

Sélection participative du blé tendre en agroécologie : valorisation de la diversité gérée dans un réseau de fermes à travers les statistiques bayésiennes

Pierre Rivière¹ Julie Dawson² Nathalie Galic¹ Sophie Pin¹
Yannick de Oliveira¹ Olivier David³ Patrick deKochko⁴
Isabelle Goldringer¹

¹ INRA Le Moulon ferme du Moulon F-91190 Gif sur Yvette France

² University of Wisconsin-Madison, 393 Plant Sciences / Moore Hall, 1575 Linden Dr. Madison, WI 53706, USA

³ INRA Unité Mathématiques et Informatique Appliquées UR 341 Domaine de Vilvert F-78352 Jouy en Josas France

⁴ Réseau Semences Paysannes 3, avenue de la Gare F-47190 Aiguillon France



- 1 Introduction
- 2 Matériels et méthodes
 - La co-construction entre les acteurs
 - Les dispositifs expérimentaux
 - Les méthodes d'analyses
- 3 Résultats
 - Sortie du modèle hiérarchique
 - Comparaison d'un modèle hiérarchique et d'un modèle non hiérarchique
 - Influence du nombre d'environnements dans l'estimation des paramètres
- 4 Discussion
 - Le modèle
 - Impact sur le programme de sélection participative
- 5 Conclusion et perspectives

1 Introduction

2 Matériels et méthodes

- La co-construction entre les acteurs
- Les dispositifs expérimentaux
- Les méthodes d'analyses

3 Résultats

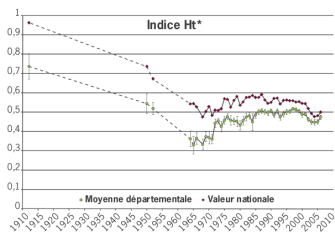
- Sortie du modèle hiérarchique
- Comparaison d'un modèle hiérarchique et d'un modèle non hiérarchique
- Influence du nombre d'environnements dans l'estimation des paramètres

4 Discussion

- Le modèle
- Impact sur le programme de sélection participative

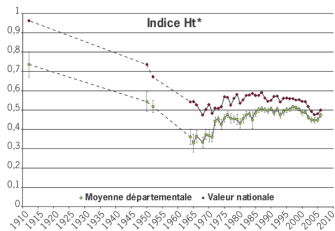
5 Conclusion et perspectives

- La biodiversité cultivée diminuée



(Goffaux et al, 2011)

- La biodiversité cultivée diminue

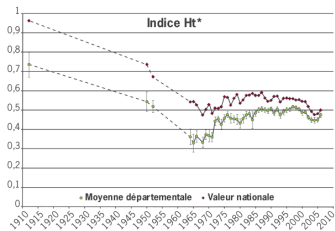


(Goffaux et al, 2011)



populations
hétérogènes,
adaptées
localement,
évolutives

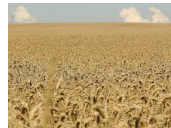
● La biodiversité cultivée diminue



(Goffaux et al, 2011)

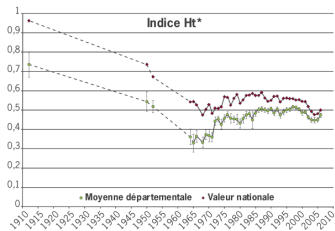


populations
hétérogènes,
adaptées
localement,
évolutives



variétés homo-
gènes : lignées
pures, hybrides

- La biodiversité cultivée diminue



(Goffaux et al, 2011)



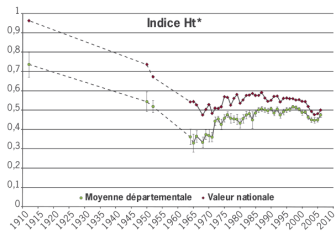
populations
hétérogènes,
adaptées
localement,
évolutives



variétés homo-
gènes : lignées
pures, hybrides

- Il y a un manque de variétés pour l'agriculture biologique

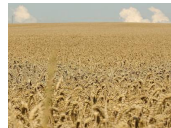
- La biodiversité cultivée diminue



(Goffaux et al, 2011)



populations
hétérogènes,
adaptées
localement,
évolutives



variétés homo-
gènes : lignées
pures, hybrides

- Il y a un manque de variétés pour l'agriculture biologique

=> Créer de nouvelles variétés adaptées localement pour l'AB et qui contribuent à plus de biodiversité.

- AB : grandes variabilités dans les environnements.
- Variété population : capacité d'adaptation temporelle, résilience, stabilité, diversité associée (Wolfe et al., 2008 ; Tooker et Franck, 2012 ; Chateil et al., 2013)
- Valoriser les interactions génotype \times environnement ($G \times E$) (Murphy et al, 2007, Desclaux et al., 2008)

=> Décentraliser la sélection sur les environnements cibles : les fermes.
Prise en compte du contexte agroécologique et socio-économique.



Associer les acteurs à la sélection : la sélection décentralisée devient participative (Ceccarelli et Grando, 2007).

Approche multi-disciplinaire : génétique des populations, génétique quantitative, agronomie, statistique, sociologie, bioinformatique.



Notre projet ...



Des agriculteurs du Réseau Semences Paysannes souhaitent de nouvelles populations de blé tendre adaptées à leurs systèmes agroécologiques.

Notre projet ...



Des agriculteurs du Réseau Semences Paysannes souhaitent de nouvelles populations de blé tendre adaptées à leurs systèmes agroécologiques.



En 2005 commence une collaboration avec l'équipe DEAP de l'INRA du Moulon.

Notre projet ...



Des agriculteurs du Réseau Semences Paysannes souhaitent de nouvelles populations de blé tendre adaptées à leurs systèmes agroécologiques.



En 2005 commence une collaboration avec l'équipe DEAP de l'INRA du Moulon.

- Création de diversité par croisements pour ensuite sélectionner :
- => 90 croisements réalisés (entre VM, VA, VP)
 - => début d'un programme de sélection participative en 2005

RSP, Maisons des Semences, paysans, DEAP, travaillent ensemble pour atteindre 3 objectifs :

- Créer des variétés-populations adaptées aux conditions locales et aux pratiques des paysans.
- Développer des méthodes et des outils opérationnels pour la gestion et la sélection de la biodiversité cultivée à la ferme.
- Renforcer l'apprentissage et l'autonomie des agriculteurs en matière de gestion et de sélection des semences.

1 Introduction

2 Matériels et méthodes

- La co-construction entre les acteurs
- Les dispositifs expérimentaux
- Les méthodes d'analyses

3 Résultats

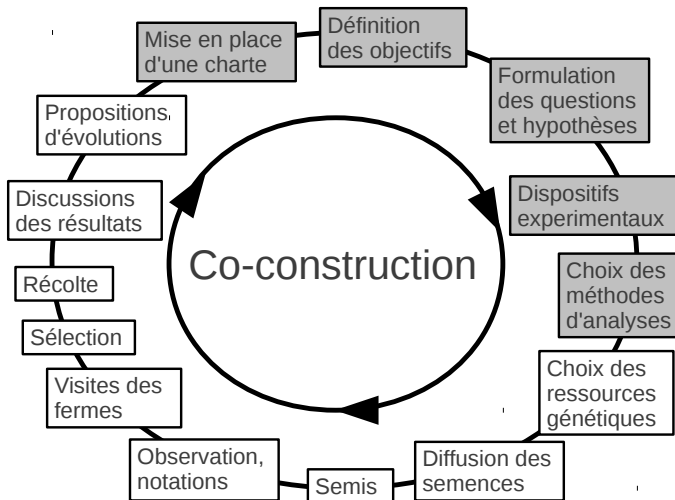
- Sortie du modèle hiérarchique
- Comparaison d'un modèle hiérarchique et d'un modèle non hiérarchique
- Influence du nombre d'environnements dans l'estimation des paramètres

4 Discussion

- Le modèle
- Impact sur le programme de sélection participative

5 Conclusion et perspectives

La méthodologie est basée sur la co-construction entre les acteurs



Fond gris : étapes discutées au cours du temps.

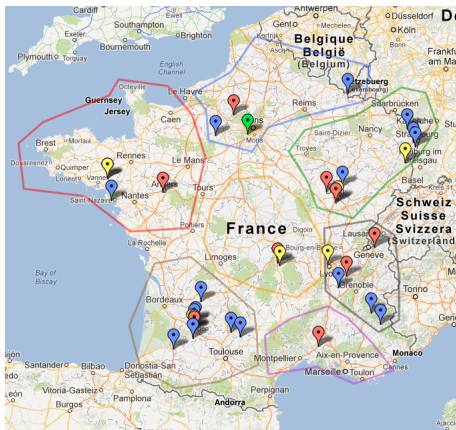
Fond blanc : étapes qui suivent le cycle de la plante.

Dispositifs expérimentaux

- Les attentes/contraintes des paysans :
 - le manque de place et de temps pour gérer des essais complètement répétés
 - l'envie d'optimiser l'espace disponible pour tester le plus de populations possible
 - le besoin de flexibilité : les paysans ne veulent pas rentrer dans un cadre trop strict qui les oblige à semer des populations qui ne les intéressent pas

Dispositifs expérimentaux

- Un large réseau d'essais $\Rightarrow \sim 25$ fermes par an pendant 3 ans



Dispositifs expérimentaux

- Les attentes des chercheurs : mettre en place des dispositifs expérimentaux qui permettent de valoriser statistiquement les résultats pour :
 - donner des résultats fiables aux paysans
 - publier une analyse scientifique du travail

Dispositifs expérimentaux

attentes des paysans
un large réseau de fermes
attentes des chercheurs

} Quels dispositifs expérimentaux ?

Dispositifs expérimentaux

Témoins : **Rouge-du-Roc** **C14** **C21** **Renan**

Fermes régionales

Rouge-du-Roc	pop1	pop2	pop3
pop4	pop5	C21	pop6
pop7	C14	pop8	pop9
pop10	pop11	pop12	Renan

pop13	C21	pop14	pop15
Renan	pop16	pop17	pop18
pop19	pop20	pop21	Rouge-du-Roc
pop22	pop23	C14	pop24

4 témoins dans 2 blocs

24 populations non répétées

Fermes satellites

Rouge-du-Roc	pop1
pop2	pop3
pop4	pop5
pop6	pop7
pop8	Rouge-du-Roc

pas de blocs ; un témoin répété deux fois

8 populations non répétées

Dispositifs expérimentaux

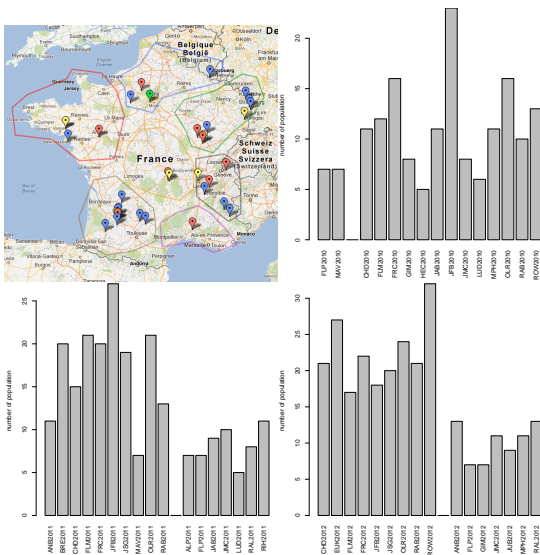
Mesures globales (1 par population) :

- hauteur
- protéine
- pmg

Mesures individuelles (25 plantes par population) :

- couleur
- barbes
- courbe
- hauteur
- distance dernière feuille base de l'épi (last leaf spike distance : LLSD)
- longueur de l'épi
- nb d'épillets
- nb d'épillet stériles
- poids de l'épi

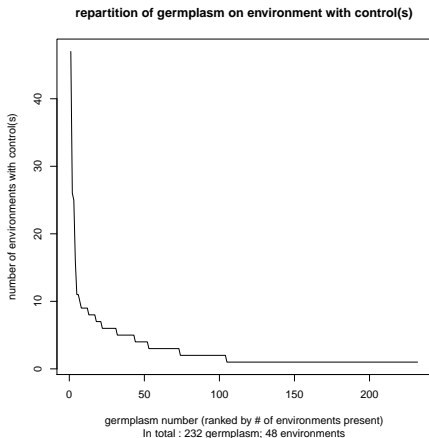
Dispositifs expérimentaux



Dispositifs expérimentaux

Définitions :

- Un environnement est une combinaison ferme \times année
- Un germplasm est une entité génétique regroupant des plantes issues d'un même processus de création, partageant une même origine : un croisement, une variété de pays, une variété ancienne, un mélange, une variété moderne, une variété issue d'une sélection massale dans les types précédemment cités.



Choix des méthodes d'analyses

Objectifs : Différencier les populations évaluées au sein de chaque ferme pour appuyer la sélection des paysans dans leurs fermes.

Sachant que l'on a :

- peu de populations répétées donc peu de degrés de liberté à la résiduelle,
- un large réseau d'essais.

Choix des méthodes d'analyses

Vraisemblance

$$Y_{ijk} = \mu_{ij} + \beta_{jk} + \varepsilon_{ijk} \quad \varepsilon_{ijk} \sim N(0, \sigma_j^2)$$

- Y_{ijk} la valeur phénotypique moyenne pour une variable donnée pour la population i , l'environnement j et le bloc k ;
- μ_{ij} l'effet de la population i dans l'environnement j ;
- β_{jk} l'effet lié au bloc k dans l'environnement j ;
- ε_{ijk} la résiduelle dans l'environnement j issue d'une loi normale centrée sur 0 et de variance σ_j^2 , propre à l'environnement j .

Choix des méthodes d'analyses

Priors

$$\mu_{ij} \sim N(\mu_{.j}, 10^6) \quad \beta_{jk} \sim N(0, 10^6)$$

On fait l'hypothèse que les variances environnementales de chaque ferme sont issues d'une distribution commune à toutes les variances environnementales.

Avec $\tau_j = \sigma_j^{-2}$ on a :

$$\tau_j \sim \text{Gamma}(\nu, \rho)$$

Afin de profiter de l'information portée par le réseau d'environnements, on fait une modélisation hiérarchique :

$$\nu \sim \text{Uniforme}(2, 10) \quad \rho \sim \text{Gamma}(10^{-6}, 10^{-6})$$

Choix des méthodes d'analyses

Posteriors

- Échantillonneur de Gibbs avec JAGS et rjags dans R.
- burn-in : 1000 itérations.
- 2 chaînes MCMC de 100 000 simulations chacune.
- 10 000 points conservés dans chaque chaîne et concaténés.
- Test de Gelman entre les deux chaînes : pas de problèmes de convergence.

Choix des méthodes d'analyses

Comparaison de moyennes

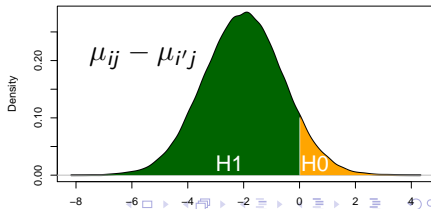
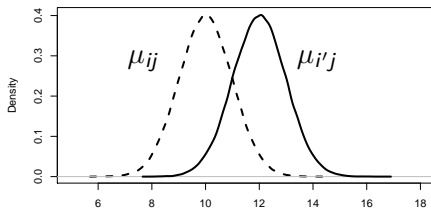
Dans un environnement donné, la moyenne de chaque population est comparée à la moyenne des autres populations.

Soit H_0 et H_1 les hypothèses telles que :

$$H_0 : \mu_{ij} - \mu_{i'j} \geq 0,$$

$$H_1 : \mu_{ij} - \mu_{i'j} < 0.$$

La probabilité est estimée par la proportion des simulations MCMC des postérieurs qui vérifie les hypothèses.



Choix des méthodes d'analyses

Comparaison de moyennes

On considère une probabilité comme significative si elle est inférieure au seuil α .

On fixe le seuil α tel que $\alpha = 0.1/I$, avec I le nombre de populations dans l'environnement.

Cela permet de corriger pour l'effet des comparaisons multiples.

C'est une alternative à la correction de Bonferroni qui nous paraissait trop stricte.

Choix des méthodes d'analyses

plan des analyses

On explore le modèle sur nos données

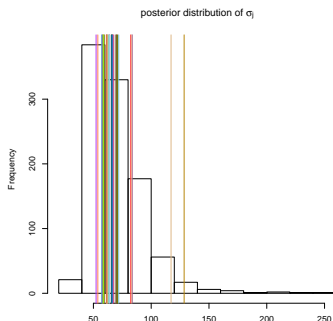
- 1 Sortie du modèle hiérarchique
- 2 Comparaison d'un modèle hiérarchique et d'un modèle non hiérarchique
- 3 Influence du nombre d'environnements dans l'estimation des paramètres

- 1 Introduction
- 2 Matériels et méthodes
 - La co-construction entre les acteurs
 - Les dispositifs expérimentaux
 - Les méthodes d'analyses
- 3 Résultats
 - Sortie du modèle hiérarchique
 - Comparaison d'un modèle hiérarchique et d'un modèle non hiérarchique
 - Influence du nombre d'environnements dans l'estimation des paramètres
- 4 Discussion
 - Le modèle
 - Impact sur le programme de sélection participative
- 5 Conclusion et perspectives

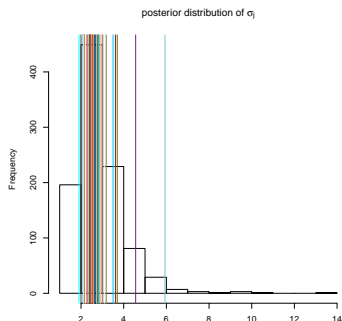
Résultats

Sortie du modèle hiérarchique

plant height



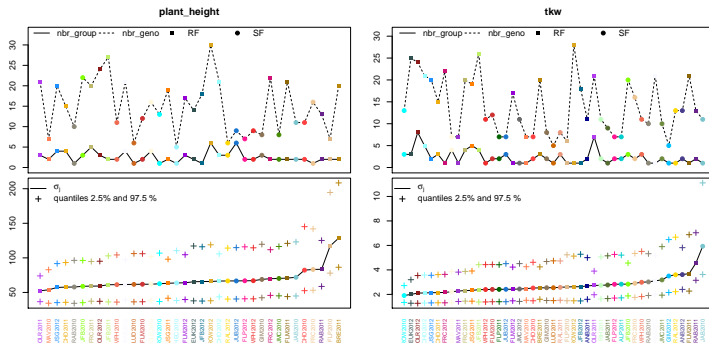
tkw



Histogramme de la distribution des σ_j basée sur les valeurs de ν et ρ . Les barres verticales sont les valeurs estimées pour chaque environnement j .

Résultats

Sortie du modèle hiérarchique



Nombre de population, nombre de groupes détectés et valeurs de σ_i estimée avec leur intervalle de confiance pour chaque environnement pour la hauteur et le pmg. RF : fermes régionales ; SF : fermes satellites.

Résultats

Comparaison d'un modèle hiérarchique et d'un modèle non hiérarchique

Le modèle

$$Y_{ijk} = \mu_{ij} + \beta_{jk} + \varepsilon_{ijk} \quad \varepsilon_{ijk} \sim N(0, \sigma_j^2)$$

- Y_{ijk} la valeur phénotypique moyenne pour une variable donnée pour la population i , l'environnement j et le bloc k ;
- μ_{ij} l'effet de la population i dans l'environnement j ;
- β_{jk} l'effet lié au bloc k dans l'environnement j ;
- ε_{ijk} la résiduelle dans l'environnement j issue d'une loi normale centrée sur 0 et de variance σ_j^2 , propre à l'environnement j .

Résultats

Comparaison d'un modèle hiérarchique et d'un modèle non hiérarchique

Le modèle

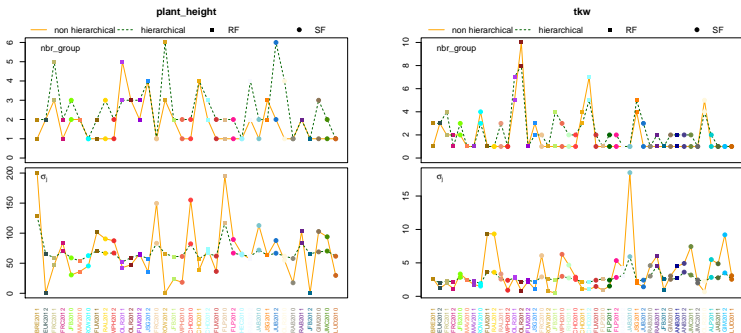
$$Y_{ijk} = \mu_{ij} + \beta_{jk} + \varepsilon_{ijk} \quad \varepsilon_{ijk} \sim N(0, \sigma_j^2)$$

Les priors

Le modèle hiérarchique	Le modèle non hiérarchique
$\mu_{ij} \sim N(\mu_{.j}, 10^6)$	$\beta_{jk} \sim N(0, 10^6)$
Avec $\tau_j = \sigma_j^{-2}$ on a :	
$\tau_j \sim \text{Gamma}(\nu, \rho)$	$\tau_j \sim \text{Gamma}(10^{-6}, 10^{-6})$
$\nu \sim \text{Uniforme}(2, 10)$	
$\rho \sim \text{Gamma}(10^{-6}, 10^{-6})$	

Résultats

Comparaison d'un modèle hiérarchique et d'un modèle non hiérarchique



Comparaison des variance résiduelles (σ_j) et nombre de groupes pour la hauteur et le pmg entre le modèle non hiérarchique et le modèle hiérarchique. RF : fermes régionales ; SF : fermes satellites.

Le modèle hiérarchique diminue les variances résiduelles.

Résultats

Comparaison d'un modèle hiérarchique et d'un modèle non hiérarchique

difference in the number of groups	no	↑	↓
color	8	20	5
awns	7	17	10
curve	18	13	2
nbr sterile spikelets	19	12	4
plant height	13	21	2
LLSD	3	26	2
spike length	15	12	1
nbr spikelets	8	10	4
spike weight	28	12	4
tkw	17	21	7
protein	10	14	1
% of the total	36.1	52.6	11.3

Occurrence des changements dans le nombre de groupe pour chaque variable. ↑ : le nombre de groupes est plus grand avec le modèle hiérarchique comparé au modèle non hiérarchique. ↓ : le nombre de groupes est plus petit avec le modèle hiérarchique comparé au modèle non hiérarchique.

Le modèle hiérarchique détecte plus de groupe que le modèle non hiérarchique.

Résultats

Influence du nombre d'environnements dans l'estimation des paramètres

On étudie combien d'environnements sont nécessaires pour avoir une analyse fiable. On ne peut pas considérer toutes les combinaisons d'environnements possibles :

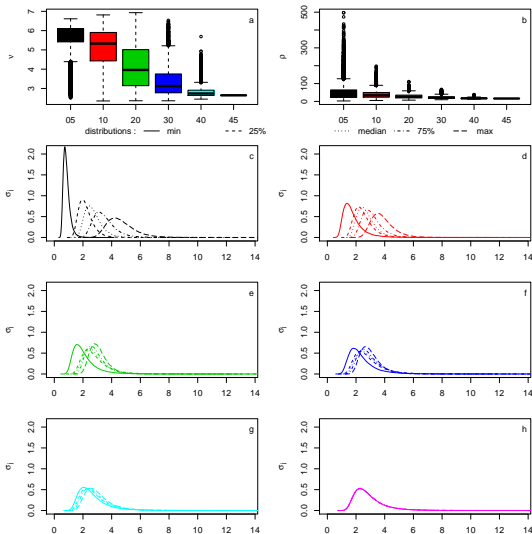
$\binom{1}{n} + \binom{2}{n} + \dots + \binom{n-1}{n} + \binom{n}{n}$ combinaisons !

On génère plusieurs échantillons avec un certains nombres d'environnements.

pmg	
Nombre d'environnements	Nombre d'échantillons
5	20000
10	12000
20	8000
30	4000
40	2000
45	1

Résultats

Influence du nombre d'environnements dans l'estimation des paramètres



Pour le pmg, évolution de l'estimation de ν (a) et ρ (b) avec le nombre d'environnements pris dans l'analyse (5, 10, 20, 30, 40 et 45) et la distribution respective des σ_j pour les différents quartiles de ν et ρ (c, d, e, f, g, h, i).

- 1 Introduction
- 2 Matériels et méthodes
 - La co-construction entre les acteurs
 - Les dispositifs expérimentaux
 - Les méthodes d'analyses
- 3 Résultats
 - Sortie du modèle hiérarchique
 - Comparaison d'un modèle hiérarchique et d'un modèle non hiérarchique
 - Influence du nombre d'environnements dans l'estimation des paramètres
- 4 Discussion
 - Le modèle
 - Impact sur le programme de sélection participative
- 5 Conclusion et perspectives

Discussion

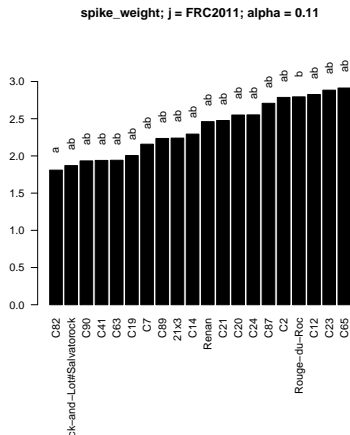
Le modèle

- L'effet G et $G \times E$ sont confondus dans l'estimation de l'effet population. Intéressant pour les paysans qui recherchent une adaptation spécifique et non une adaptation plus large qui minimiserait les interactions
- Le modèle hiérarchique permet de diminuer le taux de faux positifs résultant d'une mauvaise estimation des variances environnementales, en prenant en compte la totalité de l'information du réseau d'essais.
- Il est important d'avoir assez d'informations disponibles pour avoir une estimation plus robuste des paramètres.

Discussion

Le modèle

- La valeur de l'erreur α dépend de l'objectif agronomique. Différencier les populations est important pour sélectionner entre les populations. Le seuil peut être ajusté pour avoir au moins deux groupes plutôt que les paysans choisissent au hasard les populations.



Discussion

Impact sur le programme de sélection participative

Trois avantages de notre méthode :

- Les paysans ne sont pas contraints par les répétitions et le nombre de populations à tester, chacun peut donc participer selon ses moyens.
- Les paysans ne sont pas contraints par les choix des autres paysans du réseau. Ils ont une grande liberté dans le choix des populations qu'ils vont semer : une grande diversité de populations peut être évaluée sur le réseau d'essais.
- La méthode permet d'estimer les effets génétiques sur chaque ferme. Cela permet aux paysans d'avoir accès à une information sur la valeur spécifique de chaque population sur leur ferme. Cette information leur permet de sélectionner parmi les populations.

- 1 Introduction
- 2 Matériels et méthodes
 - La co-construction entre les acteurs
 - Les dispositifs expérimentaux
 - Les méthodes d'analyses
- 3 Résultats
 - Sortie du modèle hiérarchique
 - Comparaison d'un modèle hiérarchique et d'un modèle non hiérarchique
 - Influence du nombre d'environnements dans l'estimation des paramètres
- 4 Discussion
 - Le modèle
 - Impact sur le programme de sélection participative
- 5 Conclusion et perspectives

Conclusion et perspectives

- Ce travail participe à :
 - développer des méthodes pour la gestion et la sélection de la biodiversité cultivée à la ferme ;
 - renforcer l'apprentissage des agriculteurs sur le comportement des populations ;
 - la création de variétés-populations adaptées aux conditions locales et aux pratiques des paysans.
- La co-construction entre les acteurs permet d'explorer de nouveaux modèles.
- Ce travail est possible grâce à une collaboration inter-disciplinaire.
- Des simulations sont en cours pour mieux comprendre le modèle et explorer différents plans d'expériences dans un réseau de fermes.
- Un travail de ce type est en cours pour étudier les interactions $G \times E$.



Merci de votre attention ! 😊

Pour aller plus loin :

J.C. Dawson et al., Collaborative plant breeding for organic agricultural systems in developed countries. *Sustainability*, 3 :1206–1223, **2011**.

P. Rivière et al., Response to farmer mass selection in early generation progeny of bread wheat landrace crosses. *Renewable Agriculture and Food Systems*, in press, **2013**.

P. Rivière et al., Mise en place d'une méthodologie de sélection participative sur le blé tendre en France. *Innovations Agronomiques*, in press, **2013**.