018)
  - ***NIRS***: Grain & Leaves; FOSS spectrometers (visible and NIR)
  - ***Phenotyping***: Yield and Heading date

Diapos from : atelier – prédiction-génomique
https://github.com/timflutre/atelier-prediction-genomique

Phenomic Selection :

Results : part of variance spectra explained by genomics

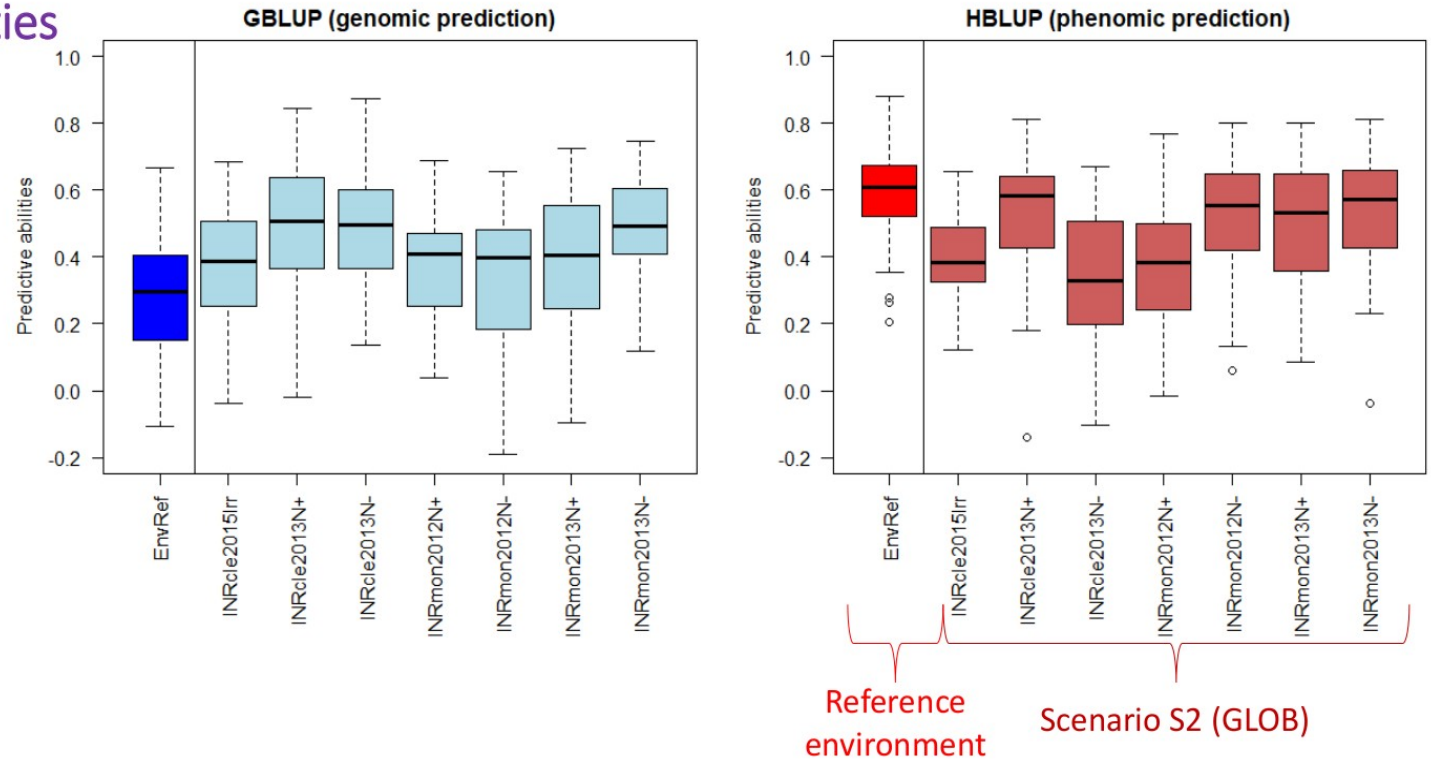


Genomic heritability is high for most wavelengths, so absorbance can be considered as a polygenic trait.
Spectra are mainly driven by genetics (and GxE). Can we use it for predictions ?

Diapos from : atelier – prédiction-génomique
https://github.com/timflutre/atelier-prediction-genomique

Phenomic Selection :

Predictive abilities



- Predictive abilities are intermediate for GS and PS.
- Predictive abilities are higher in the reference environment than in the other environments (probably because NIRS capture GxE interactions).

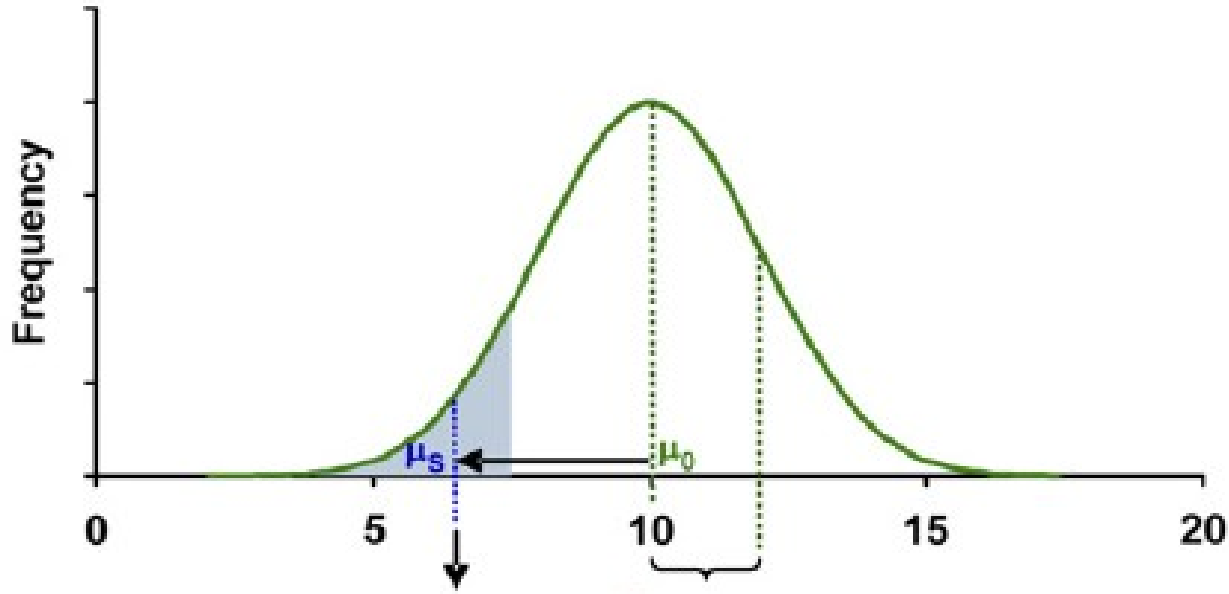
Diapos from : atelier – prédiction-génomique
https://github.com/timflutre/atelier-prediction-genomique

Phenomic Selection :

Conclusions

- Phenomic prediction is **easy** to implement !
- It is **worth pre-treating** the spectra (and doing some **spatial adjustments for each wavelength** if possible is strongly recommended).
- Phenomic predictions are **accurate for polygenic traits** (but not for oligogenic traits, see Zhu et al. 2021, TAG).
- On this dataset, PS is more accurate than GS in the reference environments **and in the environments without any NIRS data** (scenario S2, GLOB).
- The gain in comparison to GS is higher in the reference environment, probably because of **GxE interactions captured by the spectra** (see Robert et al. 2022b TAG for an analysis on GxE).
- **It is not necessary to grow the test set** to apply phenomic selection, as it can be applied with NIRS collected on seeds (or the previous year in nurseries).

Breeder's GOAL



Genetic Gain = $\Delta G = h^2 \sigma_P i / L$ → length of cycle interval (usually 1 generation)

h^2 → heritability

i → Selection Intensity: proportion of population selected to produce the next generation

σ_P → phenotypic variability in population

$$\sigma^2_{Genotype} - \sigma^2_{Dominance} - \sigma^2_{Epistasis} = \sigma^2_{Additive}$$

$$\sigma^2_{Genotype} + \sigma^2_{Environment} + \sigma^2_{GxE} + \sigma^2_{error} = \sigma^2_P$$

THANKS !