



# Analyse et prédiction des IGE

## Recherche de bouquets variétaux productifs et stables

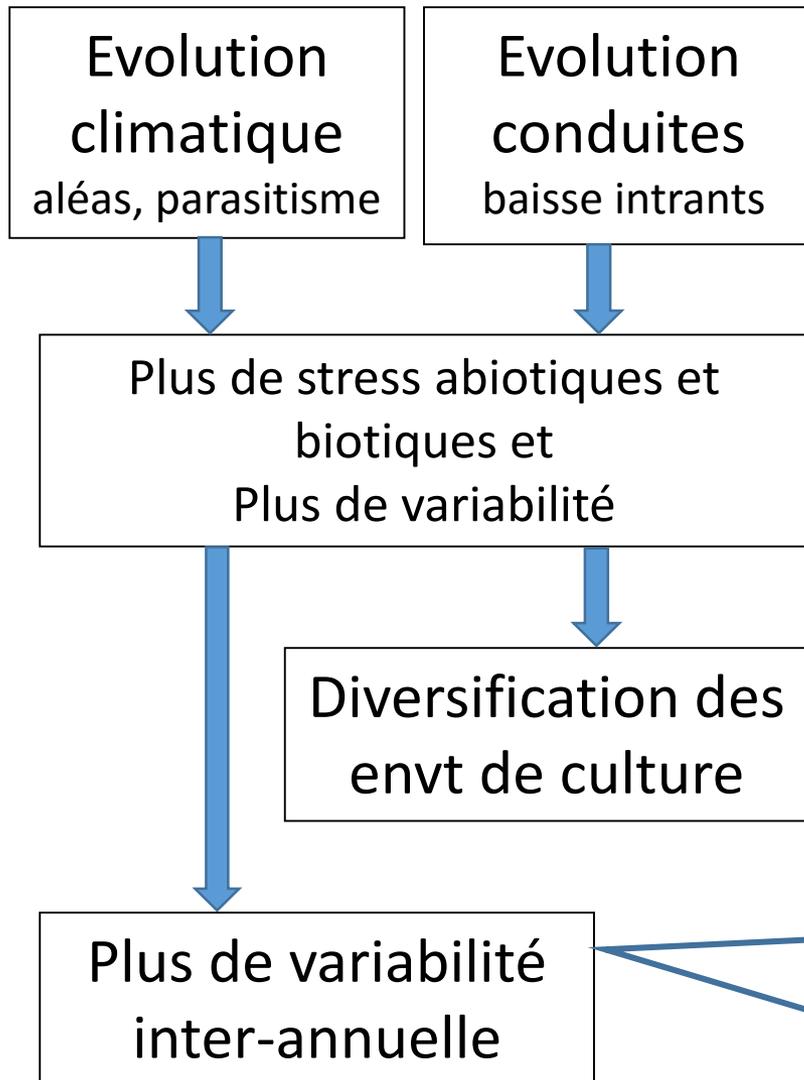
### Vers la préconisation de variétés

M. D'Orchymont, A. Gauffreteau, J. Lorgeou,  
F. Piraux, G. Grignon, C. Pontet, F. Maupas, H. Escriou  
S. Lassalvy, MH Bernicot

Journées du GIS GC HP2E  
17 janvier 2017



# Contexte



- Adaptation des variétés aux environnements de culture
  - Besoin de variétés :
    - Diversifiées** et adaptées à une large gamme d'environnements
    - Caractérisées** : Performances moyennes et réponse aux différentes conditions de cultures au champ
- Modéliser les IGE dans les réseaux d'essais variétés (MET)
- Période 2012 – 2015 sur données post-inscription blé, Betterave, Tournesol, PdT

- Importance de la stabilité des variétés
  - Réflexion à l'échelle de bouquets de variétés
- Estimation et prédiction de la stabilité de bouquets variétaux

# Objectifs sur période 2015 - 2016

- Quelles variétés composent les meilleurs bouquets?
- Peut-on les repérer a priori?

Performance et stabilité des  
**Bouquets variétaux**

*Bouquet : ensemble de variétés cultivées sur une exploitation*

**Bases orthogonales**  
Toutes les variétés dans tous les environnements

- Peut-on prédire la stabilité d'un bouquet?
- La qualité de prédiction dépend-elle du niveau de trous de la base de données?

- Capacité des modèles statistiques à prédire les IGE?
- Quel Impact de la non-orthogonalité sur cette qualité prédictive?

Des bases de données des essais variétés non orthogonales  
→ **Bases glissantes**

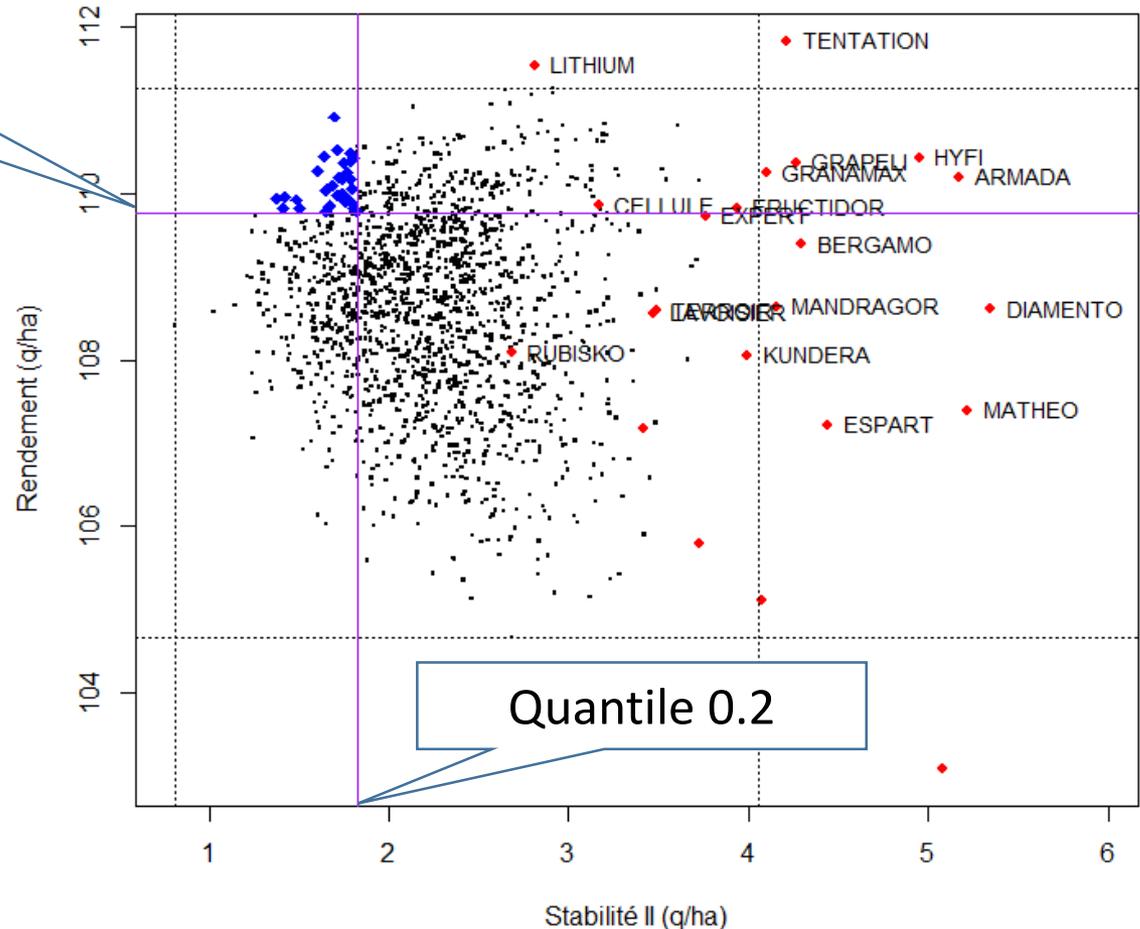


# Quelles variétés composent les meilleurs bouquets?

Quantile 0.85

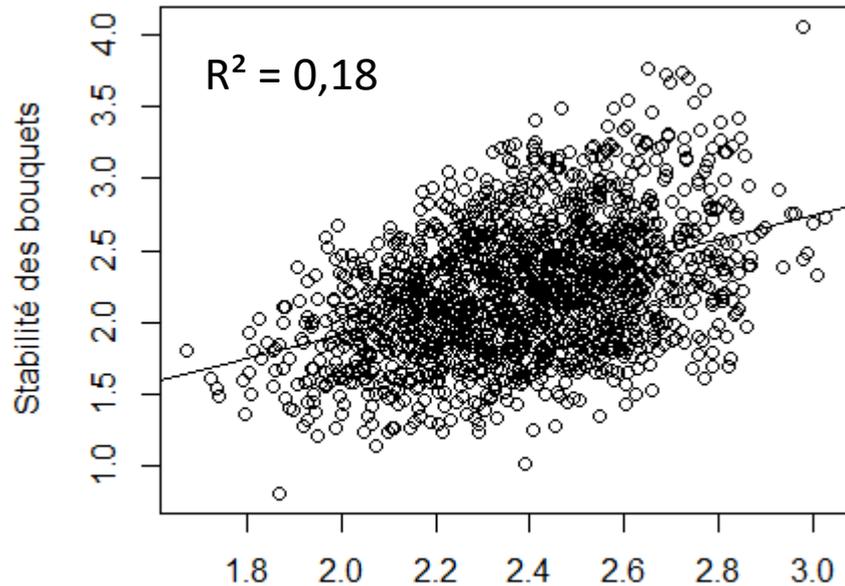
## BdD orthogonale

- Blé (post-inscription)
- Zone Nord
- 2014 (17 sites) – 2015 (2 sites)
- 22 variétés
- Bouquets de 3 variétés

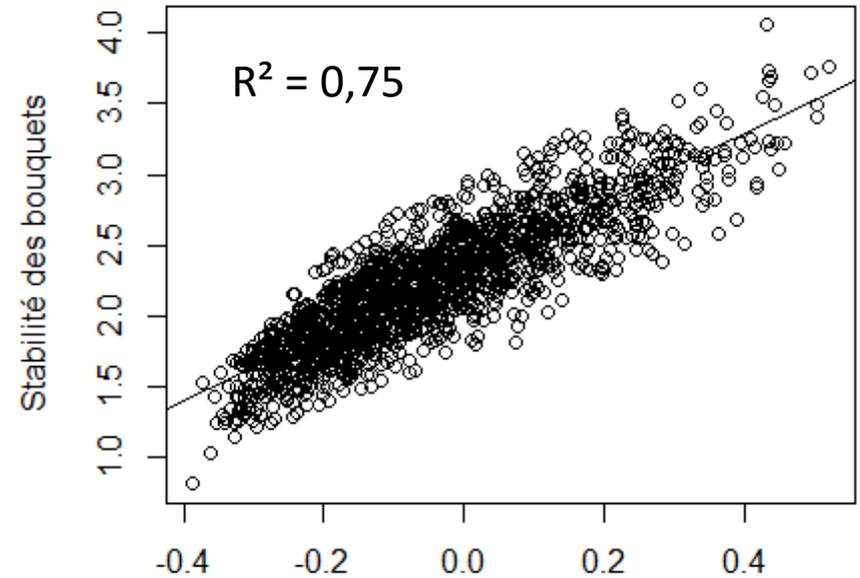


Stabilité = racine carrée de l'écovalence (moyenne des IGE au carré) –  
 Capacité à maintenir son rang dans les environnements du réseau d'essais

# Quelles variétés composent les meilleurs bouquets?



Stabilité des 3 variétés associées  
 racine carrée de la moyenne des  
 écovalences des 3 variétés



Moyenne des corrélations associées  
 2 variétés complémentaires sont  
 corrélées négativement pour l'IGE

→ La complémentarité des variétés d'un bouquet semble influencer plus sur la stabilité du bouquet que la stabilité des variétés elle-même.

# Peut-on repérer des variétés complémentaires selon leurs caractéristiques?

## Les caractéristiques phénotypiques

précocité à épiaison, résistances à : piétin verse, fusariose, verse, oïdium, rouille jaune et brune, septoriose tritici

- On calcule par couple de variétés la corrélation entre les caractéristiques phénotypiques des deux variétés, que l'on met en relation avec la corrélation des IGE (complémentarité des variétés).
  - On s'intéresse aux différences au carré par caractéristique variétale et par couple de variétés. On cherche à expliquer la complémentarité des variétés par ces différences à l'aide d'un modèle linéaire : on a 8 variables explicatives, une pour chaque caractéristique phénotypique, qui sont peu corrélées.
- Ces deux approches ne permettent pas de caractériser la complémentarité des variétés.

# Peut-on prédire la stabilité d'un bouquet à partir de bases glissantes ?

## Méthodes :

$$1. \quad eco(V_B) = \frac{1}{N} \sum_{j=1}^N IGE_{Bj}^2 = \frac{1}{N} \sum_{j=1}^N \left( \frac{1}{3} \sum_{i=1}^3 IGE_{ij} \right)^2$$

Contrainte : connaître la valeur des IGE pour les 3 variétés dans tous les essais

$$2. \quad eco(V_B) = \frac{2^2}{3^2} [eco(V_1V_2) + eco(V_1V_3) + eco(V_2V_3) - (eco(V_1) +$$

# Peut-on prédire la stabilité d'un bouquet à partir de bases glissantes ?

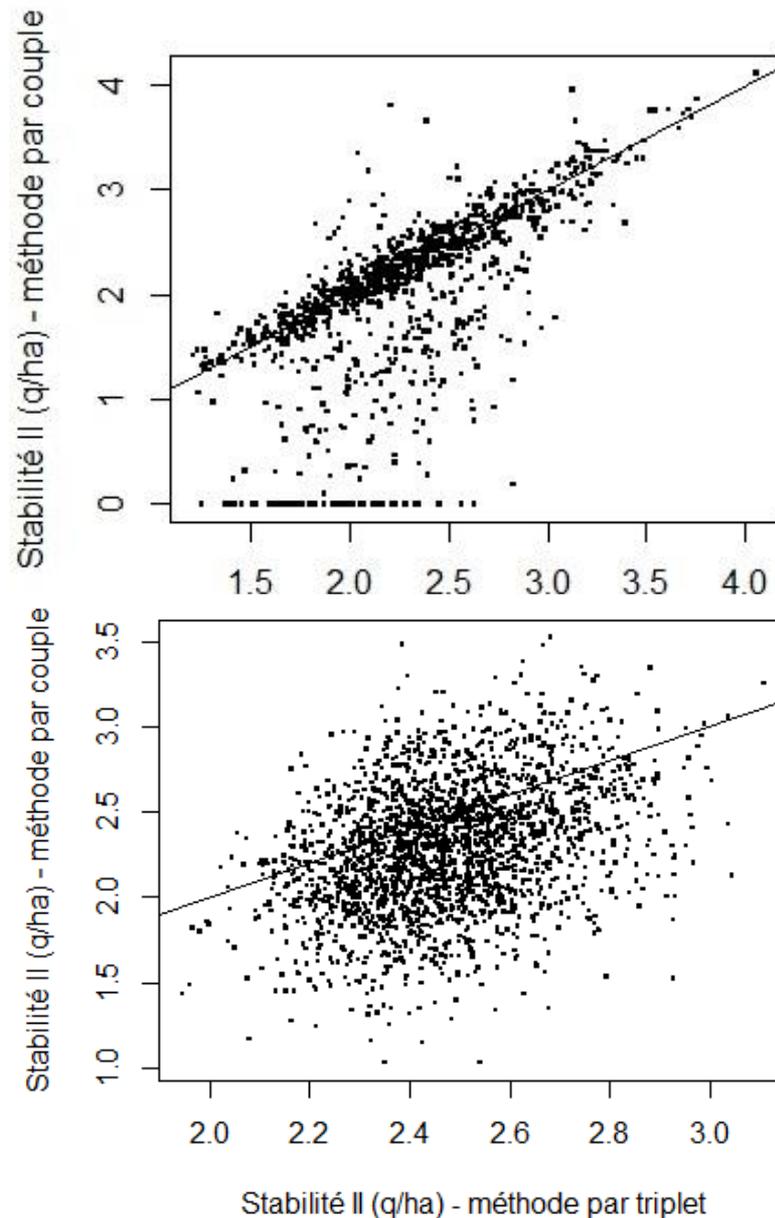
- Nb de Bouquets
  - 1330 (BdD 2014-2015)
  - 15180 (BdD générée)
- Non estimables :
  - ~20% si 2 ans et 80% de conservation
  - >80% si 5 ans et 50% de conservation

## → BdD 2014-2015

- RMSE (estimation couples) = 0,66 q/ha
- Ecart-type (estimation triplets) = 0,47 q/ha

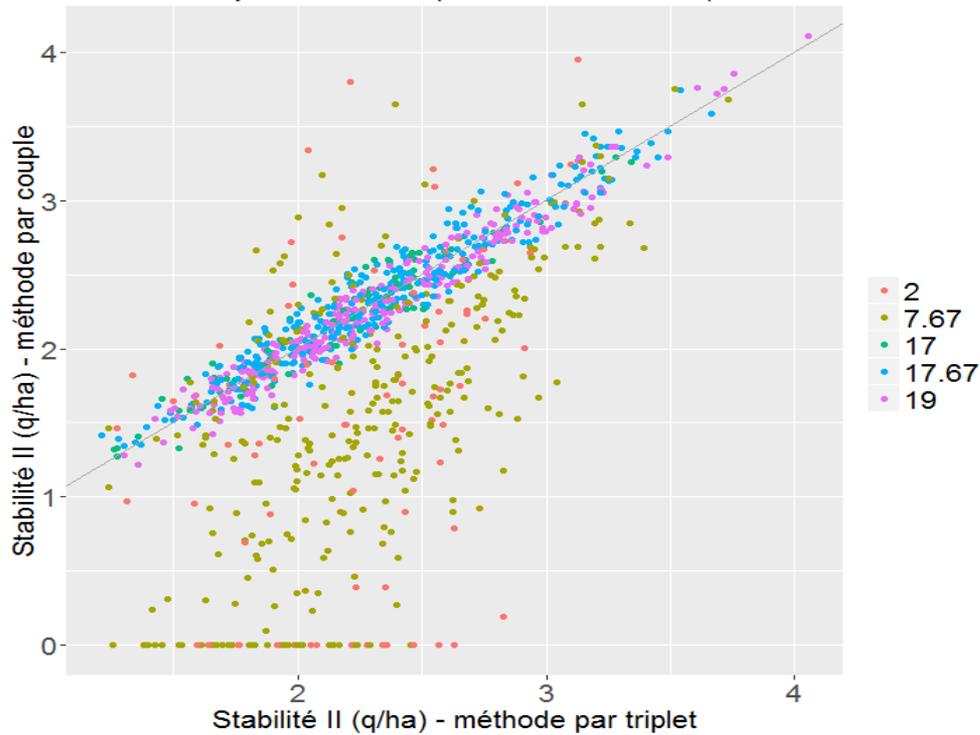
## → BdD générée

- RMSE (estimation couples) = 0,40 q/ha
- Ecart-type (estimation triplets) = 0,18 q/ha

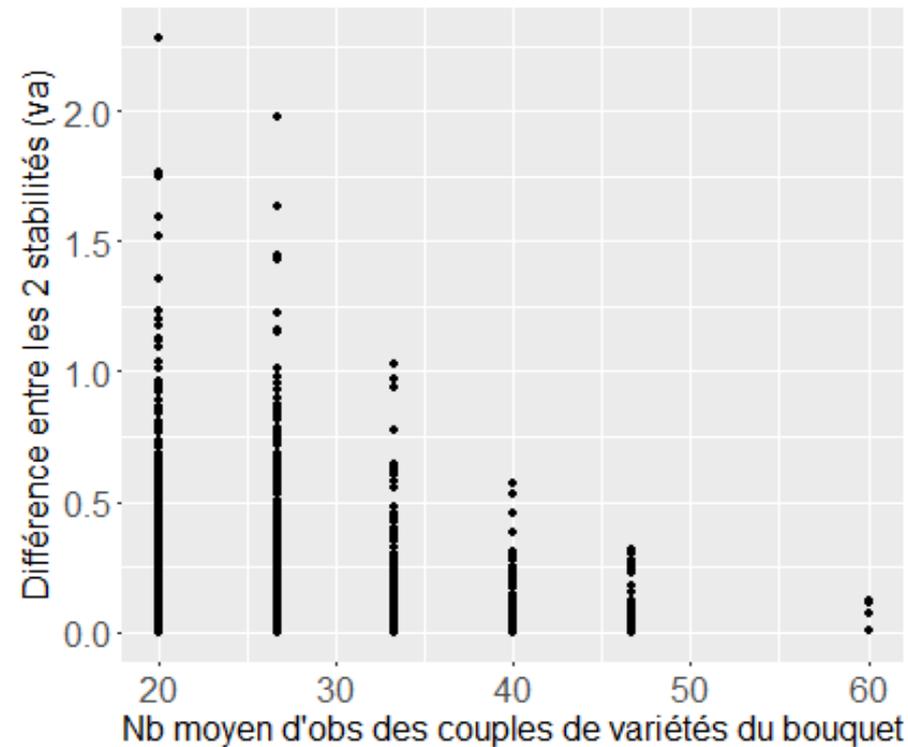


# La qualité de la prédiction dépend-elle du niveau de trous dans la base de données?

Base 2014-2015



Base générée (5ans – 50% de conservation)



# Conclusion sur les bouquets

## Base orthogonale

- Les bouquets variétaux stables sont constitués de variétés assez stables et complémentaires.
- La stabilité des bouquets peut être prédite à partir de la stabilité et de la complémentarité des variétés qui les composent.
- L'étude de la complémentarité des variétés ne permet pas de dégager de caractéristiques phénotypiques pouvant expliquer cette complémentarité.

## Base glissante

- Certains bouquets ne peuvent pas être constitués du fait des données manquantes.
- L'écovalence calculée d'un bouquet dépend fortement du nombre d'observation des couples de variétés de ce bouquet dans la base glissante.
- En dessous d'un certain pourcentage de conservation des variétés d'une année à l'autre dans la base, la proportion de bouquets construits devient très faible et les erreurs de calcul de l'écovalence très importantes.

➔ Intérêt de compléter les bases de données avec des modèles de prédiction des IGE



# Capacité des modèles statistiques à prédire les IGE?

## Contexte

- Qualité prédictive des modèles d'IGE modeste en comparaison du modèle additif (RMSEP, Spearman, erreurs III)
  - Observée sur des réseaux d'essais variétés d'espèces variées : Blé tendre, Tournesol, Betterave, Pomme de Terre
  - Pourquoi n'y arrive-t-on pas?
    - ✓ Imprécisions des données des essais (variabilité résiduelle trop importante, stress non notés, covariables peu précises...)?
    - ✓ réseaux d'essais trop déséquilibrés en matière de listes de variétés et de diversité de lieux / an?
    - ✓ Les modèles (méthodes + covariables) ne sont pas pertinents pour analyser l'IGE?
- ➔ Travail sur des données générées pour vérifier la seconde hypothèse

CV 2010	<b>Spear</b>
RF (Forward)	0.41
PLS	0.38

Modèle additif **0.42**

CV 2009	<b>Spear</b>
RF (Forward)	0.29
PLS	<b>0.38</b>

Modèle additif 0.35

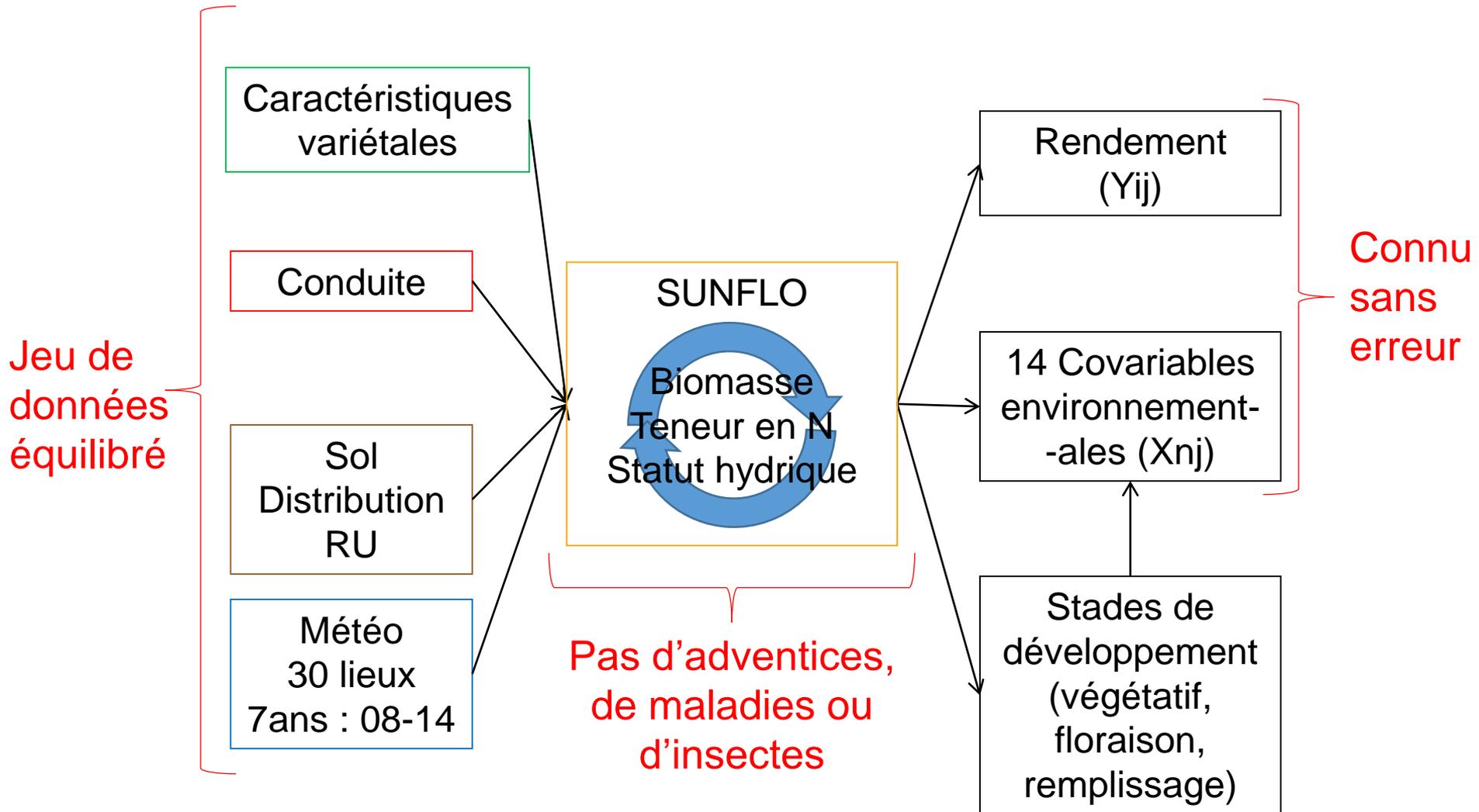
CV 2008	<b>Spear</b>
RF (Forward)	<b>0.33</b>
PLS	0.29

Modèle additif **0.33**

*Voir aussi  
Intervention Séminaire ESPERA, 2013  
Communications écrites*

# Capacité des modèles statistiques à prédire les IGE?

Travail sur données générées



# Capacité des modèles statistiques à prédire les IGE?

Cas des données simulées avec Sunflo.

Validation croisée sur An x Lieu

Modèle	Perte de Rdt	Baisse (% Add)	% Env > Add
Additif	0.155	-	-
Rég. Factorielle	0.093	40	66
PLS1	0.09	42	72
Lasso	0.087	44	73
PLS2	0.086	45	71
Random Forest	0.084	46	80

- Tous les modèles GxE ont de meilleures qualités prédictives que le modèle additif sur la base de données simulées avec Sunflo
- L'avantage de la random forest est moins clair en moyenne mais améliore les prédictions du modèle additif dans plus d'environnements
- La part d'IGE prédite par les modèles GxE reste limitée!
- Les erreurs sont plus fortes pour les environnements de forts stress précoces

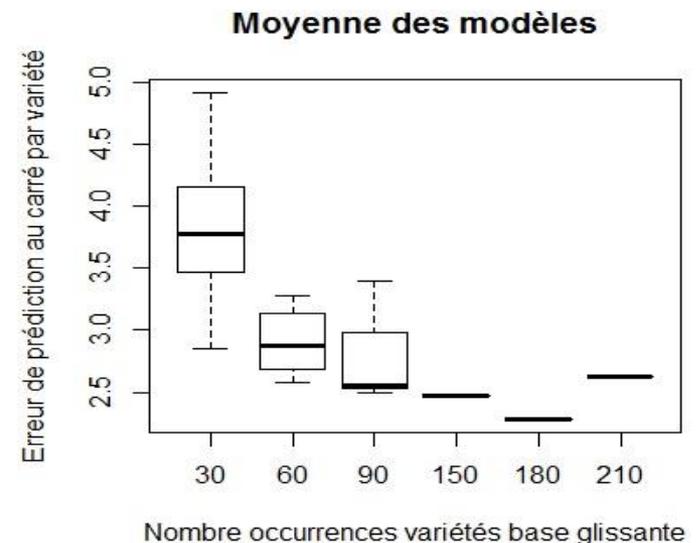
# Quel Impact de la non-orthogonalité sur la qualité prédictive des modèles d'IGE?

- Désorthogonalisation de la Base de données simulée avec Sunflo selon méthode glissante : 8 variétés annuellement, 50% de variétés conservées inter-annuellement. Nombre de variétés par nombre d'occurrence :

30	60	90	150	180	210
15	7	4	1	1	1

- Sous bases orthogonales associées, avec même nombre d'observations que les bases glissantes par réduction du nombre d'environnements (61 environnements)

- ➔ Des prédictions légèrement moins bonnes sur Bases glissantes
- ➔ Une qualité prédictive variétale plus faible pour les variétés moins présentes



# Conclusions sur la prédiction des IGE

- Les modèles d'IGE améliorent la qualité prédictive du modèle additif
  - La random forest fait mieux que les régressions linéaires, pénalisées et PLS : caractère non linéaire et interactif de l'interaction
  - La part d'IGE non prédite reste importante (>50%)
  - Les moins bonnes prédictions s'observent pour des environnements à fort stress précoces
  - Les modèles calibrés sur bases glissantes presque aussi performants que ceux établis sur bases orthogonales
  - Mais une meilleure prédiction des variétés les plus présentes dans la base
- ➔ Boucher les trous des bases générées et réelles,
- ➔ faire des bouquets et évaluer leur pertinence



# Communications écrites

- Gauffreteau A., Grignon G., Pachot P., Lorgeou J., Piraux F., Maupas F., Escriou H., Pontet C., Salvi F. 2014. Assessing the predictive accuracy of various statistical methods (PLS, random forest and factorial regression) that use environmental covariates to model Genotype x Environment Interactions in multi-environment trials. Tenth Working Seminar on Statistical Method in Variety Testing. Poland.
- Gauffreteau A., Grignon G., Pachot P., Lorgeou J., Piraux F., Maupas F., Escriou H., Pontet C., Salvi F., 2015. Assessing the predictive accuracy of various statistical methods (PLS, random forest and factorial regression) that use environmental covariates to model genotype x environment interactions in multi-environment trials. Biultyn Oceny Odmian. 34
- Gauffreteau A., D'Orchimond M., Pontet C., Debaeke P., 2016. Can Genotype x Environment Management Interactions (GEMI) be predicted in Sunflower multi-environment trial? 19<sup>th</sup> International Sunflower Conference. 29/05/2016-03/06/2016. Edirne – Turkey.
- Lorgeou J., Piraux F., Grignon G., Gauffreteau A. 2015. Interactions Génotypes – Environnements, une complexité à interpréter en blé tendre. Perspectives Agricoles. 426.





# Une démarche itérative

- Identifier les variétés les plus appropriées en rendement aux différents environnements ciblés

- ⇒ modéliser les IGE sur le rendement avec des méthodes statistiques et des covariables G et E à partir des données des essais variétés historiques
- ⇒ effectuer des prédictions de rendements des variétés pour différents environnements (sol, données météo historiques)

2012-2015

- Identifier les variétés dont les IGE sur le rendement seraient complémentaires et dont la complémentarité augmenterait la stabilité relative

- ⇒ Définir les méthodes
- ⇒ Estimer les données manquantes avec les modèles IGE

2015-2016

Tests de plusieurs modèles avec données :

- observées en Post-Inscription sur blé tendre, betteraves, pdt et tournesol ou
- simulées avec modèle de culture en tournesol

⇒ Evaluation des qualités de prédiction

⇒ Recherches sur les origines des écarts prédicts-observés

# Projet de valorisation par Arvalis pour la préconisation de variétés en blé tendre

**Les caractéristiques des 3 parcelles à semer** (contexte agro-pédo-climatique (avec ses FL), dates de semis, précédents, gestion résidus, type de sol) ;

## FILTRES sur les VARIETES

- Filtres sur des caractéristiques souhaitées et rédhibitoires
- Compromis entre caractéristiques : fonction du contexte pédoclim + IT prévisionnel

Variétés Parcelle 1

Variétés parcelle 2

Variétés Parcelle 3

**Proposition de 3 listes de variétés** avec informations sur caractéristiques des variétés, indicateurs technico-éco, environnementaux (RDT, MBP, IFT, etc.)

**Choix des variétés à prendre en compte dans chacune des 3 parcelles**

Calculs de la **stabilité et des rendements** de toutes les combinaisons de 3 variétés possibles avec les variétés filtrées et choisies des listes des 3 parcelles

Proposition de **bouquets de variétés avec données de RDT et de stabilité, informations sur les caractéristiques individuelles des variétés et moyennes des bouquets**