Évaluation de la biodiversité en utilisant la stratégie d'identification ciblée du matériel génétique pour des caractères agronomiques/fonctionnels

Abdallah Bari
Kenneth Street, Michael Mackay, Eddy De Pauw, Dag Endresen and Ahmed Amri

Colloque 2011
Centre de la Science de la Biodiversité du Québec, Montréal
9 Décembre 2011

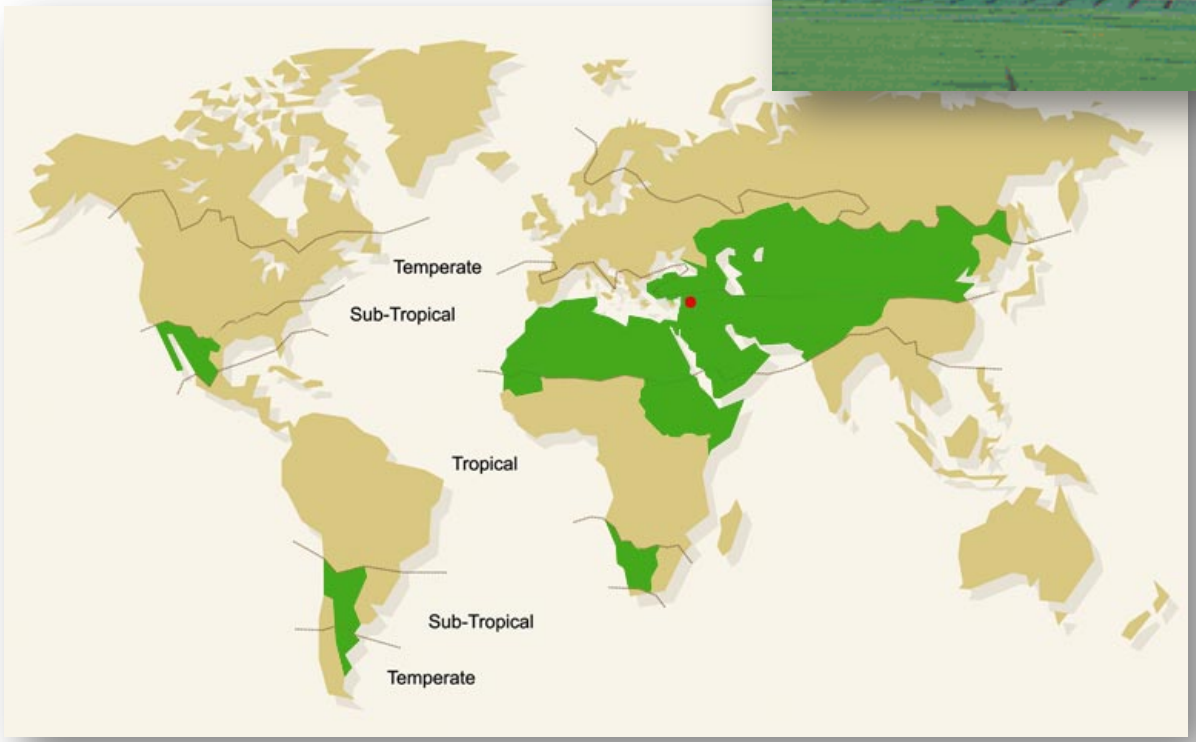
Grain Research & Development Corporation

contenu

- **Objectif**
  - Développer une information a priori
  - Développer le meilleur sous-ensemble possible d'accessions avec des caractères recherchés
- **Ensemble de données**
  - Données environnementales
  - Données de caractères
- **Méthodologies**
  - Préparation des données
  - Techniques de modélisation
- **Résultats/Discussion**
  - FIGS sous-ensembles
  - caractères
- **Conclusion**

Grain Research & Development Corporation

ICARDA



International Center for Agricultural Research in the Dry Areas

GRDC
Grain Research & Development Corporation

apport de la diversité intra spécifique (ressources phyto-génétiques)

- Augmentation de 1-2% de rendement annuel pendant les 30 - 40 années passées
- Contribution génétique~ 50%

Nouvelles variétés de blé en Ethiopie

Grain Research & Development Corporation

apport de la diversité intra spécifique (ressources phyto-génétiques)

Les ressources phyto-génétiques (caractères)

- l'adaptation phénologiques (courte durée de croissance)
- l'utilisation efficiente de l'eau,
- résistance aux stress biotiques (maladies et insectes),
- tolérance aux stress abiotiques comme la sécheresse et la salinité),
- grain de meilleure qualité

Evaluation au champ

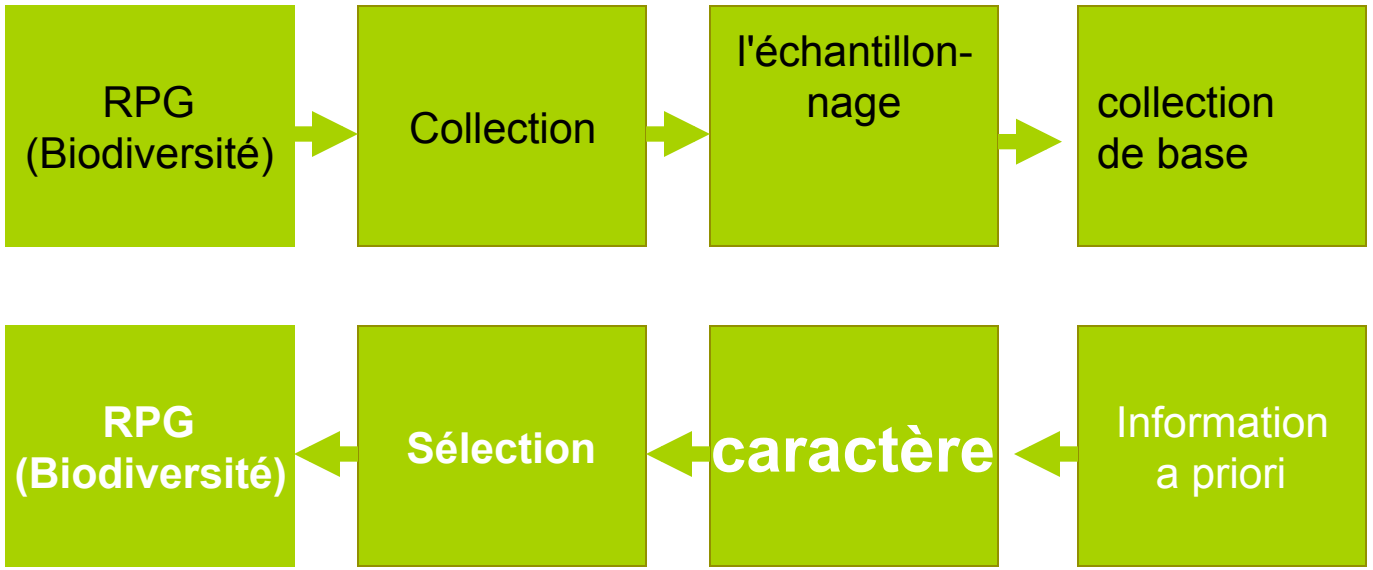
Grain Research & Development Corporation

défi à surmonter

- 50 - 60 000 caractères (loci)
- 7 million d'échantillons
- 1400 banques de gènes

Seed samples

approche FIGS

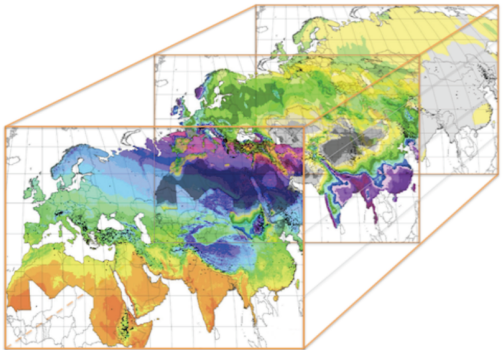


FIGS applique aux ressources phytogénétiques (collections conservées) la « même » pression de sélection que celle exercée sur les plantes par l'évolution.

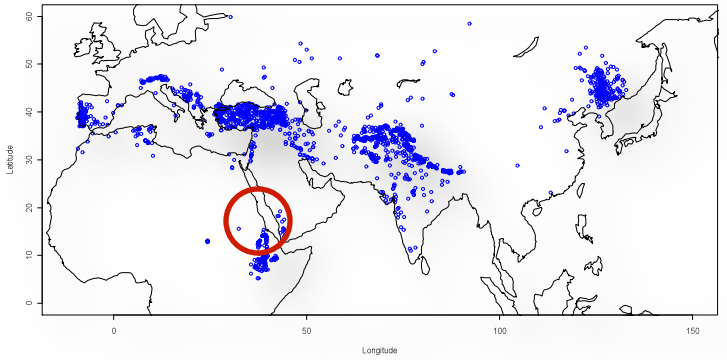
exploration de la variation naturelle

En reliant les caractères, les environnements (et les pressions de sélection qui y sont associés) avec les accessions des banques de gènes (par exemple variétés locales et les espèces apparentées), nous pouvons cibler les accessions les plus susceptibles d'avoir la variation génétique spécifique.

caractères



environnement



les accessions ciblées

Grain Research & Development Corporation

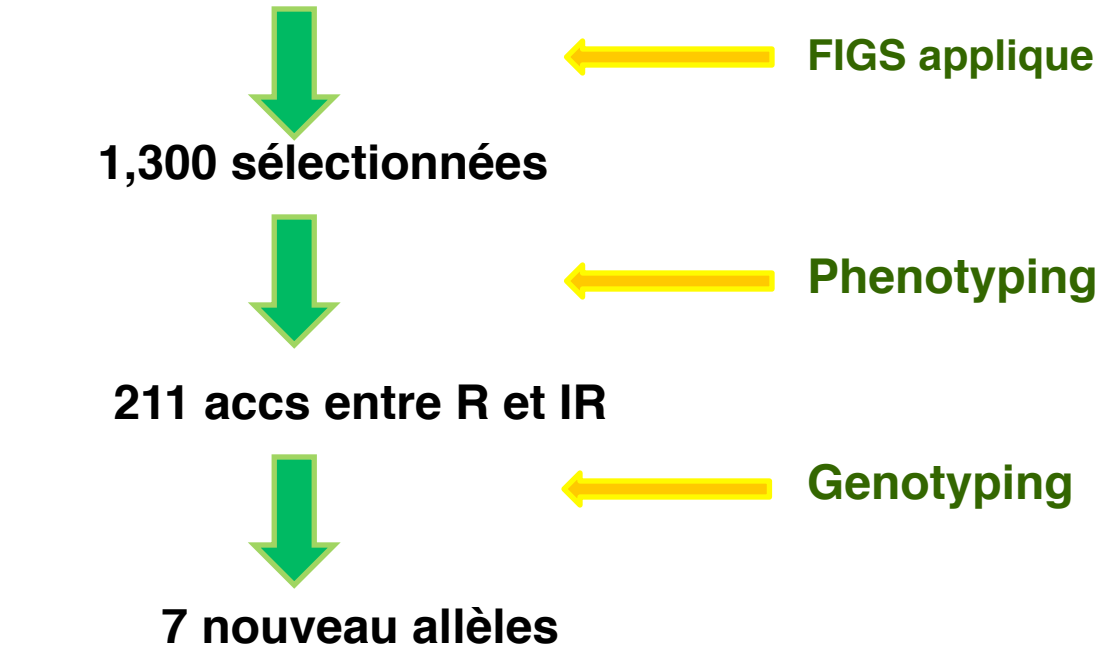
approche FIGS

FIGS a permis d'identifier les caractères chez les plantes qui étaient pendant si longtemps convoités par les améliorateurs telles que la résistance à:

- L'oïdium (powdery mildew)
- Puceron russe du blé (Russian Wheat Aphid)
- Punaises de blé (sunn pest)

Braidotti, G.2009. Keys to the gene bank, Biotechnology. *Partners in Research for Development* 16-17.

approche FIGS



Au moins 2 ont la spécificité de race nouvelle

100 ans de génétique classiques = 7 allèles

GRDC
Grain Research & Development Corporation

Kaur K; Street K; Mackay M; Yahiaoui N; Keller B (2008). Allele mining and sequence diversity at the wheat powdery mildew resistance locus Pm3. 11th IWGS, 24-29 Aug., Brisbane)

l'idée

G x E --> variation génétique

Pouvons-nous utiliser les mêmes principes de l' évolution dans le sens inverse pour identifier les environnements qui «engendrent« le caractère de variation génétique spécifique

Variation génétique ← E x G

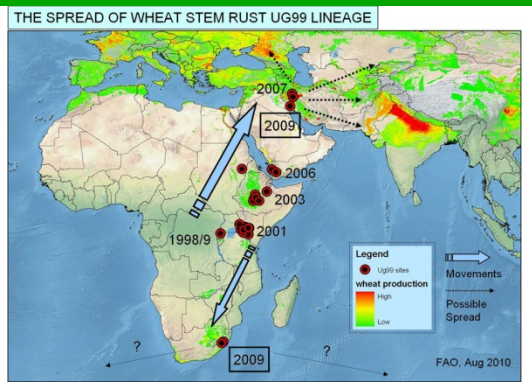
Grain Research & Development Corporation

exemples de variation de caractères sous l'influence de l'environnement (Source: M. Mackay)

caractère	Espèce	Environnent influence	Référence
tolérance à la toxicité du bore	Le blé tendre	Le type de sol	Mackay (1990)
résistance au puceron russe du blé (RWA)	**Le blé tendre**	**Altitude, la température en hiver, la distribution de RWA**	**Bohssini, et al 2009**
résistance à la sécheresse	*Triticum dicoccoides*	Température, aridité	Peleg, Fahima et al. 2005
Couleur et la longueur des glumes	Le blé dur	Altitude	chere, Belay et al. 1996
Date d'épiaison, la longueur chaume, biomasse, rendement en grains et ses composants	*Triticum dicoccoides*	Climat, sol et disponibilité en eau	Bhârat and Nievo 2004
la diversité en gluténine	Le blé dur	Les précipitations, la température minimale au mois de Janvier. l'altitude.	Vanhintum and Elings 1991

caractère (rouille de la tige = Y)

Données de caractère (Y comme variable dépendante)



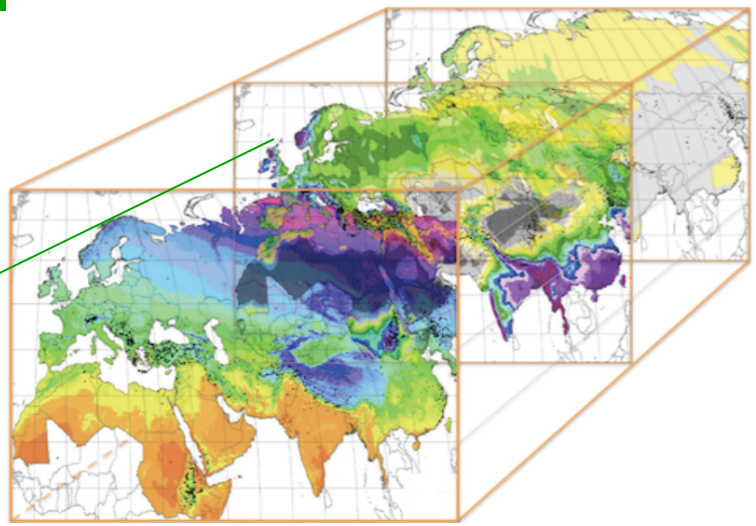
http://www.news.cornell.edu/

site_code1	R_state0	R_state1	R_state2	R_state3	R_state4	R_state5	R_state6	R_state7	R_state8	R_state9
ETH-S893	0	0	0	0	0	0	0	0	1	0
ETH-S1222	0	0	0	0	0	0	0	0	0	1
NS_339	0	0	0	0	0	0	1	0	1	0
ETH-S1153	0	0	0	0	2	1	3	0	0	0
NS_415	0	0	0	0	0	0	1	0	0	0
NS_424	0	0	0	1	0	0	0	0	0	0
ETH64:55	0	0	1	0	0	0	0	0	0	0
NS_525	0	0	0	0	0	0	1	0	0	0
NS_526	0	1	2	1	2	0	3	0	0	0
NS_559	2	5	1	0	0	2	0	0	0	0
ETH64:53	0	0	1	0	0	0	0	0	0	0

Source: (USDA) National Genetic Resources Program (NGRP) GRIN database

données éco-climatiques (X)

Les bases de données éco-climatiques (ICARDA): température moyenne annuelle (devant), les précipitations annuelles (au milieu) et les temperatures (derrière) (De Pauw 2008)



Climate data (X as independent variables)

site_code1	prec01	prec02	prec03	prec04	prec05	.....	ari01	ari02	ari03	ari04	ari05
ETH-S893	25	36	72	154.22	148.88		0.167	0.246	0.439	1.098	1.169
ETH-S1222	29	44	92	167.46	168		0.223	0.344	0.646	1.354	1.612
NS_339	44	67	130.43	177.96	185.74		0.351	0.552	0.949	1.457	1.751
ETH-S1153	36	48	86	140.92	131.94		0.28	0.39	0.609	1.108	1.078
NS_415	32	46.61	95.42	150.3	157		0.271	0.419	0.732	1.289	1.437
NS_424	31.94	45	90	143.62	150		0.257	0.38	0.641	1.146	1.272
ETH64:55	28	38.26	57	97.57	81		0.247	0.344	0.45	0.834	0.662
NS_525	28	39	57	97.13	80.78		0.248	0.352	0.452	0.836	0.669
NS_526	27	39	57	97.01	80.77		0.241	0.354	0.455	0.842	0.68
NS_559	23	40	61.89	129.04	102		0.226	0.397	0.511	1.206	0.998

Source: International Center for Agricultural Research in the Dry Areas (ICARDA)

plateforme

Language R
(Development des algorithmes)

```
> Data transformation (λ)
> Model <- model(caractere ~ climate)
> Measuring accuracy metrics
> ….
```

Système d'Information Géographique (SIG)

Arc Gis
Environnemental data/layers
(surfaces)

Modeling purpose

Generation des doneness environment ales

Grain Research & Development Corporation

preparation des données

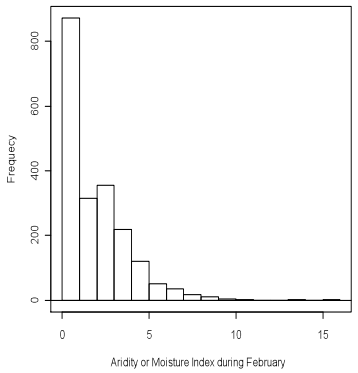
Power relationship $\mu \sim \sigma^{2(p)}$ (spread)

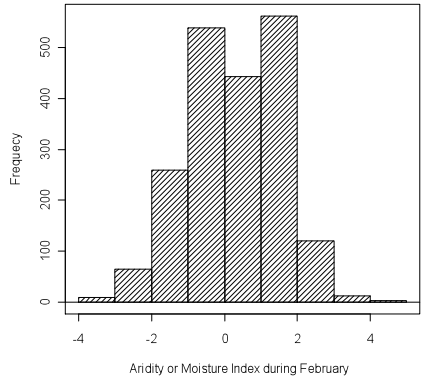
$$f_\lambda(x) = \begin{cases} \dfrac{x^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \log(x) & \lambda = 0 \end{cases}$$

$$l(\lambda) = -\frac{n}{2}\log_e\left[\frac{1}{n}\sum\left(x_j^\lambda - \overline{x^\lambda}\right)^2\right] + (\lambda - 1)\sum_{j=1}^{n}\log_e(x_j)$$

Climate data (X as independent variables)

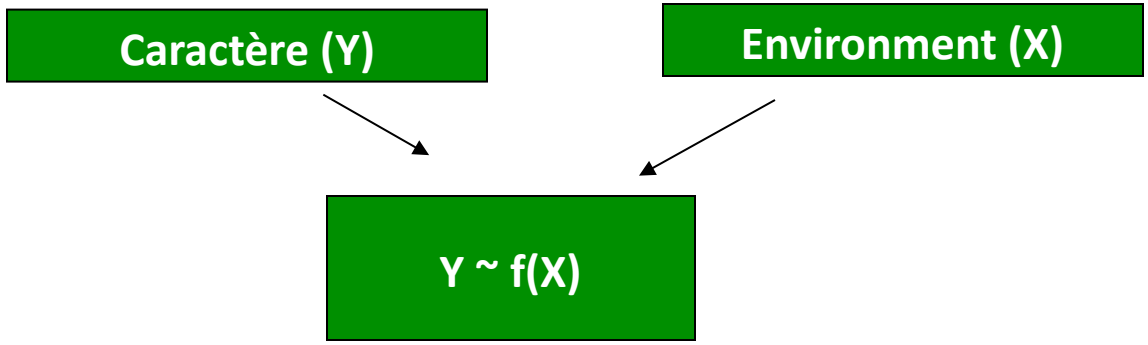
site_code	.....	**ari02**	.....
ETH-S893		**0.246**	
ETH-S1222		**0.344**	
NS_339		**0.552**	
ETH-S1153		**0.390**	
NS_415		**0.419**	
NS_424		**0.380**	
ETH64:55		**0.344**	
NS_525		**0.352**	
NS_526		**0.354**	
NS_559		**0.397**	





GRDC
Grain Research & Development Corporation

cadre de modélisation



D'abord l'approche linéaire indépendamment de la distribution sous-jacente décrivant les données

$$Yi \sim N\left(\beta_0 + \sum_{i=1}^{n} \beta_i X_i \,.\, \sigma_\epsilon^2\right)$$

$$Yi \sim B\left(1, \Phi(\beta_0 + \sum_{i=1}^{n} \beta_i X_i)\right)$$

X est l'ensemble des variables qui contient des variables explicatives ou prédicateurs (données sur le climat) où $X \in Rm$, $Y \in Y$ qui est soit une réponse catégorique (label) ou un numérique (caractères / états des descripteurs).

GRDC
Grain Research & Development Corporation

cadre de modélisation

L'analyse en composantes principales (ACP)
Partial Least Square (PLS)
Partitionnement récursif (RF)
Support Vector Machines (SVM)
Réseaux de Neurones (NN)

Bari A., Street K., Mackay M., Endresen D.T.F., De Pauw E. & Amri A. (2011) Focused identification of germplasm strategy (FIGS) detects wheat stem rust resistance linked to environmental variables. Genetic Resources and Crop Evolution (in press)

vient de paraître cette semaine
http://www.springerlink.com/content/m7140x68v2065113/fulltext.pdf

Grain Research & Development Corporation

mesures de précision

Les paramètres qui fournissent des informations sur la spécificité («caractère agro-climatique»)

Confusion matrix (2-by-2 contingency table)

		Observed	
		Resistant	Susceptible
Predicted	Resistant	a	b
	Susceptible	c	d

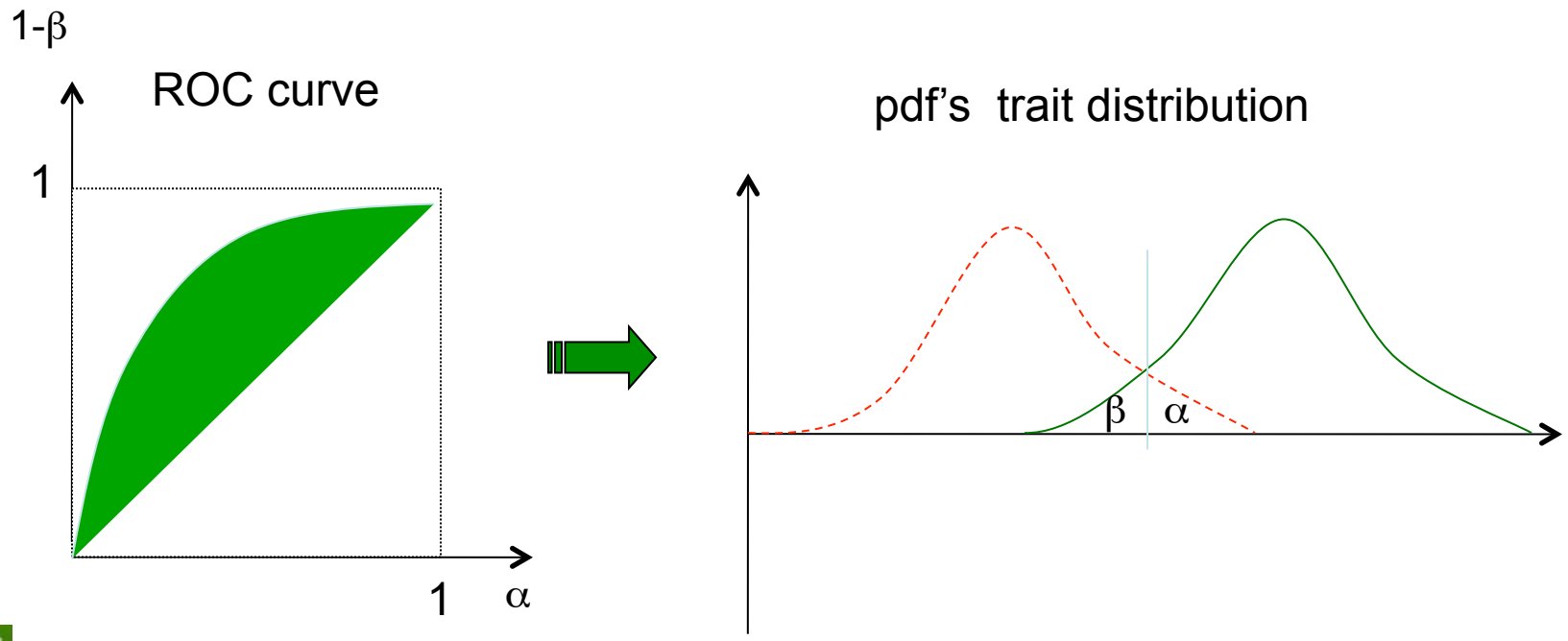
Sensitivité a/ (a + c) = α
Spécificité d/(b + d) = β

α et β sont des indicateurs de la capacité des modèles de classer correctement les observations.

GRDC
Grain Research & Development Corporation

mesures de précision

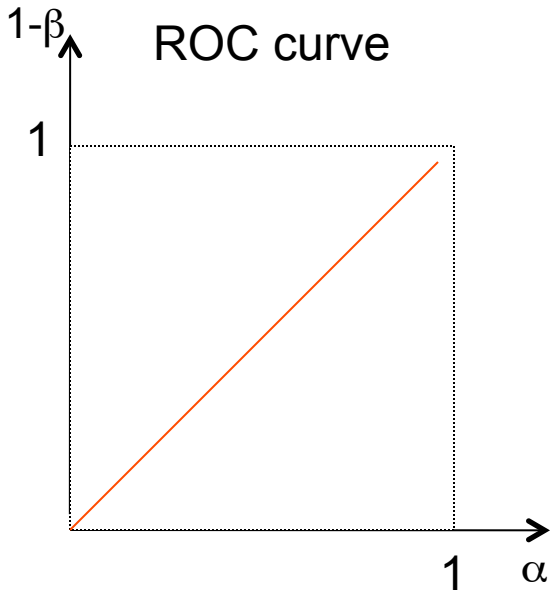
Les paramètres qui fournissent des informations sur la spécificité («caractère agro-climatique») ..

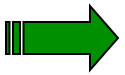


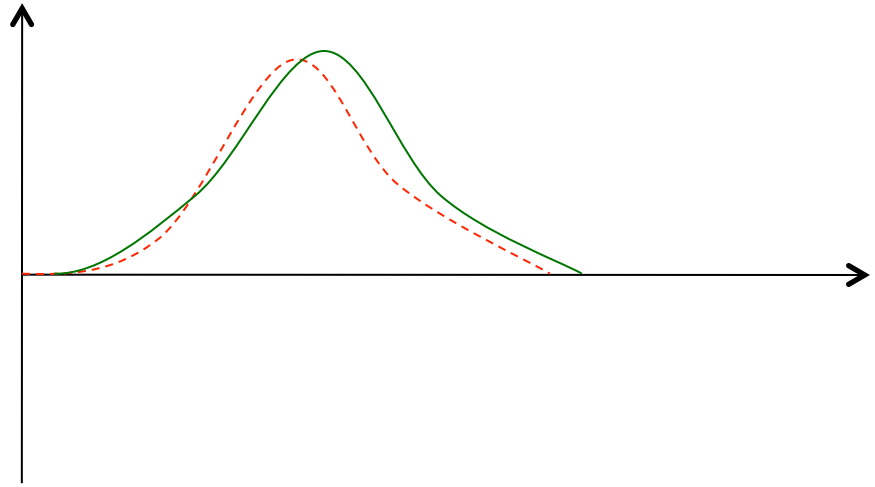
La courbe ROC et le PDF résultants de la distribution de caractère (trait state)

mesures de précision

L'aleatoire

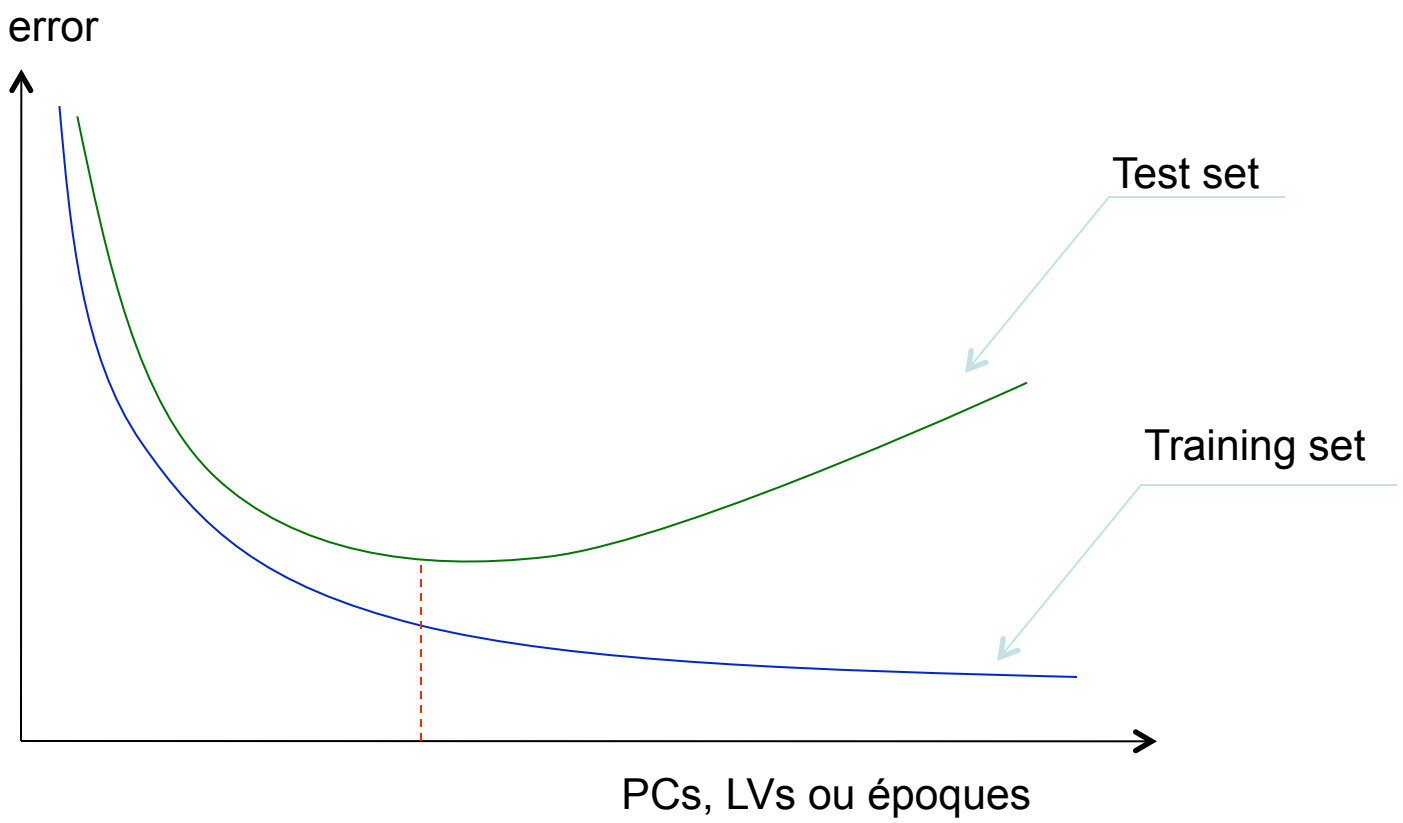






Grain Research & Development Corporation

optimisation/tuning



Tendance de l'erreur par rapport au nombre de composantes (PC / LV) ou époques (NN)

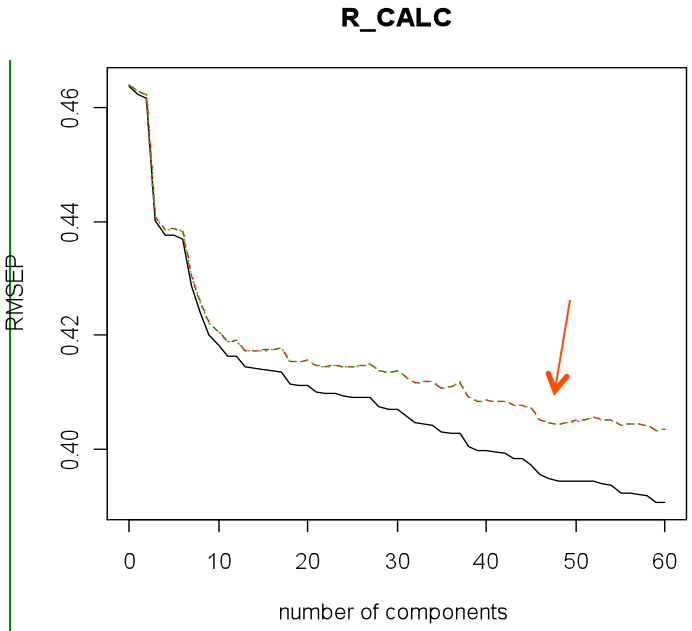
les résultats

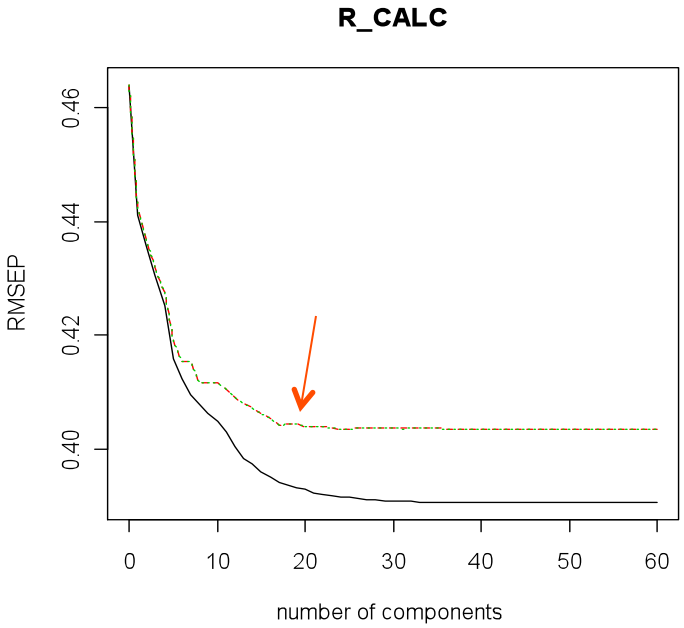
Table: Précision / Accord pour les différents modèles
μ: : moyenne, L: Basse CI limite, U: la limite CI limit

Model	*	AUC	Overall	Kappa	LK+
pls	μ	**0.69**	0.75	**0.40**	3.86
	L	**0.68**	0.75	**0.38**	3.49
	U	**0.70**	0.76	**0.42**	4.24
rf	μ	**0.70**	0.76	**0.42**	3.81
	L	**0.69**	0.75	**0.40**	3.55
	U	**0.72**	0.77	**0.45**	4.08

Grain Research & Development Corporation

les résultats

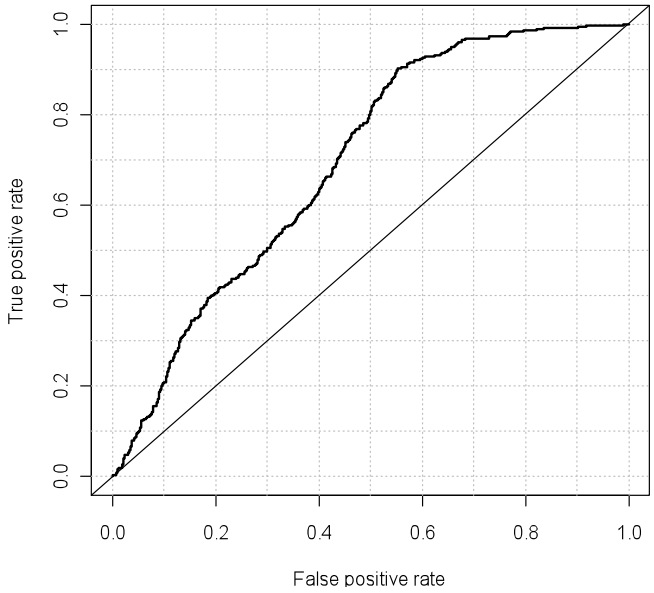


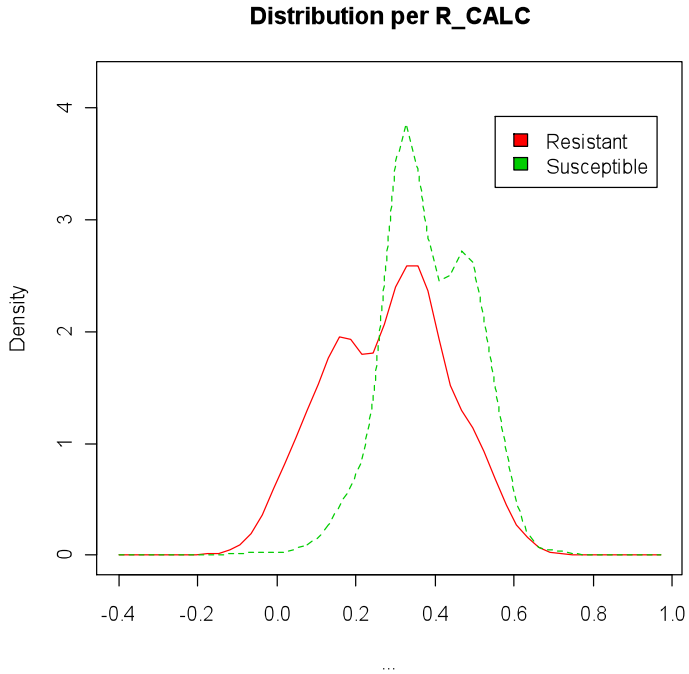


Les flèches indiquent la où le nombre de composantes (PC et LV) a été sélectionné pour la prédiction

PCA

PC5

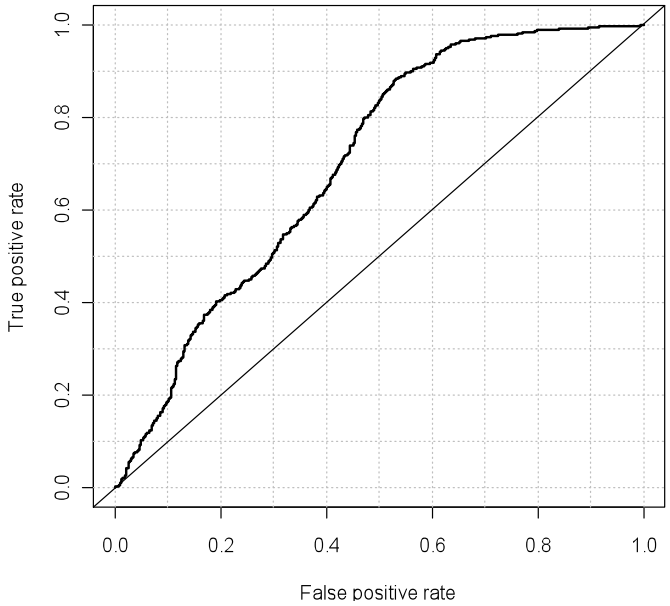


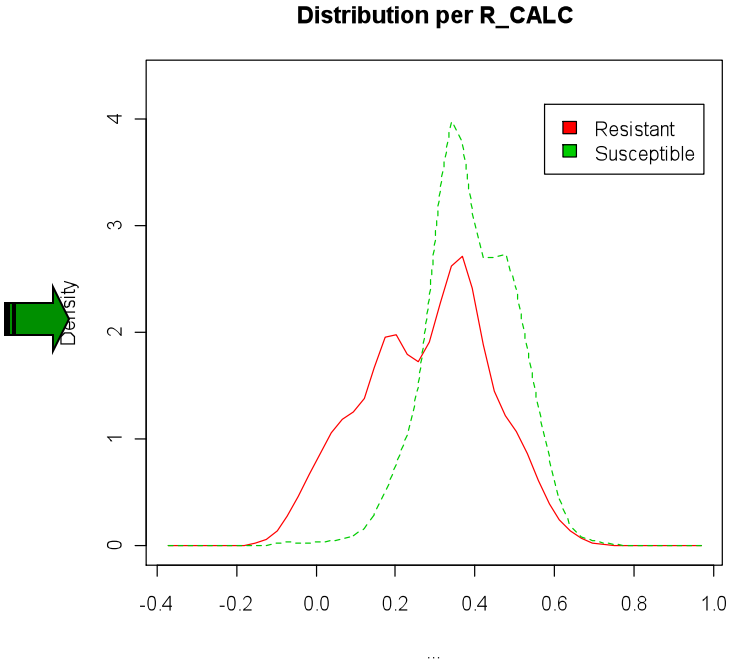


Grain Research & Development Corporation

PLS

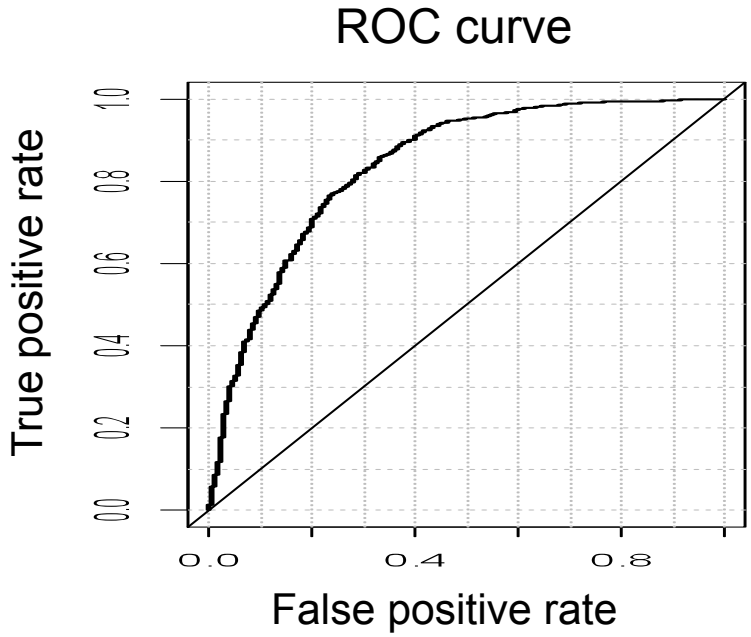
LV2

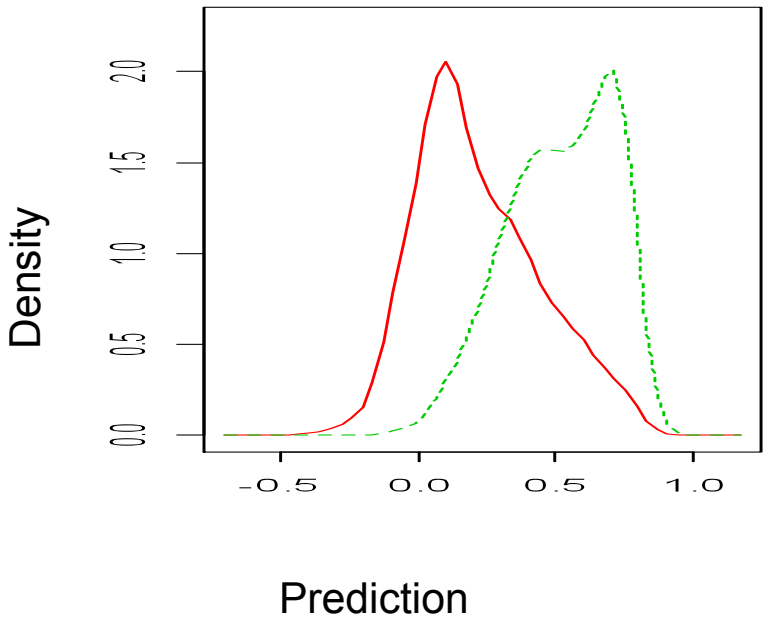

True positive rate
False positive rate
0.0
0.2
0.4
0.6
0.8
1.0


Distribution per R_CALC
Resistant
Susceptible
Density
0
1
2
3
4
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0

PLS (optimisée)

- Principal component analysis (PCA)
- **Partial Least Square (PLS)**
- Random Forest (RF)
- Support Vector Machines (SVM)
- Neural Networks (NN)

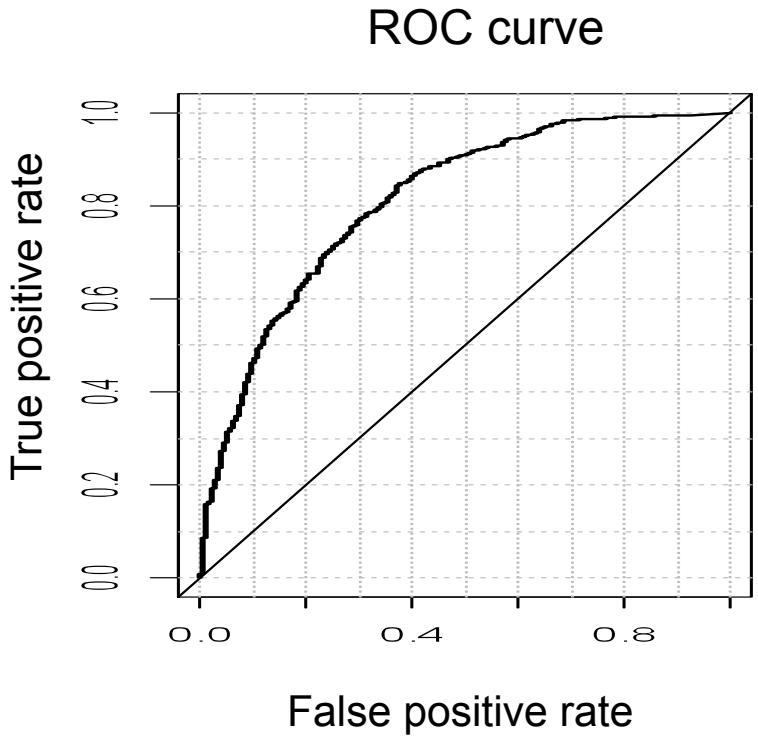


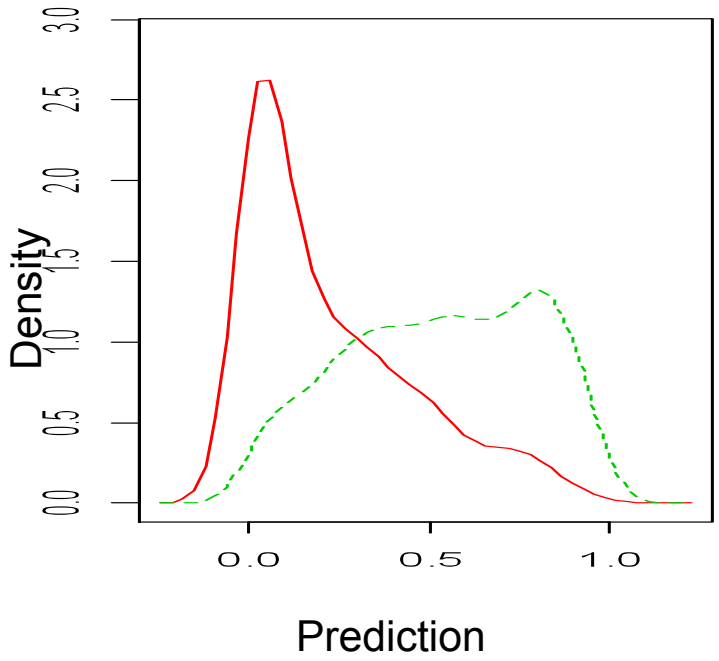


GRDC
Grain Research & Development Corporation

RF

- Principal component analysis (PCA)
- Partial Least Square (PLS)
- **Random Forest (RF)**
- Support Vector Machines (SVM)
- Neural Networks (NN)

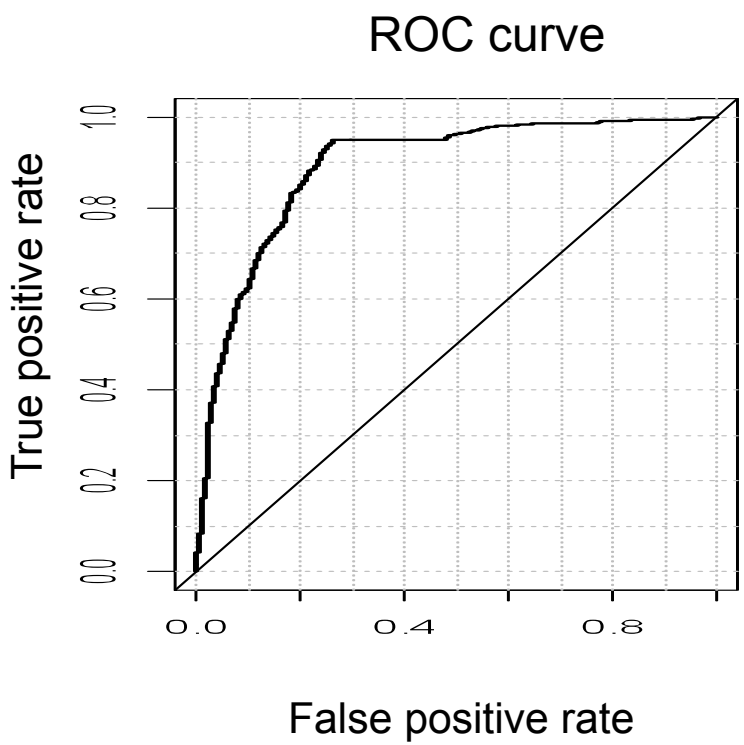


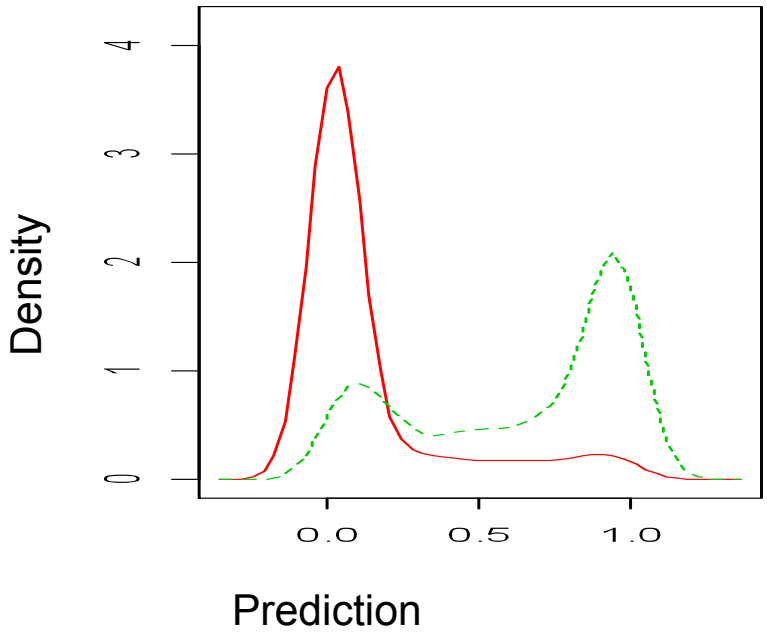


GRDC
Grain Research & Development Corporation

SVM

- Principal component analysis (PCA)
- Partial Least Square (PLS)
- Random Forest (RF)
- **Support Vector Machines (SVM)**
- Neural Networks (NN)





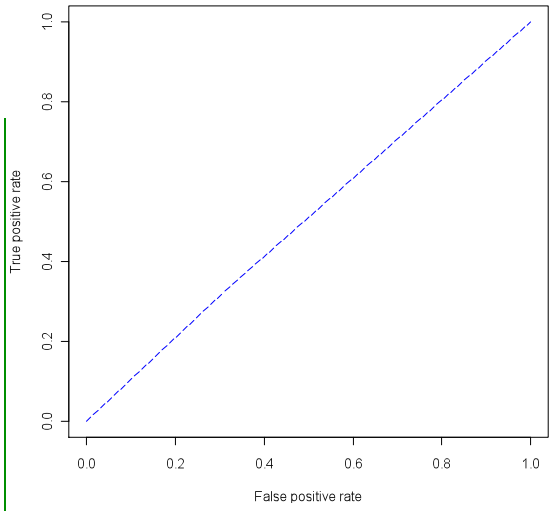
Grain Research & Development Corporation

les résultats

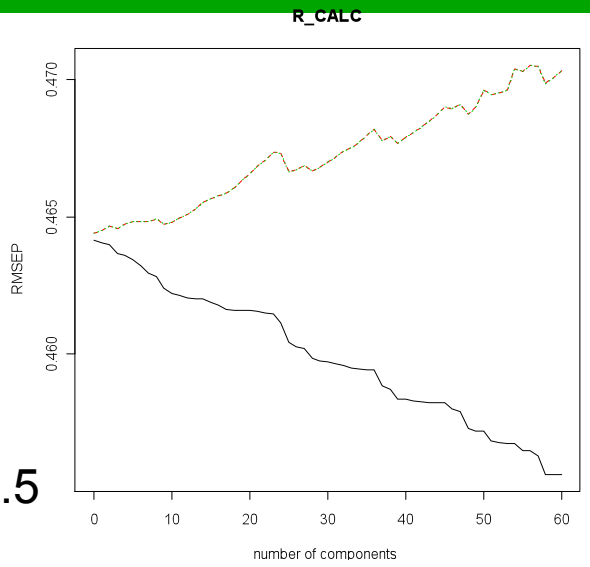
Tableau: Précision / Accord métriques d'une distribution aléatoire des caractères de résistance μ: : moyenne, L: Basse CI limite, U: la limite supérieure de CI

Model	*	AUC	Overall	Kappa	LK+
pls	μ	**0.51**	0.67	**0.02**	1.50
	L	**0.50**	0.66	**0.00**	0.99
	U	**0.51**	0.69	**0.03**	2.01
rf	μ	**0.53**	0.65	**0.06**	1.36
	L	**0.52**	0.64	**0.05**	1.27
	U	**0.53**	0.65	**0.07**	1.46

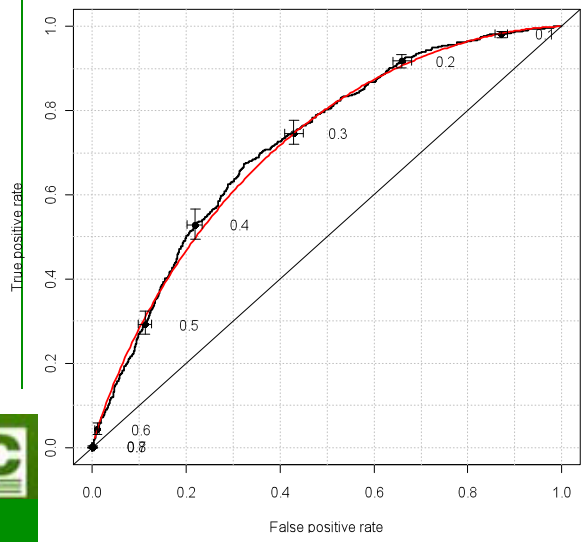
aléatoire (hypothetical)

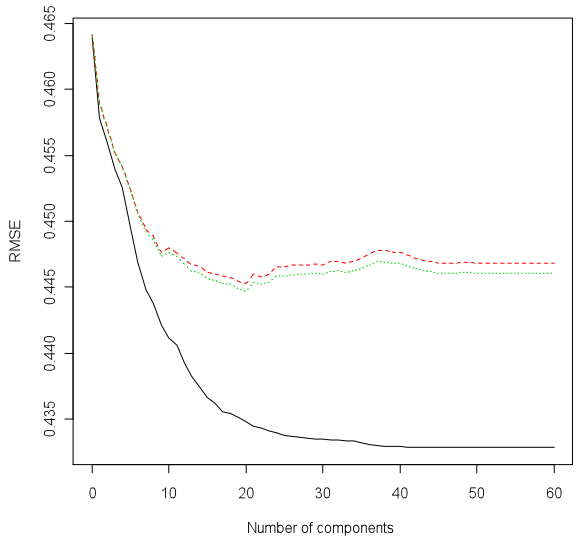


AUC ~ 0.5



distribution aléatoire totale de caractère de la résistance à la tige de la rouille





distribution aléatoire partielle de caractère de la résistance à la tige de la rouille

Grain Research & Development Corporation

les résultats

- L'approche FIGS s'est avérée très efficace pour discerner entre les environnements qui engendre des accessions ayant le caractère de résistance vis-à-vis des accessions qui n'en ont pas [1].

- FIGS a pu aider identifier des échantillons avec une probabilité plus élevée de contenir le caractère de résistance a la rouille de la tige [2].

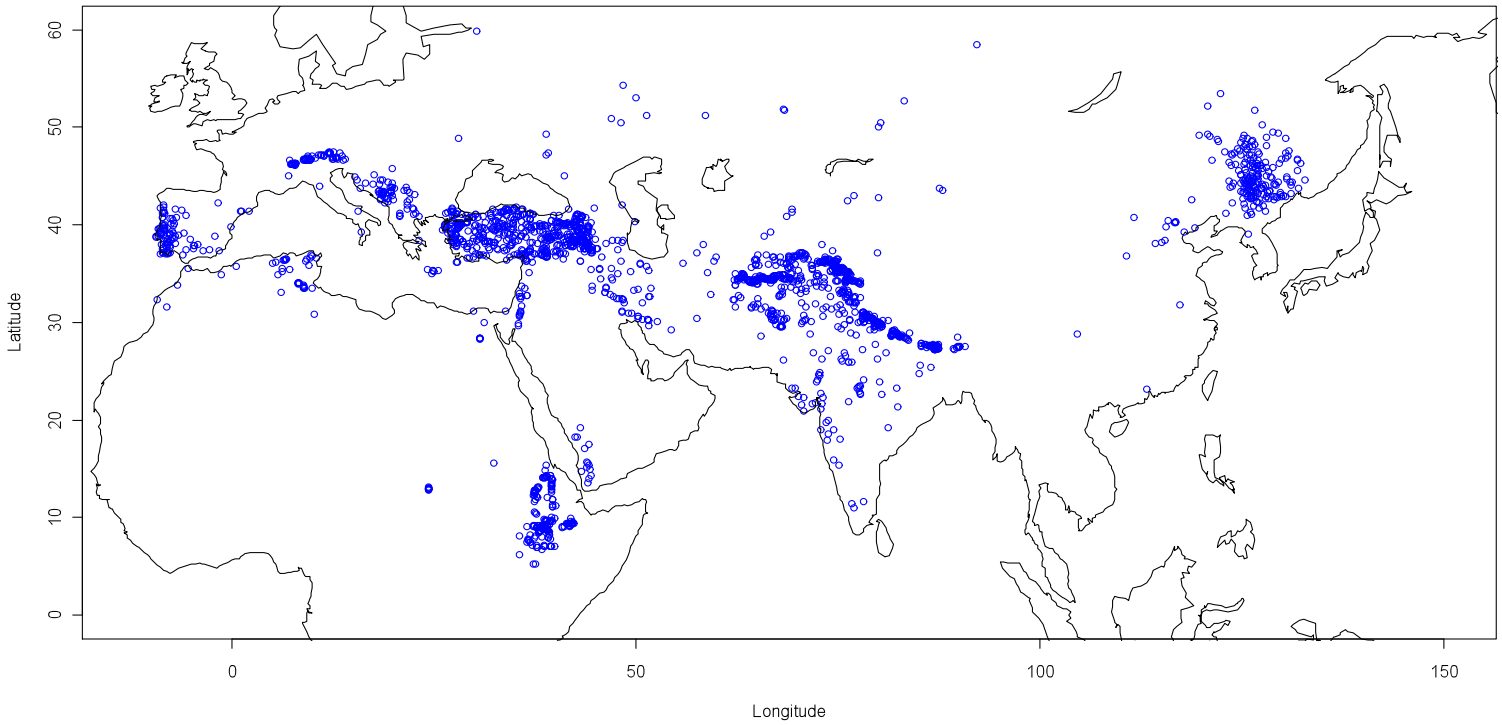
3728 accs (inconnues) -> 500 -> 129 accs (25.8%)

1) Bari A., Street K., Mackay M., Endresen D.T.F., De Pauw E. & Amri A. (2011) Focused identification of germplasm strategy (FIGS) detects wheat stem rust resistance linked to environmental variables. Genetic Resources and Crop Evolution. http://www.springerlink.com/content/m7140x68v2065113/fulltext.pdf

2) Endresen D.T.F., Street K., Mackay M., Bari A. De Pauw E. & Amri A. (in press) Sources of resistance to stem rust (Ug99) in bread wheat and durum wheat identified using Focused Identification of Germplasm Strategy (FIGS). Crop Science

Grain Research & Development Corporation

rouille de la tige (stem rust)

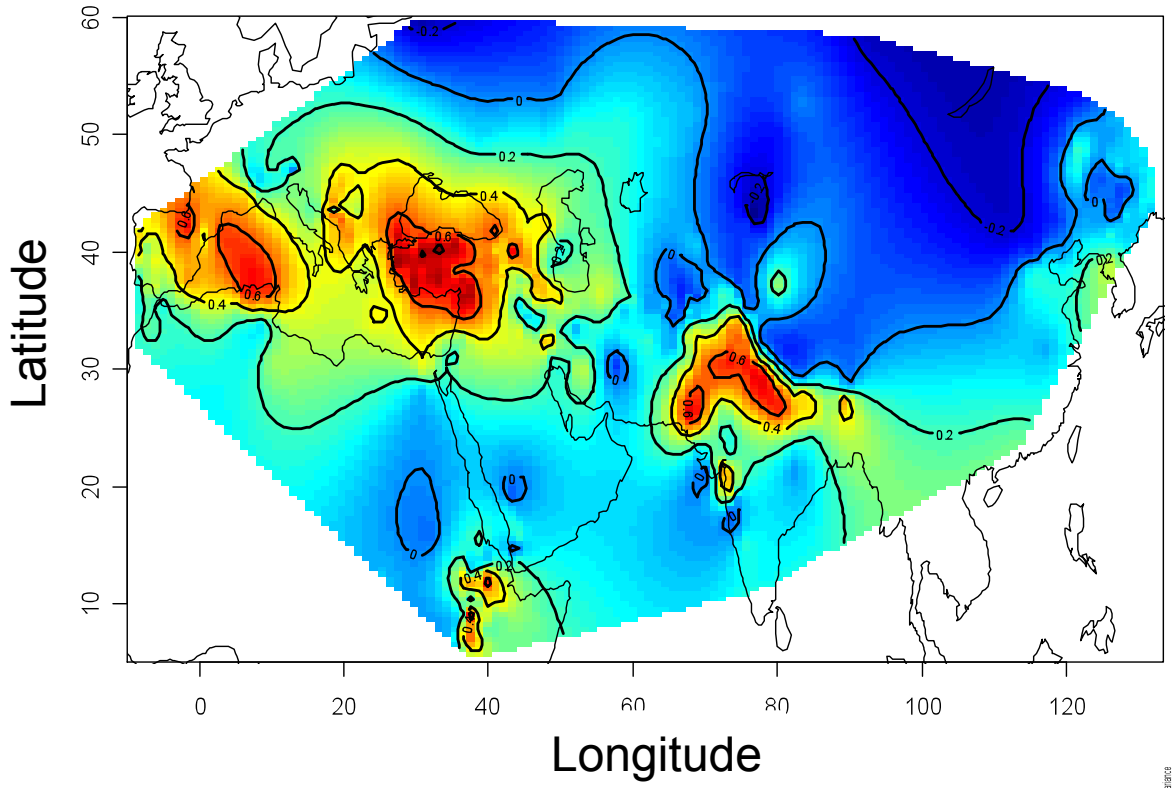


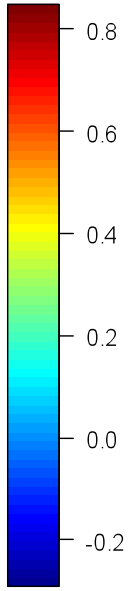
Distribution des accessions de blé en relation avec la rouille de la tige

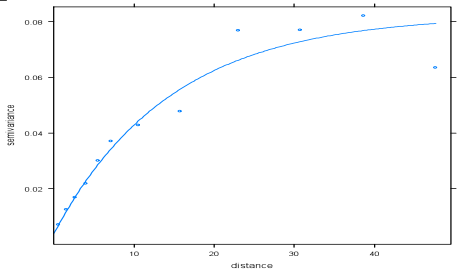
Grain Research & Development Corporation

PLS

Régions où la résistance est susceptible de se produire (rouge foncé)



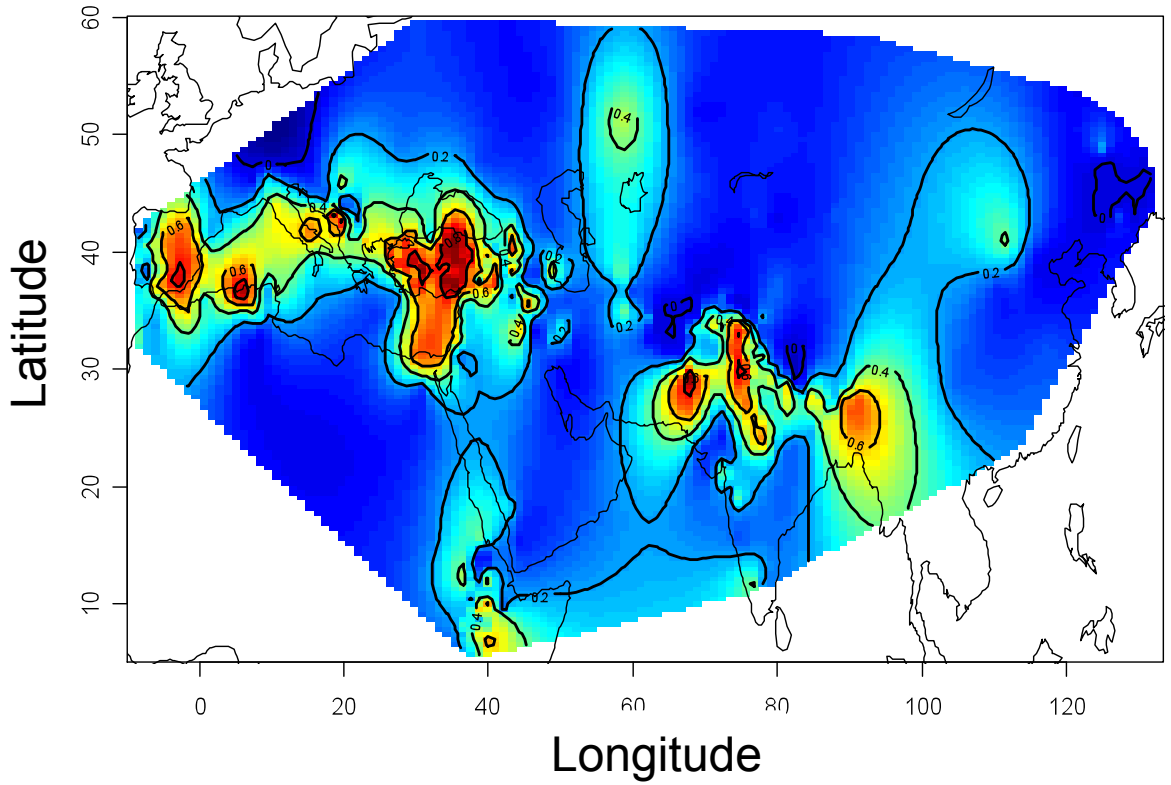


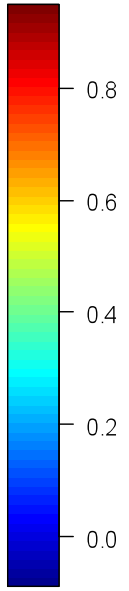


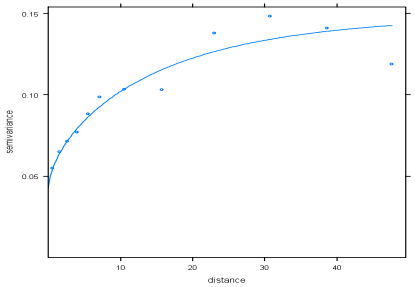
GRDC
Grain Research & Development Corporation

Random Forest (RF)

Régions où la résistance est susceptible de se produire (rouge foncé)







GRDC
Grain Research & Development Corporation

outputs

- **FIGS plateforme/outils**
  - Le cadre conceptuel (modélisation)
  - Des algorithmes
- **Publications**
  - Endresen D.T.F, Street K., Mackay M., Bari A., De Pauw E (***2011***). Predictive association between biotic stress caracteres and ecogeographic data for wheat and barley landraces. https://www.crops.org/publications/cs/new-articles
  - Bari A., Street K., Mackay M., Endresen D.T.F., De Pauw E. & Amri A. ((2011***)*** Focused identification of germplasm strategy (FIGS) detects wheat stem rust resistance linked to environmental variables.
  - Bari A., Street K., Mackay,M., Endresen D.T.F., De Pauw E., & Amri A. Mining genetic resource collections for useful traits using Focused Identification of Germplasm Strategy (FIGS). http://atlas-conferences.com/cgi-bin/abstract/cbcr-78
  - Sources of resistance to stem rust (Ug99) in bread wheat and durum wheat identified using Focused Identification of Germplasm Strategy (FIGS). (***submitted)***

Grain Research & Development Corporation

consultation lancée / initiée

ICARDA International Center for Agricultural Research in the Dry Areas

For immediate release

A new approach to mining agricultural gene banks promises to speed the pace of research innovation for food security

Research team for 'Focused identification of Germplasm Strategy (FIGS)' opens global consultation to enrich this new tool

December 4, 2011. An innovative new approach to rapidly identifying plant genetic material that can produce new crop varieties – to reduce hunger, fight crop disease and other stresses such as excessive drought and heat – is now set to serve agricultural researchers worldwide. The FIGS method is an innovative alternative to traditional gene bank searching. It increases the speed of innovation and is a strategic new approach for crop researchers and plant breeders worldwide, who are looking to improve crop yields and combat the negative effects of climate change.

GRDC Grain Research & Development Corporation

l équipe FIGS

Kenneth Street
International Center for Agricultural Research in the Dry Areas (ICARDA)

Michael Mackay
Biodiversity Internationa, Italie

Eddy De Pauw
ICARDA

Dag Terje Filip Endresen
Nordic Genetic Resources Center (NordGen), Suède

Ahmed Amri
ICARDA

Abdallah Bari
ICARDA


Ken


Michael


Eddy


Dag


Ahmed


Abdallah

Grain Research & Development Corporation