

RMT Modelia, Paris, 16 novembre 2009

L'analyse ROC en agronomie

David Makowski

makowski@grignon.inra.fr



Pourquoi s'intéresser aux erreurs des modèles ?



Pourquoi s'intéresser aux erreurs des modèles ?

i. Les modèles font plusieurs types d'erreur

- Erreur de prédiction
- Erreur de classement
- Erreur de décision
- Erreur d'estimation



Pourquoi s'intéresser aux erreurs des modèles ?

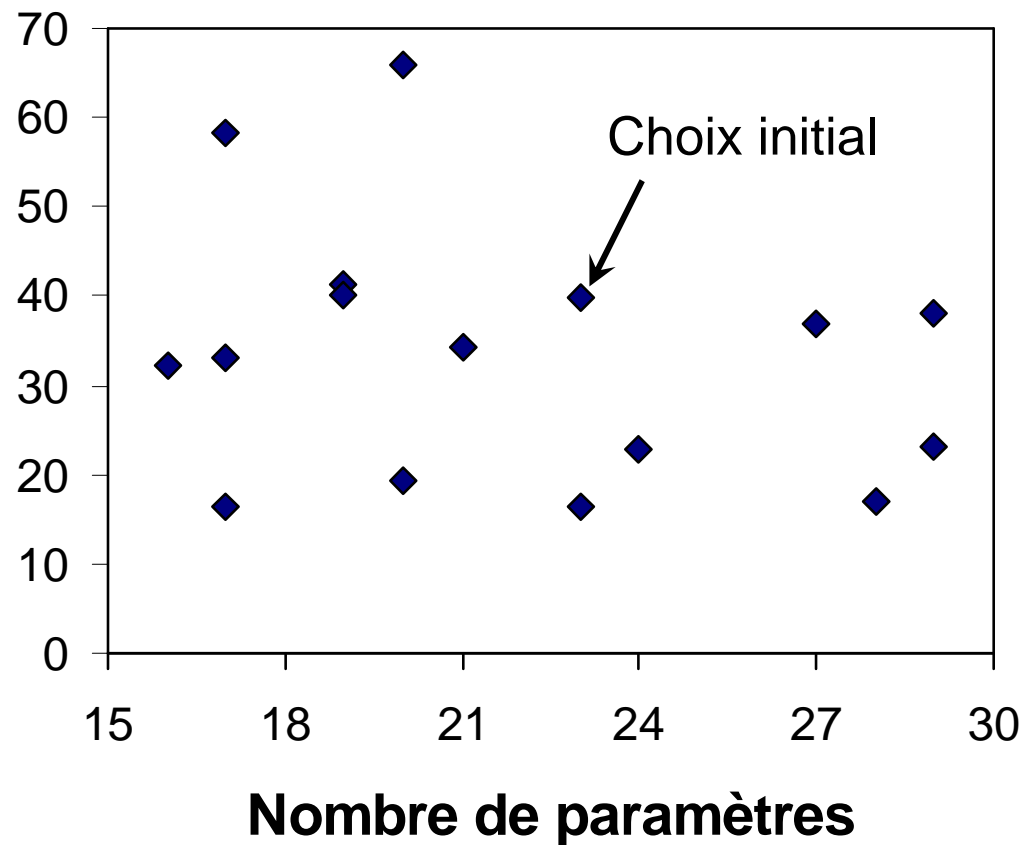
ii. Évaluer les erreurs pour faciliter le choix d'un modèle

- Plusieurs modèles sont souvent disponibles pour un usage donné.
- Difficile de déduire le niveau d'erreur d'un modèle à partir de sa structure ou de son niveau de complexité.
- Résultats des comparaisons de modèles parfois inattendus.



Evaluation des erreurs de 16 modèles prédisant l'incidence du piétin échaudage du blé au stade GS33.

RMSEP (%)



Pourquoi s'intéresser aux erreurs des modèles ?

iii. Donner aux utilisateurs une information sur les risques d'erreurs

- Les modèles agronomiques jouent un rôle de plus en plus important dans les expertises scientifiques à diverses échelles.
- Important de fournir une information sur les erreurs liées à l'utilisation d'un modèle donné.



Pourquoi s'intéresser aux erreurs des modèles ?

iv. Orienter les recherches futures

- Quantifier l'intérêt d'une amélioration des modèles.
- Faire des hypothèses sur les points faibles des modèles et sur la meilleure façon de les améliorer.



Pourquoi s'intéresser aux erreurs des modèles ?

- i. Les modèles peuvent conduire à différents types d'erreur
- ii. Evaluer les erreurs pour faciliter le choix d'un modèle
- iii. Donner aux utilisateurs une information sur les risques d'erreurs
- iv. Orienter les recherches futures



Comment évaluer les erreurs des modèles ?

- Analyses d'incertitude et de sensibilité
- Evaluation de la qualité prédictive
- **Evaluation des erreurs de classement**



« Le MSEP de mon modèle est très mauvais mais ce n'est pas grave. Mon modèle n'est pas utilisé pour prédire mais pour classer ».

Un agronome anonyme.



Règle de décision binaire

« Si la prédiction \hat{I} du modèle est supérieure au seuil de décision S alors je considère que la variable d'intérêt Y est égale à 1. Sinon je considère que $Y=0$. »

« Si $\hat{I} > S$, alors $Y=1$, sinon $Y=0$ »



Exemples d'utilisation de règles de décision binaires en agronomie

Si la prédiction est supérieure à S

- le risque de pollution par les nitrates sera élevé
- la teneur en protéines du blé sera forte
- il y aura une perte de rendement due au sclérotinia
- cette espèce d'oiseau sera présente dans cette prairie



Deux types d'erreurs de classification

Faux positif: $I > S$ mais $Y=0$

Faux négatif: $I < S$ mais $Y=1$



Décision optimale

		$Y = 0$	$Y = 1$
Décision basée sur le modèle	$I < S$	Vrai négatif	Faux négatif
	$I \geq S$	Faux positif	Vrai positif

Le but de l'analyse ROC est d'estimer les fréquences de faux négatif et de faux positif pour tous les seuils de décision S .



Un exemple détaillé



Objectif

**Evaluer plusieurs indicateurs de risque de pollution
nitrique à partir de données expérimentales**



Les indicateurs

Indicator	Expression
I_1	$I_1 = \text{amount of applied nitrogen}$
I_2	$I_2 = \text{applied nitrogen} + \text{soil mineral nitrogen in winter}$
I_3	$I_3 = [\text{applied nitrogen} + \text{soil mineral nitrogen in winter}]/\text{grain yield}$
I_4	$I_4 = \text{applied nitrogen} - \text{nitrogen content in grain} * \text{grain yield}$
I_5	$I_5 = \text{residual soil mineral nitrogen predicted by a dynamic model}$



Données

Bassin de Bruyères

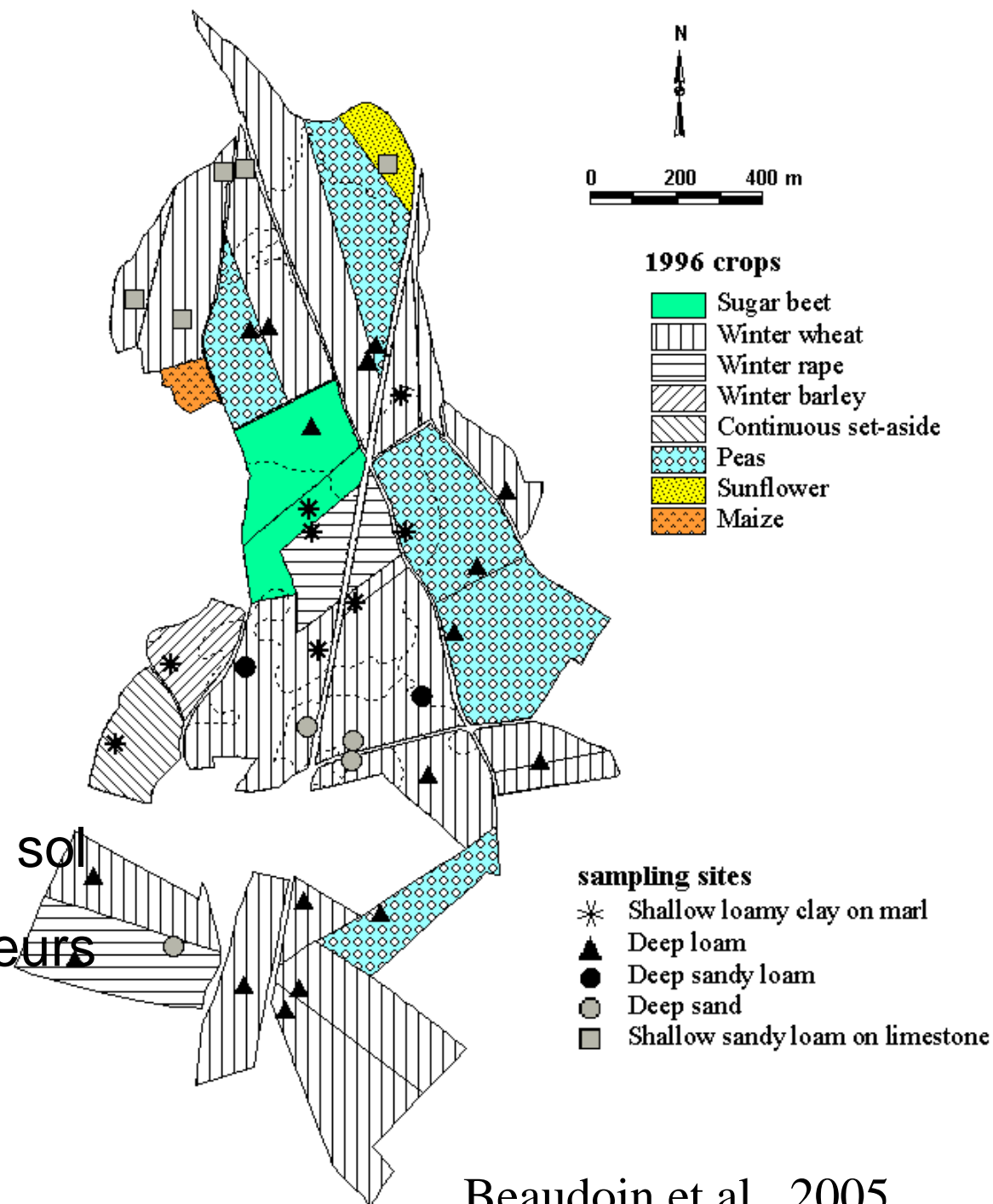
36 sites = 145 ha

10 années = 1991-2000

92 sites-années de blé

Mesures

- Reliquat N récolte dans le sol
- Entrée de tous les indicateurs



Beaudoin et al., 2005



Analyse ROC

		Décision optimale	
		$Y = 0$	$Y = 1$
Décision basée sur le modèle	$I < S$	Vrai négatif	Faux négatif
	$I \geq S$	Faux positif	Vrai positif

Sensibilité = Nombre de vrai positif / Nombre total de situations avec $Y=1$

Spécificité = Nombre de vrai négatif / Nombre total de situations avec $Y=0$



Estimation de la sensibilité et de la spécificité pour un indicateur et tous les seuils de décision S

n sites-années avec $Y=0$ (faible reliquat récolte: $RR < D_{\text{thresh}}$)

m sites-années avec $Y=1$ (fort reliquat récolte: $RR \geq D_{\text{thresh}}$)

(i). Déterminer la valeur I de l'indicateur pour chaque site-année.

(ii). Définir un seuil de décision S .

(iii). **Sensibilité** = $Prob(I \geq S \mid Y=1) = 1 - \text{Proba. Faux négatif}$

(iv). **Spécificité** = $Prob(I < S \mid Y=0) = 1 - \text{Proba. Faux positif}$

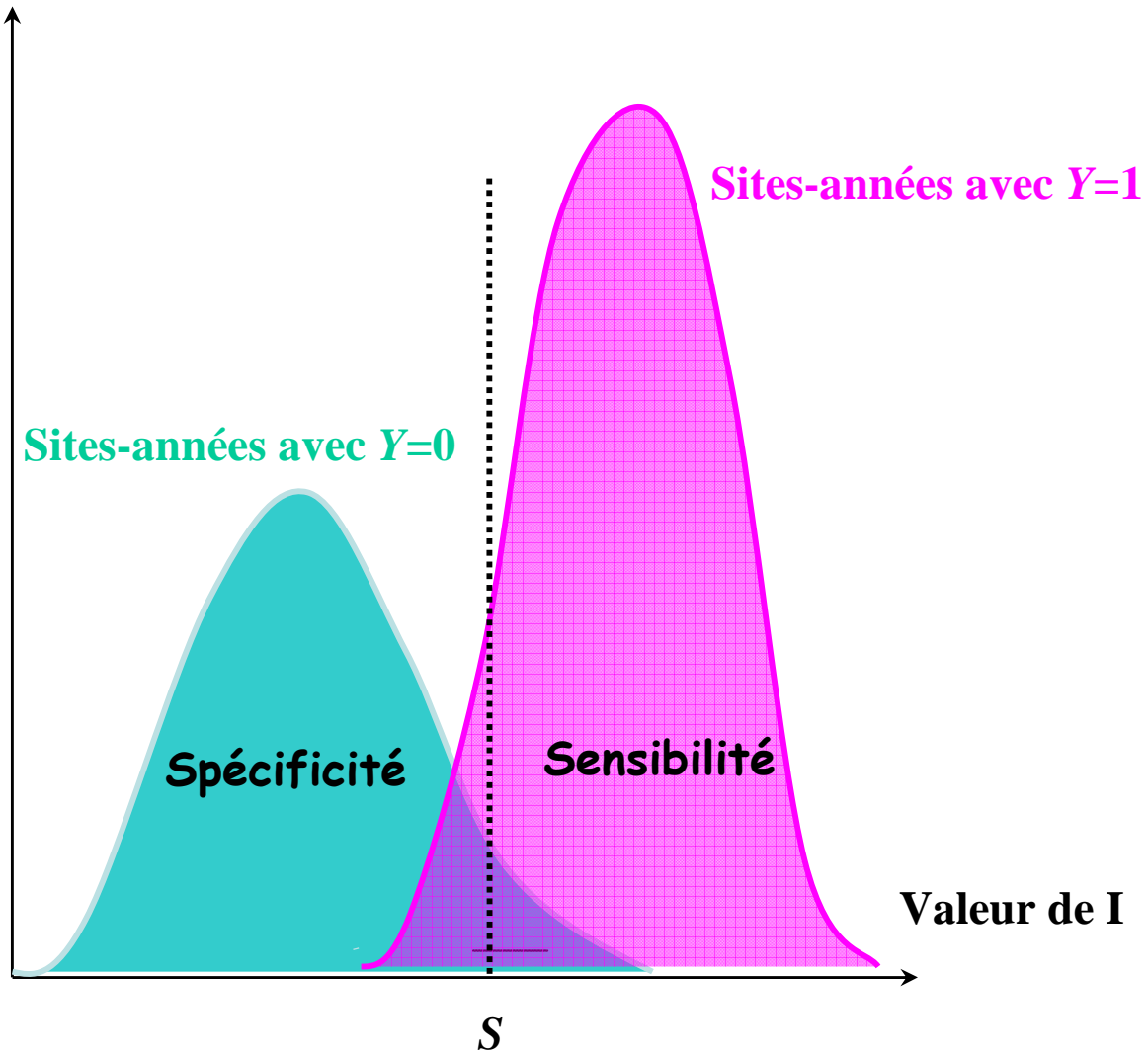
(v). **Courbe ROC**: Sensibilité (S) *versus* $1 - \text{Spécificité}$ (S)

(vi). Estimer l'aire sous la courbe ROC (**AUC**) pour l'indicateur I .

Si $AUC \sim 0.5$, l'indicateur n'est pas meilleur qu'un classement aléatoire.

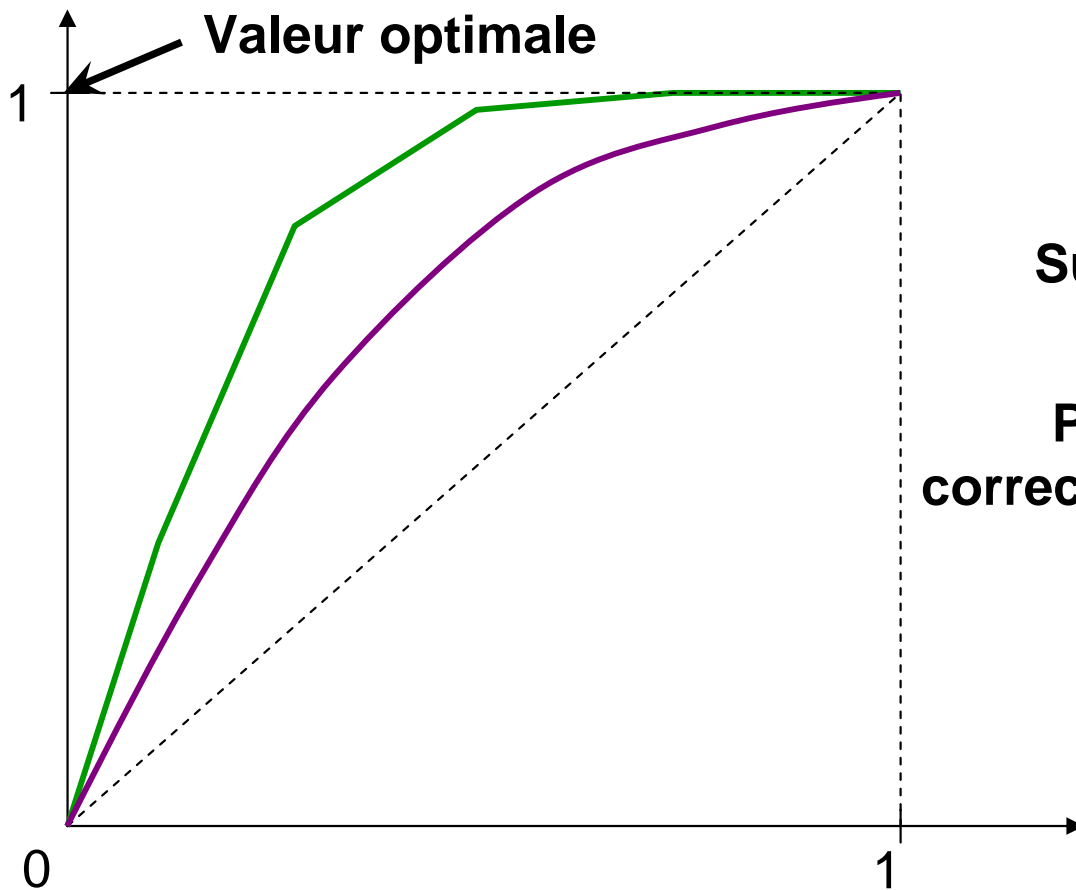


Fréquence



Courbe ROC

Sensibilité



Surface sous la courbe
=
Probabilité de classer
correctement deux sites-années

1 - Spécificité



Code R (library ROCR)

```
TAB<-read.table("f:\\David\\Projets\\BleBruyere.txt",header=T,sep="\t")
```

```
#Gold standard
```

```
RR.t<-30
```

```
Ref<-TAB$RR
```

```
Ref[Ref<RR.t]<-0
```

```
Ref[Ref>=RR.t]<-1
```

```
#ROC analysis
```

```
Ind.list<-c(1,2,3,4,5)
```

```
Auc.vec<-Ind.list
```

```
for (i in 1:length(Ind.list)) {
```

```
    Ind<-TAB[,i]
```

```
    pred<-prediction(Ind,Ref)
```

```
    perf<-performance(pred,"auc")
```

```
    auc<-perf@"y.values"
```

```
    Auc.vec[i]<-as.numeric(auc)
```

```
}
```

```
print(Auc.vec)
```

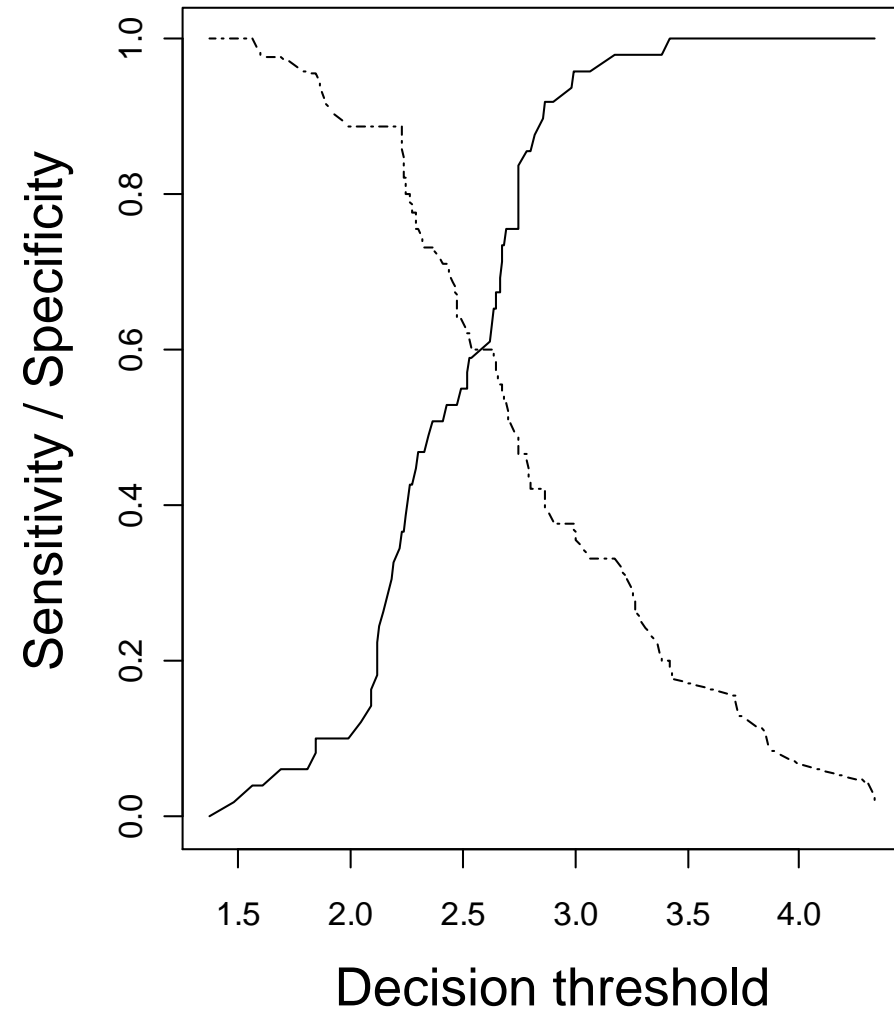
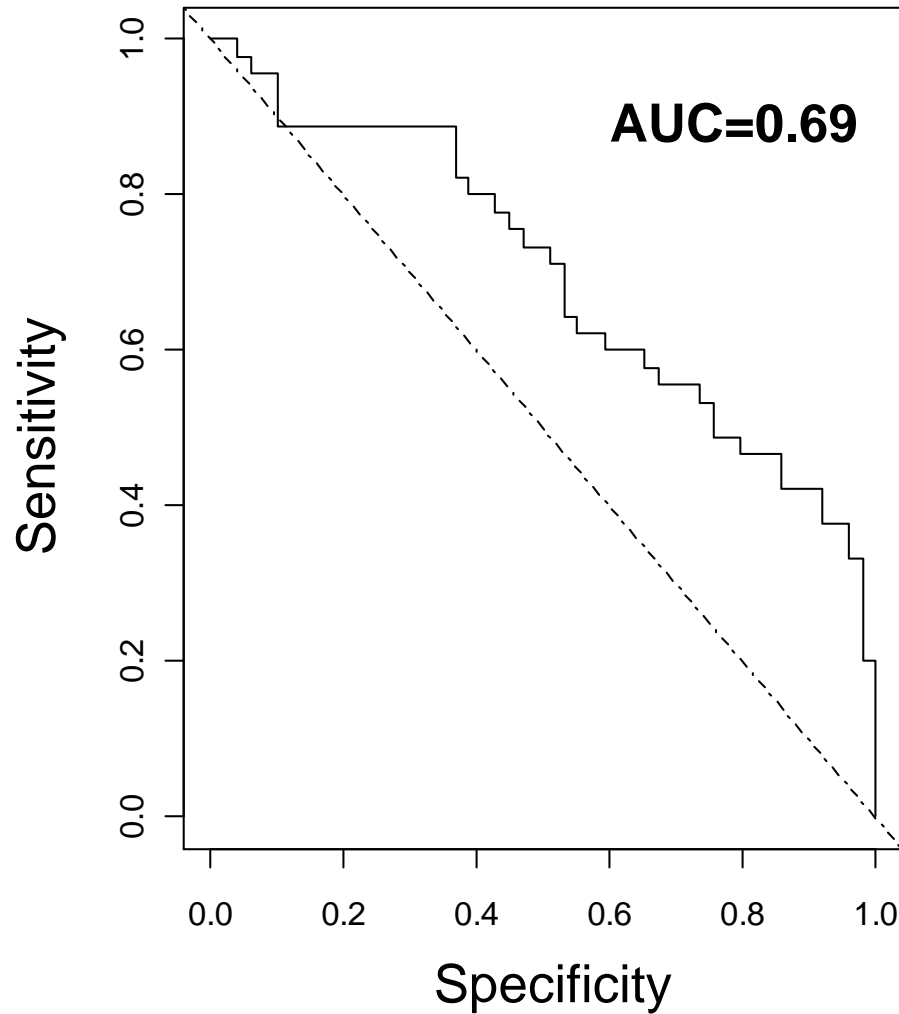


Résultats



Sensibilité, spécificité, et AUC de l'indicateur I_3 ($D_{\text{thresh}} = 30 \text{ kg/ha}$)

$I_3 = [\text{applied nitrogen} + \text{soil mineral nitrogen in winter}]/\text{grain yield}$



Capacité de 5 indicateurs à discriminer les sites-années avec des niveaux élevés et faibles de reliquat N récolte

Indicator	Area under the ROC curve (probability of correct ranking)		
	$D_{thresh}=20$ kg/ha	$D_{thresh}=30$ kg/ha	$D_{thresh}=40$ kg/ha
I_1	0.58 (*)	0.62 (**)	0.61 (**)
I_2	0.64 (**)	0.63 (***)	0.58 (.)
I_3	0.70 (***)	0.69 (***)	0.63 (**)
I_4	0.64 (***)	0.67 (***)	0.63 (**)
I_5	0.45 (.)	0.58 (.)	0.50 (.)



Indicator	Expression
I_1	$I_1 = \text{amount of applied nitrogen}$
I_2	$I_2 = \text{applied nitrogen} + \text{soil mineral nitrogen in winter}$
I_3	$I_3 = [\text{applied nitrogen} + \text{soil mineral nitrogen in winter}]/\text{grain yield}$
I_4	$I_4 = \text{applied nitrogen} - \text{nitrogen content in grain} * \text{grain yield}$
I_5	$I_5 = \text{residual soil mineral nitrogen predicted by a dynamic model}$



Capacité de 5 indicateurs à discriminer les sites-années avec des niveaux élevés et faibles de reliquat N récolte

Indicator	Area under the ROC curve (probability of correct ranking)		
	$D_{thresh}=20$ kg/ha	$D_{thresh}=30$ kg/ha	$D_{thresh}=40$ kg/ha
I_1	0.58 (*)	0.62 (**)	0.61 (**)
I_2	0.64 (**)	0.63 (***)	0.58 (.)
I_3	0.70 (***)	0.69 (***)	0.63 (**)
I_4	0.64 (***)	0.67 (***)	0.63 (**)
I_5	0.45 (.)	0.58 (.)	0.50 (.)

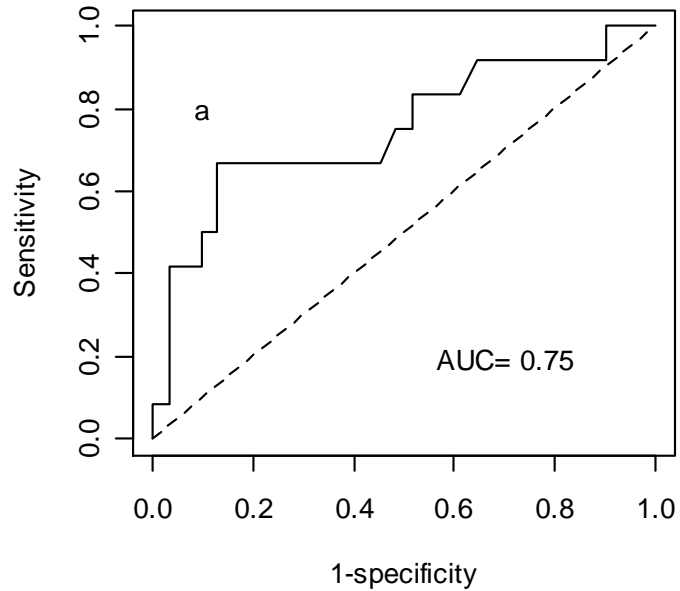


Autres exemples d'applications de l'analyse ROC

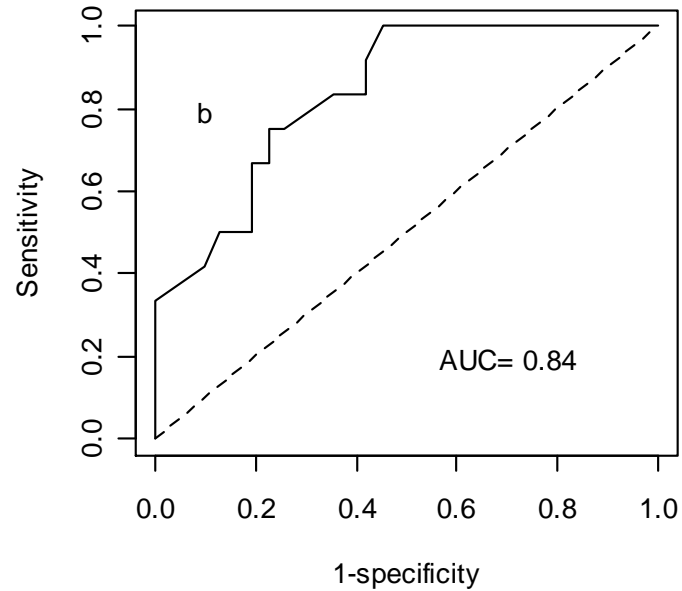


Classification par rapport à un seuil de teneur en protéines

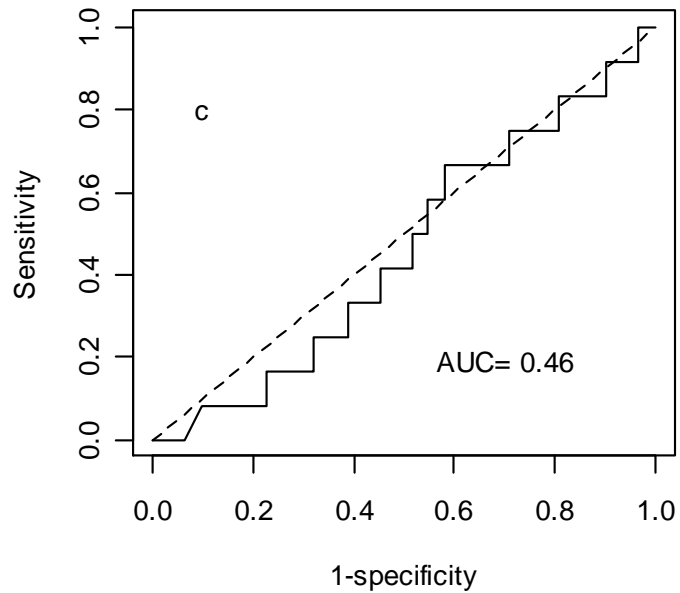
SPAD



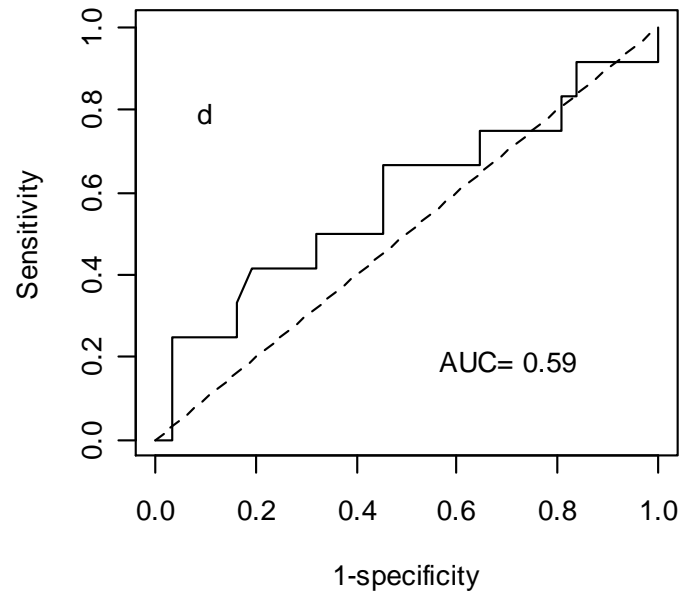
INN



Crop Model 1



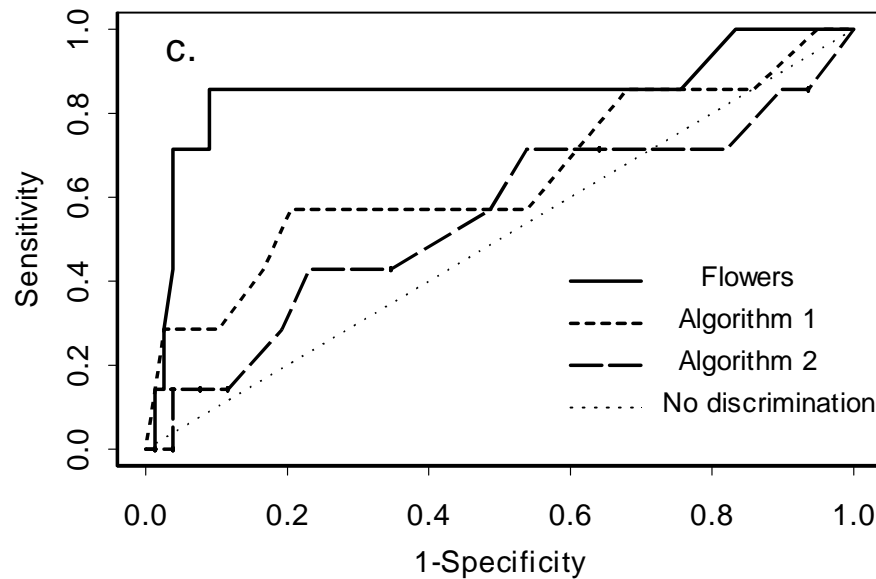
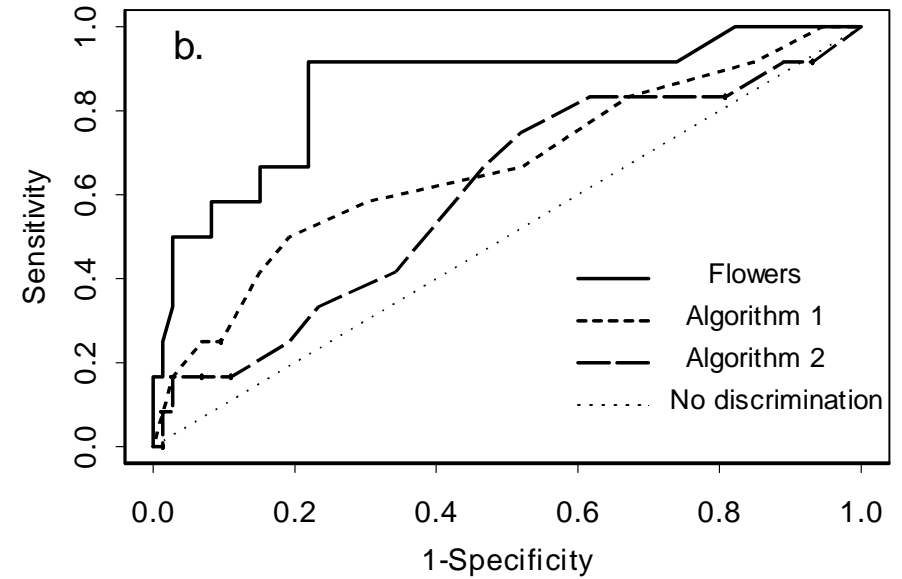
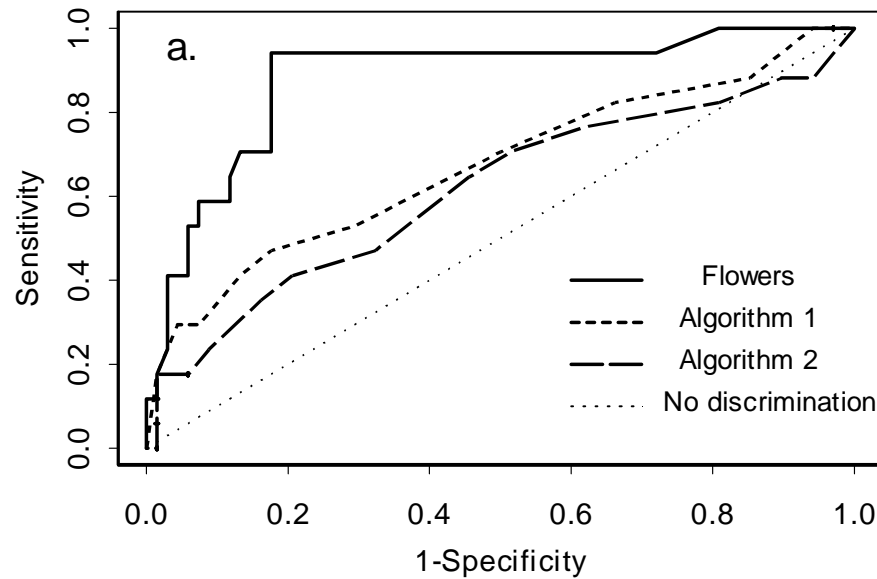
Crop Model 2



Barbottin, Makowski, LeBail,
Jeuffroy. 2008



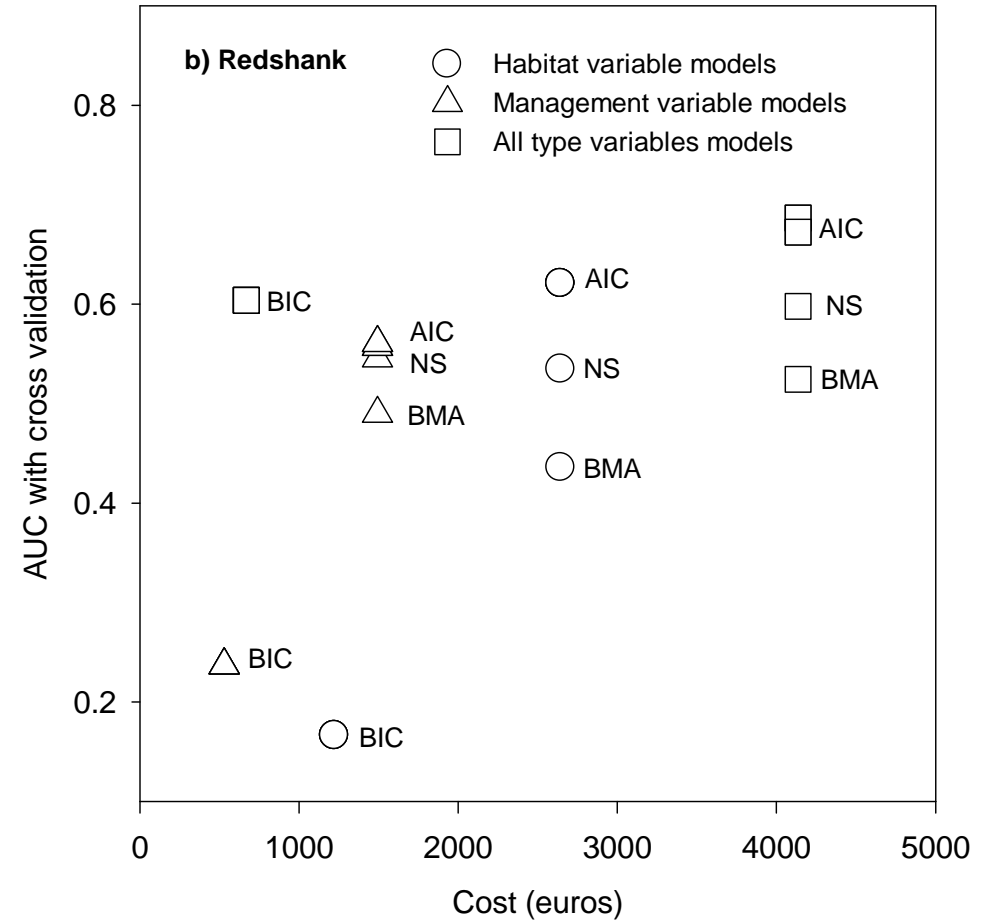
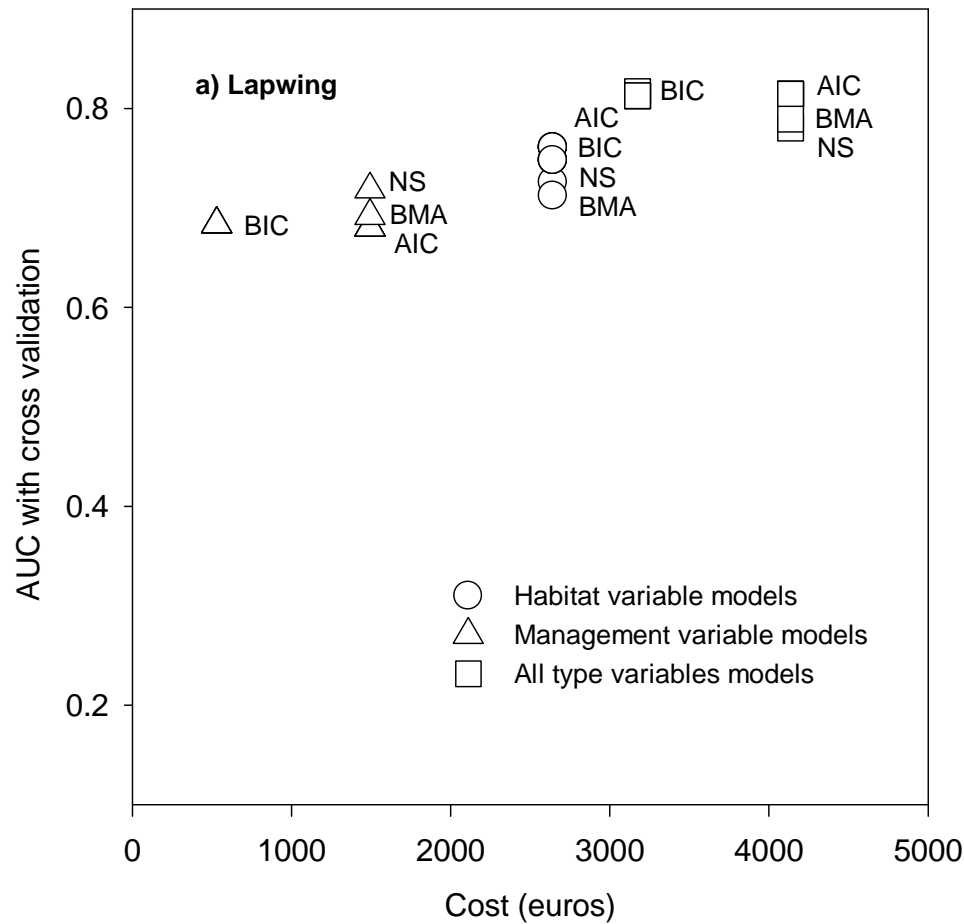
Prédiction du risque de perte de rendement associé au sclérotinia du colza



Makowski, D., M. Taverne, J. Bolomier, M. Ducarne. 2005



Prédiction de la présence/absence de deux espèces d'oiseaux dans les prairies du marais poitevin



Barbottin, Tichit, Cadet, Makowski. In press.



« Les prédictions de mon modèle sont imprécises et le modèle ne classe pas mieux que le hasard.

Mais c'est peut-être dû à l'utilisation de mesures peu fiables et pas au modèle lui-même.

On ne peut donc pas dire que le modèle est mauvais ».

Un agronome anonyme.



- **Possible en théorie**
- **Les bases de données utilisées dans les travaux présentés ici ont été choisies par les modélisateurs eux-mêmes**
- **Avec ces bases de données, tous les modèles n'ont pas le même niveau de précision**
- **S'il n'y a pas de mesures fiables disponibles pour évaluer le(s) modèle(s), il faut commencer par collecter de telles mesures !**



Quelques recommandations

- **Constituer des bases de données de référence pour évaluer les modèles**
- **Utiliser des procédures statistiques d'évaluation de modèles, comme l'analyse ROC ou autres**
- **Comparer plusieurs modèles de niveaux de complexité contrastés pour un usage donné**
- **Utiliser des méthodes d'analyse de sensibilité et d'incertitude**
- **Ne pas tout miser sur les modèles; utiliser des méthodes de synthèses de connaissance (e.g., meta-analyse de données et d'articles)**



Références

- Barbottin A, Makowski D, Le Bail M, Jeuffroy M-H, Bouchard Ch, Barrier C. 2008. Comparison of models and indicators for categorizing soft wheat fields according to their grain protein contents. *European Journal of Agronomy* 29, 159-183.
- Makowski D., Denis J-B., Ruck L., Penaud A. 2008. A Bayesian approach to assess the accuracy of a diagnostic test based on plant disease measurement. *Crop Protection* 27:1187-1193.
- Makowski, D., M. Taverne, J. Bolomier, M. Ducarne. 2005. Comparison of risk indicators for sclerotinia control in oilseed rape. *Crop Protection* 24:527-531
- Makowski D., Tichit M., Guichard L., van Keulen H., Beaudoin N. 2009. Measuring the accuracy of agro-environmental indicators. *Journal of Environmental Management* 90, S139-S146.
- Pepe MS. 1998. Three approaches to regression analysis of ROC curves for continuous test results. *Biometrics* 54, 124-135.
- Pepe MS. 2003. The statistical evaluation of medical tests for classification and prediction. Oxford Statistical Science series-28
- Primot, S., M. Valantin-Morison, D. Makowski. 2006. Predicting the risk of weed infestation in winter oilseed rape crops. *Weed Research* 46:22-33
- Swets, J.A., 1988. Measuring the accuracy of diagnostic systems. *Science* 240, 1285-1293.

