

GRANDE CULTURE  
**GCHP2E**  
à Hautes Performances Economiques et Environnementales

# Estimer le progrès génétique à partir de données historiques

**Stagiaire = Matthieu Dritzas**

**Maitre de stage : Stéphane Lassalvy**

Avec Marie-Hélène Bernicot et le groupe innovation variétale : Arnaud Gauffreteau, Josiane Lorgeou, François Piraux, Célia Pontet, Sandrine Mikol



# Pourquoi ce stage ?

## ■ Une demande forte de communiquer sur le progrès génétique

■ Progrès génétique = amélioration d'une caractéristique recherchée par rapport à l'existant (variété témoin ou ensemble de variété)

- Rendement, Valeur d'utilisation, Régularité , Résistance aux maladies et autres bioagresseurs, Résistance à la verse

■ Différentes approches existent, essais spécifiques, estimation à partir des données historiques (pontage , modèle mixte)

■ Intégrer la gestion des séries incomplètes avec le modèle mixte dans les outils et chaînes de calcul du GEVES

Dans le cadre du GIS GCHP2E ?  
Pour bénéficier de la dynamique du  
groupe innovations variétales

# Quelques références

- A. Luciani – Etude du progrès génétique chez différentes espèces de grande culture – Rapport GEVES, 2004
- H.P. Piepho & al. – Dissection genetic and non-genetic sources of long-term yield trend in german ocial variety trials, 2013
- H.P. Piepho & J. Möhring – Selection in cultivar trials, is it ignorable ? Crop-Science 46:192-201, 2006
- F.X. Oury & al. – A study of genetic progress due to selection reveals a negative effect of climate change on bread wheat yield in France
- R. Little & D. Rubin – Statistical Analysis with Missing Data, 1987
- R. Little – Selection and pattern-mixture models, 2008
- J.L. Schafer & J.W. Graham, Missing data : Our view of the state of the art, Psychological Methods Vol. 7, No. 2, 147-177, 2002
- G. Verbeke & G. Molenberghs – Linear Mixed Models for Longitudinal Data, 2000
- J. Pinheiro & D. Bates – Mixed-Effects Models in S and S-PLUS, 2000
- D. Bates, M. Mächler, B. Bolker and S. Walker – Fitting linear mixed-effects models using lme4, Journal of Statistical Software

# Les données dont dispose le CTPS/GEVES

	2000	2001	2002	2003	2004	2005
var 2000-1	x	x				
var 2000-2	x	x				
var 2000-3	x					
var 2001-1		x	x			
var 2001-2		x				
var 2001-3		x				
var 2002-1			x			
var 2002-2			x	x		
var 2002-3			x	x		
var 2003-1				x	x	
var 2003-2				x		
var 2003-3				x	x	
var 2004-1					x	x
var 2004-2					x	X
var 2004-3					x	x
var 2004-4					x	
tem a	x	x				
tem b	x	x	x			
tem c	x	x	x	x	x	
tem d			x	x	x	X
tem e				x	x	X
tem f						X
tem x	x	x	x	x	x	x

## Les variétés candidates

- 2 ans pour l'inscription (très rarement 3)
- Un taux de passage entre la 1<sup>o</sup> et 2<sup>o</sup> année très variable selon les espèces de 100% pour les avoines à 50% pour le blé tendre d'hiver, sélection sur le rendement, la qualité, la résistance aux maladies

## Des témoins

- Ces données sont du type **Missing at random (MAR)** :  $P(M|Y_{obs}, Y_{mis}) = P(M|Y_{obs})$  : la raison de la censure est liée à la variable mais n'est pas cachée. **Ce type de données manquantes peut être traité avec un modèle linéaire à effets mixte**

# Le modèle que nous avons développé pour calculer le PG :

## 1) Estimation des effets variétaux par modèle mixte

$$Y = \text{Génotype} + \text{Année} + n^{\circ} \text{essai}(\text{Année}) + \text{Génotype} \times \text{Année} + \varepsilon$$

Calcul des  
moyennes  
ajustées  
variétales

**Effets fixes**

**Effets aléatoires**

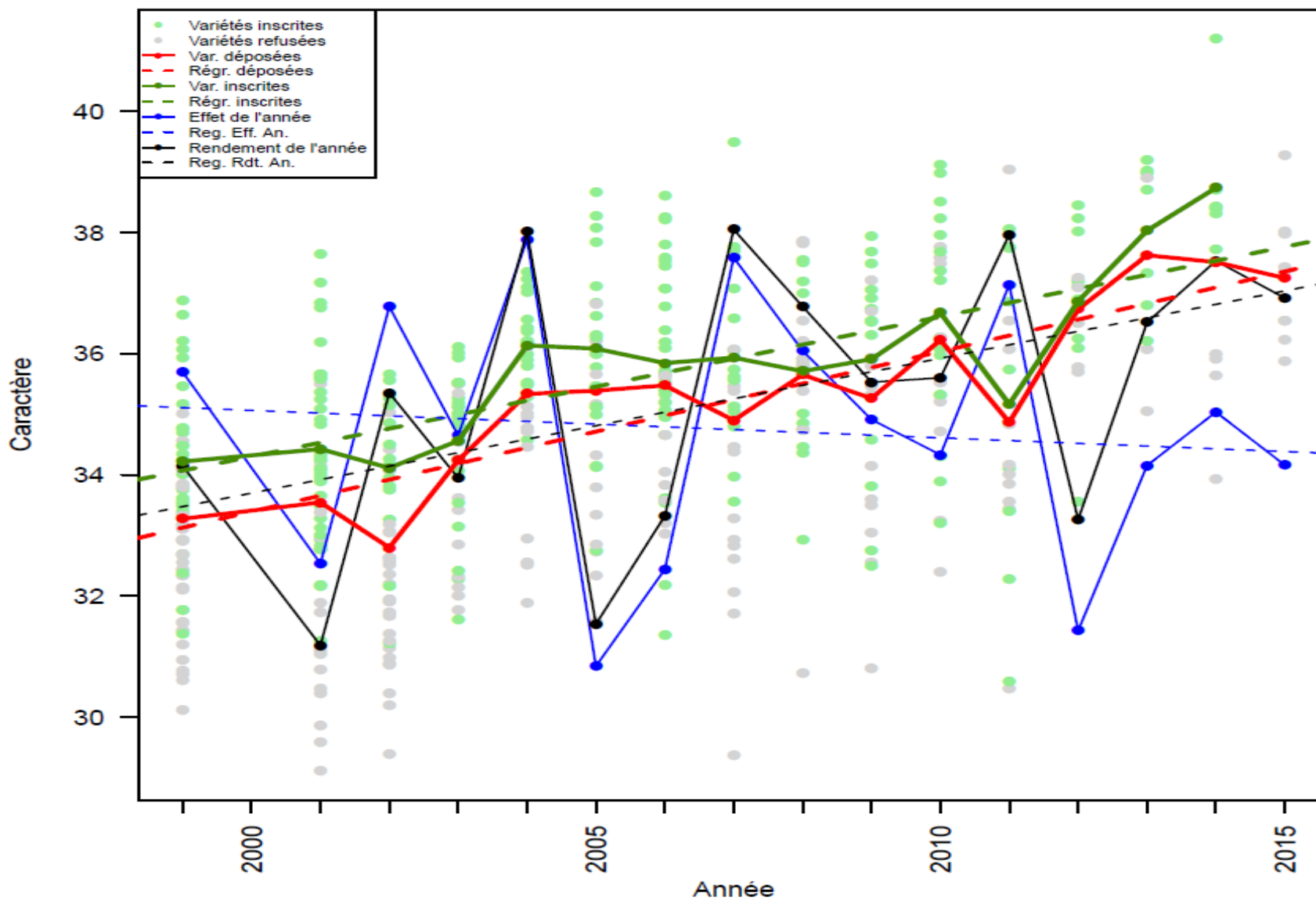
## 2) Calcul des moyennes ajustées par année de promotion pour les variétés candidates

## 3) Régression des moyennes ajustées promotionnelles par l'année de promotion

$$\text{Moy Promotions} = \alpha \times \text{Année de Promotion} + \beta + \varepsilon''$$

# Exemple tournesol série B, 1999 à 2015

Courbe de progrès génétique pour RDT GRAIN 11% Code : 760



Equation PG var. présentées :  $y = 0.26x - 495.28$

Equation PG Var. inscrites :  $y = 0.23x - 427.45$

# Les différentes façons d'évaluer le Progrès Génétique

## 1° approche : des expérimentations spécifiques

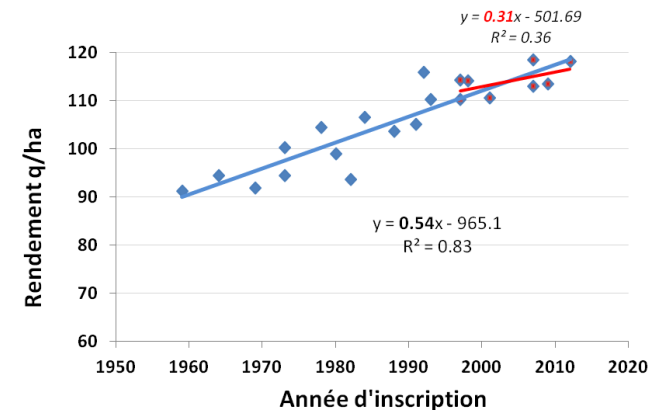
### ■ études récentes pour le blé tendre

■ **INRA** (*Brancourt, 2004*) : 14 variétés, inscription de 1946 à 1992

- HN PG rdt à 15% = 0,74 q/ha/an , BN PG= 0,53 q/ha/an

■ **Arvalis** : 21 variétés inscription de 1955 à 2012, Nord France, 5 essais

- PG = 0.54 q/ha/an de 1955 à 2012
- PG = 0.30 q/ha/an de 1997 à 2012




■ **Biogemma, INRA, Arvalis** (*P.Cormier, 2012*)


- 1985 variétés inscription de 1986 à 2010, 4 essais, 2 niveaux N
- HN PG rdt à 15% = 0,41 q/ha/an , BN PG= 0,30 q/ha/an

# Les différentes façons d'évaluer le Progrès Génétique

## 2° approche : utilisation des données historiques

	2000	2001	2002	2003	2004	2005
var 2000-1	x	x				
var 2000-2	x	x				
var 2000-3	x					
var 2001-1		x	x			
var 2001-2		x				
var 2001-3		x				
var 2002-1			x			
var 2002-2			x	x		
var 2002-3			x	x		
var 2003-1				x	x	
var 2003-2				x		
var 2003-3				x	x	
var 2004-1					x	x
var 2004-2					x	X
var 2004-3					x	x
var 2004-4					X	
tem a	x	x				
tem b	x	x	x			
tem c	x	x	x	x	x	
tem d			x	x	x	X
tem e				x	x	X
tem f						X
tem x	x	x	x	x	x	x

 **SANS ajustement par comparaison à une variété témoin présente de nombreuses années**

 **Sur données inscription de Sud de 1999 à 2014, comparaison à Soissons présent 16 les années 0,34 q/ha/an , comparaison à Apache, présent 14 années = 0,21 q/ha/an (0 si enlève la 1° année !)**



# Les différentes façons d'évaluer le Progrès Génétique

## 2° approche : utilisation des données historiques

- **AVEC le rendement moyen ajustée des variétés**
- **Calcul de l'effet année, en prenant en compte les différences entre les variétés communes aux 2 années successives**
  - INRA sur données inscription de 1986 à 2010 = 1,37 q/ha/an Nord et 0,65 q/ha/an pour le Sud (*Oury, 2012*)
  - GEVES sur données de BTH sud 1999 à 2014 = 0,93 q/ha/an
- **Avec le modèle linéaire mixte**
  - ARVALIS : réseau post-inscription selon les zones d'études
    - entre 1983 et 2012 : de 0,6 à 0,8 q/ha/an
    - entre 1997 et 2012 : de 0,4 à 0,6 q/ha/an
  - GEVES : réseau d'inscription toute France de 1989 à 2000 = 0,9 q/ha/an (*Luciani, 2004*)
  - GEVES sur données de BTH 1999 à 2104 = 0,41 q/ha/an pour sud et 0,46 nord

# Pourquoi ces différences d'estimations ?

## **Années sur lesquelles le PG est calculé**

-  Pour le blé tendre, le progrès rendement traité plus faible ces dernières années

## **Les conditions de cultures dans lequel le PG est évalué**

-  **Les estimations du PG à partir de modèle semblent souvent plus élevées que celles quantifiées directement**, *en particulier pour les méthodes corrigeant l'effet année uniquement par les variétés communes 2 années consécutives.*

# Pourquoi ces différences d'estimations ?

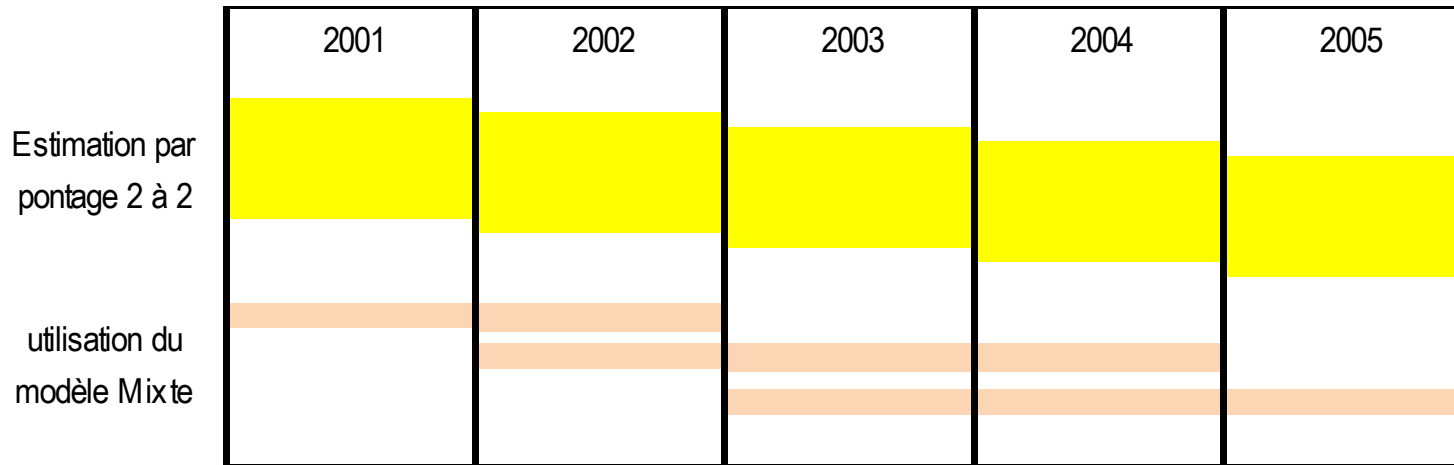
## ■ Pourquoi ?

## ■ Serait-ce un effet de nos jeux de données avec des trous qui ne sont pas complètement aléatoires ?

- Toutes les variétés déposées ne passent pas en seconde année car il y a une sélection sur le rendement, les variétés ayant eu une mauvaise performance sont éliminées, des variétés sont retirées.
- En comparant le rendement des variétés candidates entre la 1<sup>o</sup> et la 2<sup>o</sup> année, on a souvent des moins bons résultats la 2<sup>o</sup> année (en % aux témoins)

	BTH T Nord rdt	BTH T sud rdt	BD T sud rdt	TO série A rdt huile	TO série B rdt	TO série C rdt
<b>écart aux témoins A2-écart témoins A1</b> (témoins communs aux 2 )	-0.27	-0.27	-0.2	-0.55	-0.66	-0.22
Année =						

- Les modèles utilisant les variétés communes 2 années successives sont beaucoup plus sensibles à cette diminution que le modèle mixte qui utilisera également le fait que les variétés témoins (souvent plus stables) sont présents plusieurs années.



# Construction d'un jeu de données pour étudier l'influence du taux de passage en 2<sup>ème</sup> année et du taux d'inscription

## A partir des estimations calculées sur les données rendement de la série B tournesol

- Simulation des moyennes de chaque promotion :  
 $y = a * x + b$ ,  $a$  et  $b$  coefficients de la droite de progrès génétique
- Simulation des moyennes variétales (l'année de dépôt de la variété est supposée connue) :  
Moy Génotype = Moyenne promotion du Génotype +  $\varepsilon'$ ,  $\varepsilon' \sim N(0, \sigma_{\text{Variété}})$
- Simulation moyennes des génotypes par année et par essai  
-Selon la formule :  
 $Y = \text{Génotype} + \text{Année} + \text{Essai(Année)} + \text{Génotype} \times \text{Année} + \varepsilon$ 
  - Année  $\sim N(0, \sigma_{\text{Année}})$
  - Essai(Année)  $\sim N(0, \sigma_{\text{Essai(Année)}}$
  - Génotype  $\times$  Année  $\sim N(0, \sigma_{\text{interaction}})$
  - Résiduelle  $\sim N(0, \sigma_{\varepsilon})$

- 10 variétés déposées chaque année
- 11 années d'études
- 7 essais chaque année

# Influence du taux de passage en 2<sup>ème</sup> année et du taux d'inscription

Variations des taux de passage et d'inscription :

Nb témoins	Fréquence	Taux passage	Taux inscription	Pente moyenne estimée	Biais
4	1 témoin changé chaque année	100	100	<b>0.300</b>	<b>0.004</b>
4		100	50	<b>0.300</b>	<b>0.004</b>
4		50	100	<b>0.341</b>	<b>0.045</b>
4		50	50	<b>0.340</b>	<b>0.044</b>

Même avec le modèle linéaire mixte, il semble y avoir un effet du taux de sélection sur le rendement dans l'estimation du progrès génétique.

Mais même dans ce cas où le taux de sélection de 50% appliquée à des variétés ayant tout le même niveau de rendement est beaucoup plus important que la vraie vie où la sélection se fait également sur les différences de potentiel entre les variétés, et sur d'autres caractéristiques (résistance aux maladies, la surestimation est faible, de 15%

*pente estimée plus forte qu'en taux de passage plus faible, effet de l'effectif plus réduit ??*

- **Modèle linéaire mixte est le « meilleur » modèle pour estimer, pour le rendement, le PG à partir de séries de données historiques**
  
- **Pour le GEVES**
  - Finaliser les travaux
  
  - Etudier comment améliorer les estimations
    - Témoins = nombre, durée
    - Comprendre l'effet de la sélection entre 1<sup>o</sup> et 2<sup>o</sup> année
  
- **Le progrès génétique c'est aussi des progrès sur les résistances aux maladies, verse, qualités, régularités... les méthode de quantification du PG peuvent être différentes**