Novembre 2006

Exemples d'utilisation de modèles pour optimiser la lutte contre les bioagresseurs

David Makowski

UMR Agronomie INRA INA P-G
Thiverval-Grignon

Règle de décision

« Règle permettant de prendre une décision en fonction d'informations disponibles au moment de la décision ».

Règle de décision binaire

« **Règle** permettant de choisir parmi **deux décisions** ».

Exemples de décisions binaires

- **Traiter / Ne pas traiter contre une maladie.**
- **Traiter / Ne pas traiter contre des mauvaises herbes.**

Mais aussi

- **Semer la variété 1 / Semer la variété 2.**
- **Troisième apport d’engrais / Pas de troisième apport.**

Règle de décision binaire basée sur un indicateur et un seuil de décision

« Je traite si mon indicateur est supérieur/inférieur à un seuil de décision ».

« Je traite si *I* ≥ *S*. Je ne traite pas sinon ».

- Indicateur ***I*** = mesure ou prédiction d'un modèle (ex: % plantes malades).
- Seuil ***S*** = Valeur numérique (ex: 20%).

Optimisation de

« Je traite si *I* ≥ *S*. Je ne traite pas sinon ».

Deux problèmes pratiques:

• Choisir le meilleur seuil *S* pour un indicateur ***I*** donné.

• Choisir le meilleur indicateur parmi plusieurs indicateurs disponibles.

Étapes de l'optimisation

1. Définir une série d'indicateurs (mesures et/ou modèles).

2. Définir la gamme de variation possible pour le seuil S associé à chaque indicateur (*e.g* 0-100 % de fleurs malades).

3. Définir un ou plusieurs critères pour évaluer les règles de décision (*i.e* les combinaisons entre *I* et *S*).

4. Estimer les valeurs de ces critères pour chaque règle.

5. Choisir la « meilleure » règle.

Etape 1. Définir une série d’indicateurs (mesures et/ou modèles)

- <u>**Exemple 1:**</u> **Traitement du piétin échaudage du blé d’hiver.**

- <u>**Exemple 2:**</u> **Traitement du sclérotinia du colza.**

- <u>**Exemple 3:**</u> **Traitement des mauvaises herbes sur colza.**

Seize modèles simulant le % plantes attaquées par le piétin échaudage (Ennaïfar, 2006)

	Name	Type	Equation	Fsem30[z]	Selection
1	DynL0$^-$	**Dyn**amic	**L**inear	Absent	Bibliographic
2	DynL0$^+$	**Dyn**amic	**L**inear	Absent	Statistical + Bibliographic
3	DynL1$^-$	**Dyn**amic	**L**inear	Present	Bibliographic
4	DynL1$^+$	**Dyn**amic	**L**inear	Present	Statistical + Bibliographic
5	DynM0$^-$	**Dyn**amic	**M**ultiplicative	Absent	Bibliographic
6	DynM0$^+$	**Dyn**amic	**M**ultiplicative	Absent	Statistical + Bibliographic
7	DynM1$^-$	**Dyn**amic	**M**ultiplicative	Present	Bibliographic
8	DynM1$^+$	**Dyn**amic	**M**ultiplicative	Present	Statistical + Bibliographic
9	StaL0$^-$	**Sta**tic	**L**inear	Absent	Bibliographic
10	StaL0$^+$	**Sta**tic	**L**inear	Absent	Statistical + Bibliographic
11	StaL1$^-$	**Sta**tic	**L**inear	Present	Bibliographic
12	StaL1$^+$	Static	**L**inear	Present	Statistical + Bibliographic
13	StaM0$^-$	**Sta**tic	**M**ultiplicative	Absent	Bibliographic
14	StaM0$^+$	**Sta**tic	**M**ultiplicative	Absent	Statistical + Bibliographic
15	StaM1$^-$	**Sta**tic	**M**ultiplicative	Present	Bibliographic
16	StaM1$^+$	**Sta**tic	**M**ultiplicative	Present	Statistical + Bibliographic

Un des 16 modèles simulant le % plantes attaquées par le piétin échaudage (Ennaïfar, 2006)

Techniques culturales (succession, date de semis, texture du sol…)

Climat (température, pluviométrie)

$$C_1 = \mu^{(1)} + \alpha_1^{(1)} \cdot x_1 + \ldots + \alpha_n^{(1)} \cdot x_n$$

$$C_2 = \mu^{(2)} + \alpha_1^{(2)} \cdot x_1 + \ldots + \alpha_n^{(2)} \cdot x_n$$

$$y = \frac{1 - \exp^{-(C1+C2)\cdot t}}{1 + \frac{C_2}{C_1} \cdot \exp^{-(C1+C2)\cdot t}}$$

Trois indicateurs pour raisonner le traitement du sclérotinia du colza

- **Pourcentage de fleurs malades parmi *N* fleurs prélevées dans une parcelle.**

- **Niveau de risque prédit par un modèle qui ne tient pas compte du climat.**

- **Niveau de risque prédit par un modèle qui tient compte du climat.**

Niveau de risque de sclérotinia prédit par un modèle qui ne tient pas compte du climat

Risk factor	Level	Points
Number of oil-seed crops during the last ten years	>5	30
	3-5	20
	2-3	10
	1	0
Other host crops during the last five years	Yes	15
	No	0
Level of infection in the last crop	High	15
	Moderate	5
	Low	0
Type of field	Wet	10
	Dry	0
Plant density	High	10
	Normal	5
	Low	0

Vingt modèles simulant le risque d'infestation du colza par les mauvaises herbes
(Primot *et al.*, 2006)

Model	Equation	Variables	Barralis grid	Threshold of weed biomass at winter (t ha^{-1})
LIN.weed	Linear	Weed	No	-
LIN.tech	Linear	Techniques	No	-
LIN.all	Linear	All	No	-
LIN.selec	Linear	Selected	No	-
LIN.selec.b	Lineal	Selected	Yes	-
LOG.weed.10	Logistic	Weed	No	0.10
LOG.tech.10	Logistic	Techniques	No	0.10
LOG.all.10	Logistic	All	No	0.10
LOG.selec.10	Logistic	Selected	No	0.10
LOG.selec.10.b	Logistic	Selected	Yes	0.10
LOG.weed.15	Logistic	Weed	No	0.15
LOG.tech.15	Logistic	Techniques	No	0.15
LOG.all.15	Logistic	All	No	0.15
LOG.selec.15	Logistic	Selected	No	0.15
LOG.selec.15.b	Logistic	Selected	Yes	0.15
LOG.weed.20	Logistic	Weed	No	0.20
LOG.tech.20	Logistic	Techniques	No	0.20
LOG.all.20	Logistic	All	No	0.20
LOG.selec.20	Logistic	Selected	No	0.20
LOG.selec.20.b	Logistic	Selected	Yes	0.20

Modèle logistique pour prédire le risque d'infestation par des mauvaises herbes

$$P\left(y > D_{thresh}\right) = \frac{\exp\left(\alpha_0 + \sum_{i=1}^{p} \alpha_i x_i\right)}{1 + \exp\left(\alpha_0 + \sum_{i=1}^{p} \alpha_i x_i\right)}$$

y: biomasse de mauvaise herbe en hiver.

D_{thresh}: Seuil de nuisibilité.

$x1$, …, xp: variables explicatives.

$\alpha 0$, …, αp: paramètres à estimer.

Modèle linéaire pour prédire le risque d'infestation par des mauvaises herbes

$$y = \alpha_0 + \sum_{i=1}^{p} \alpha_i x_i + \varepsilon$$

y: biomasse de mauvaise herbe en hiver.

x1, …, *xp*: variables explicatives.

$\alpha 0$, …, αp: paramètres à estimer.

Variables d'entrée

Type of the variable	Variables	Minimum	Maximum	Unit
Techniques	Sowing Date (SD)	27 July (207)	7 September (249)	Days since 1st January
Techniques	Soil mineral Nitrogen at emergence (N)	26	317	kg ha^{-1}
Techniques	Oilseed rape Density at emergence (OD)	14	232	Plants.m^{-2}
Techniques	Soil Tillage (ST)	0 (deep tillage)	1 (superficial tillage)	
Weeds	Weed Density at emergence (WD)	1	532	Plants.m^{-2}
Weeds	Type of the most abundant weed at emergence	Type 1	Type 7	See weed typology in Table 2

Variables d'entrée (suite)

Type	Group	Time of emergence of the weed	Height of weed relative to rapeseed
1	Monocotyledonous	Summer	Equal or lower
2	Monocotyledonous	Indifferent	Equal or lower
3	Dicotyledonous	Summer	Equal or lower
4	Dicotyledonous	Summer	Higher
5	Dicotyledonous	Autumn	Equal or lower
6	Dicotyledonous	Indifferent	Equal or lower
7	Dicotyledonous	Indifferent	Higher

Étapes de l'optimisation

1. Définir une série d'indicateurs (mesures et/ou modèles).

2. Définir la gamme de variation possible pour le seuil *S* associé à chaque indicateur (0, 0.1, …, 0.9, 1).

3. Définir un ou plusieurs critères pour évaluer les règles de décision (*i.e* les combinaisons entre *I* et *S*).

4. Estimer les valeurs de ces critères pour chaque règle.

5. Choisir la « meilleure » règle.

Étape 3. Définir un ou plusieurs critères pour évaluer les règles de décision (*i.e* les combinaisons entre *I* et *S*).

- **Évaluer la qualité prédictive des modèles (MSEP).**

- **Évaluer le résultat de l'application de la règle de décision.**

  - **Marge brute moyenne obtenue en appliquant la règle**.

  - **Analyse ROC (sensibilité, spécificité)**.

Étape 4. Estimer les valeurs des critères pour chaque règle.

Analyse ROC

Notations

Y : une variable aléatoire prenant les valeurs 0 ou 1 pour une réponse négative ou positive respectivement.

I : une variable continue correspondant à un indicateur.

S : un seuil de décision.

Exemples

$Y = 0$ si perte de rendement due à la maladie = 0, $Y=1$ sinon.

$Y = 0$ si pourcentage de plante malade à la récolte < 10%, $Y=1$ sinon.

$Y = 0$ si biomasse de mauvaise herbe < 0.15 t/ha, $Y=1$ sinon.

Analyse ROC

n parcelles avec Y=0 (*e.g*. biomasse mauv. herbes < 0.15 t.ha^{-1}).

m parcelles avec Y=1 (*e.g.* biomasse mauv. herbes > 0.15 t.ha^{-1}).

(i). Calcul de l'indicateur I pour chaque parcelle.

(ii). Définition d'un seuil de décision, I_T .

(iii). **Sensibilité** = $Prob(I \geq I_T \mid Y=1)$

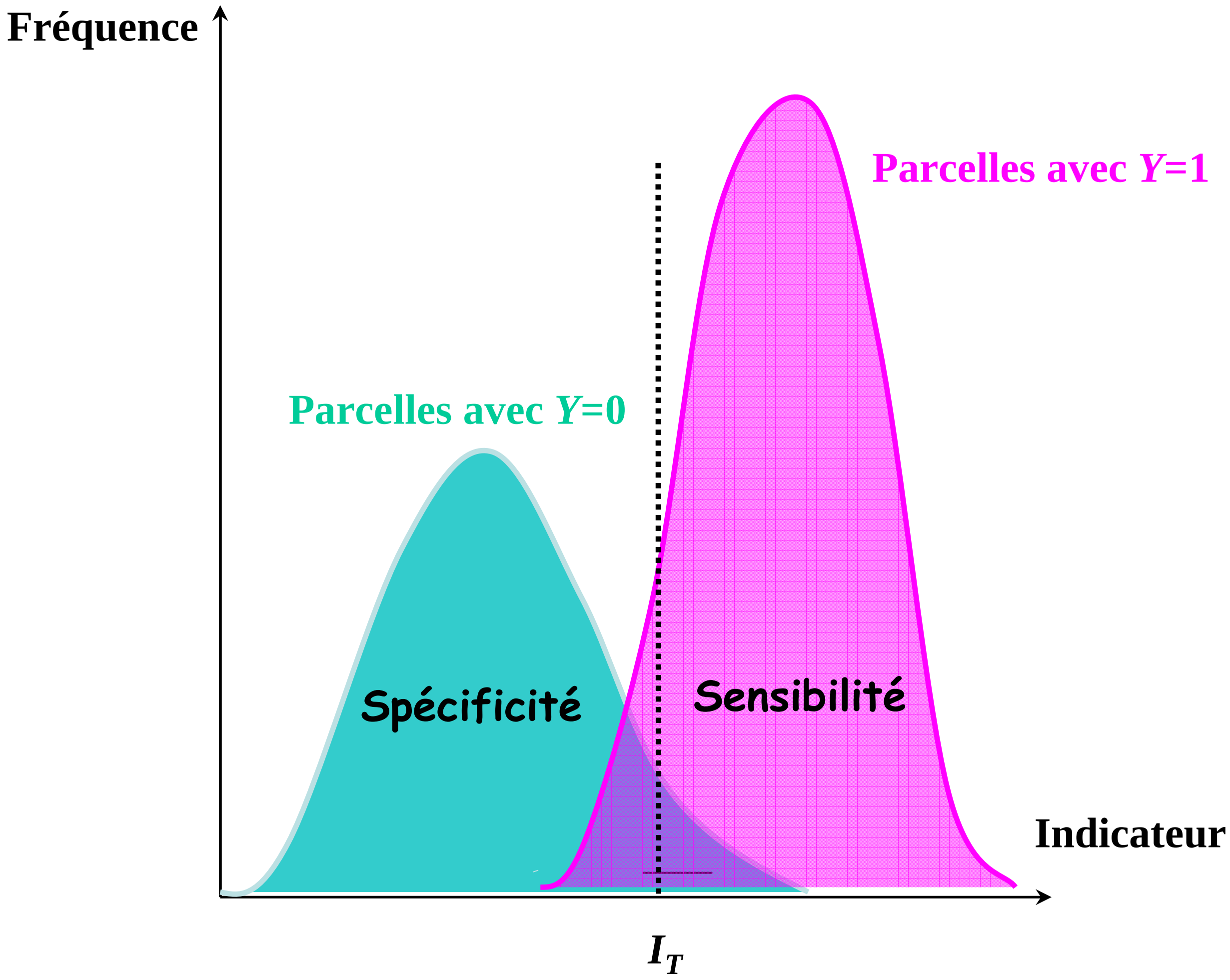
(iv). **Spécificité** = $Prob(I < I_T \mid Y=0)$

(v). **Courbe ROC**: Sensibilité (I_T) *versus* 1 – Spécificité (I_T)

(vi). Estimation la surface sous la courbe (**AUC**) pour chaque indicateur I.

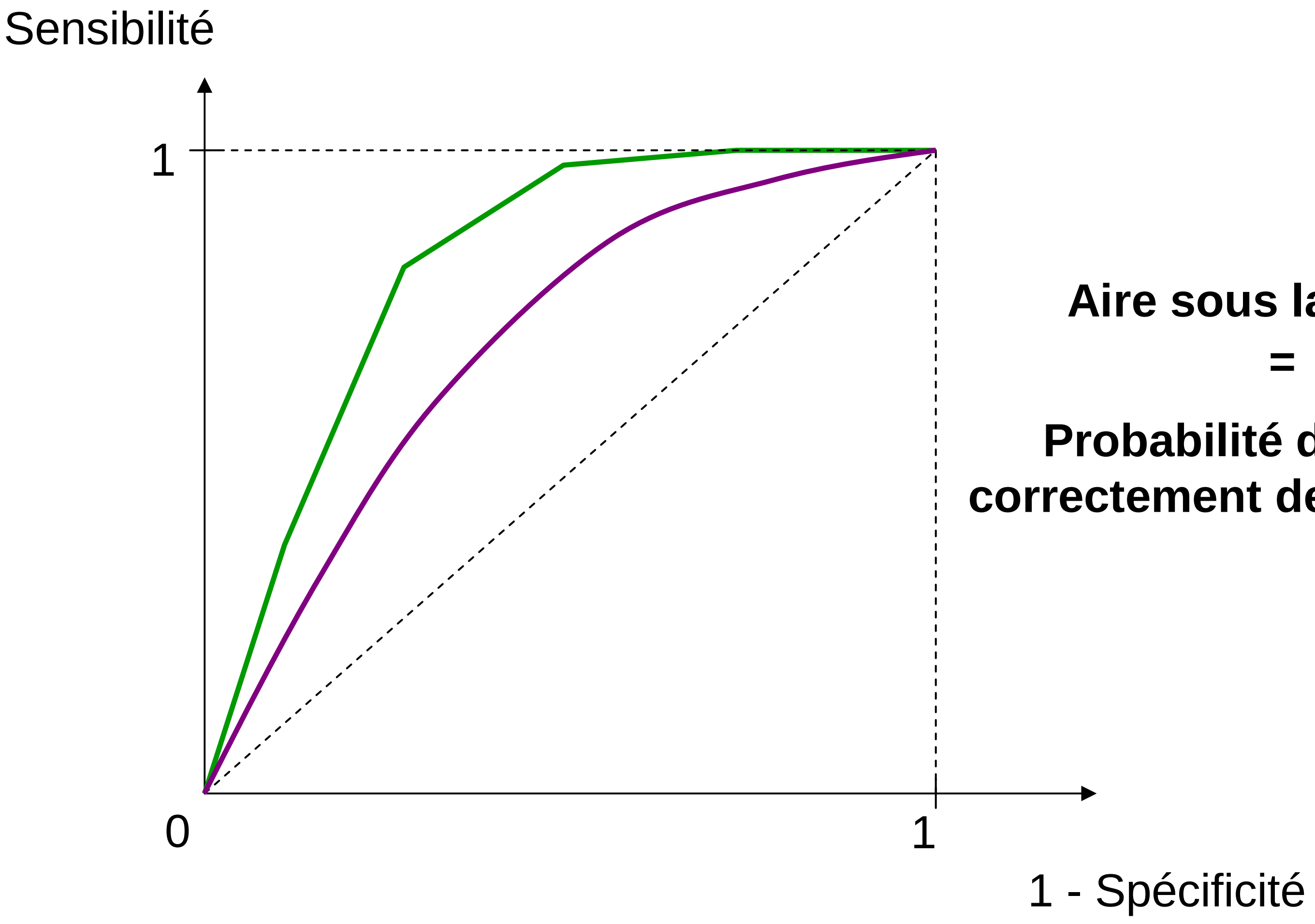
Si AUC~0.5, l'indicateur n'est pas mieux qu'une décision aléatoire.

Analyse ROC



Fréquence
Parcelles avec Y=1
Parcelles avec Y=0
Spécificité
Sensibilité
Indicateur
I_T

Analyse ROC



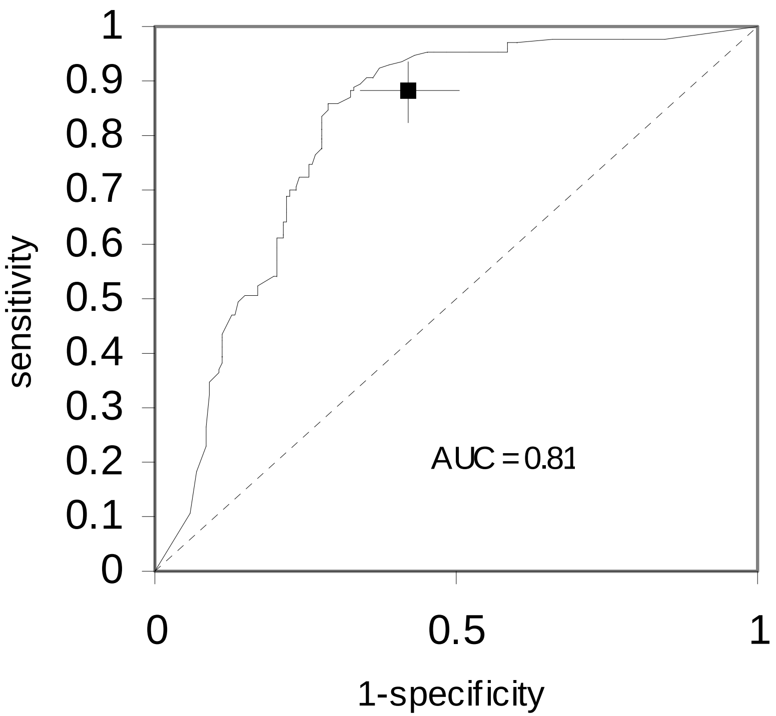
Aire sous la courbe

=

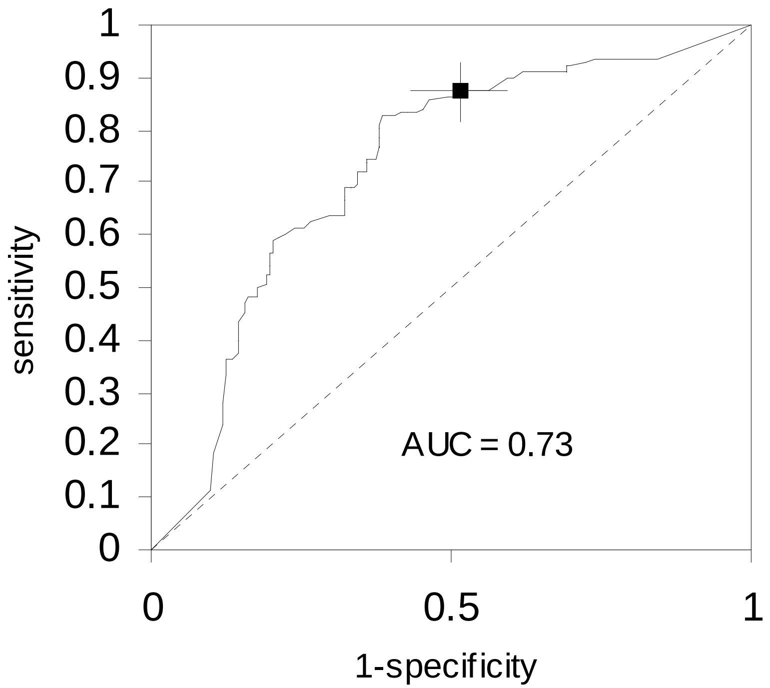
Probabilité de classer correctement deux parcelles

Modèles simulant le risque d'infestation du colza par les mauvaises herbes (Primot *et al*., 2006)

a. **Variables sélectionnées par stepwise**



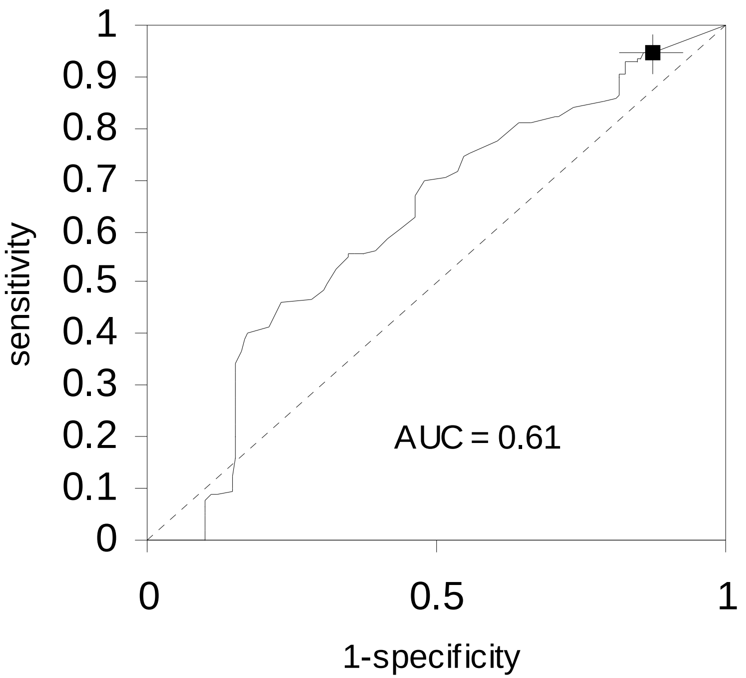
b. **Toutes les variables**



Modèles simulant le risque d'infestation du colza par les mauvaises herbes (Primot *et al.*, 2006)

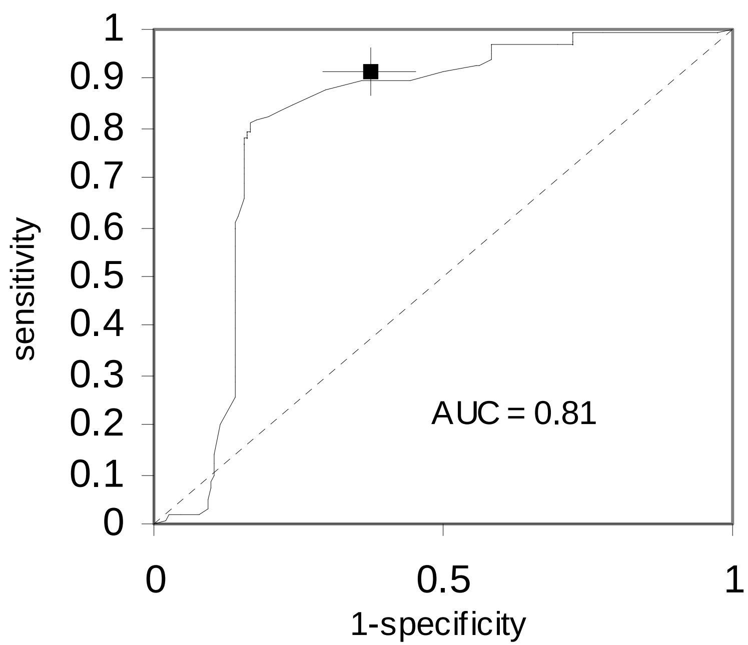
Variables « Techniques »

c.

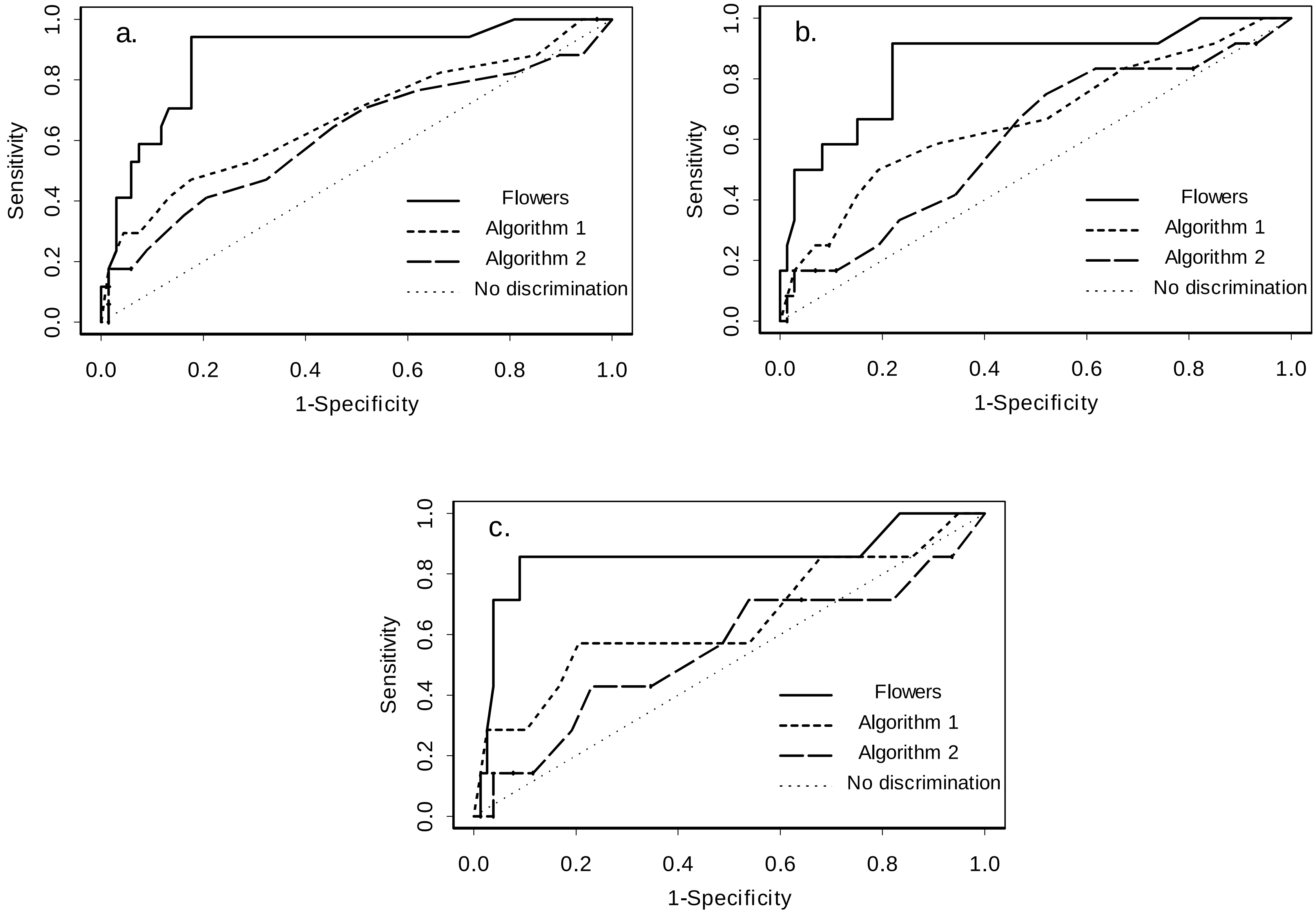


Variables « Plantes »

d.



Indicateurs de risque de sclérotinia (Makowski *et al*., 2005)



Étape 5. Choix de la meilleure règle

- Avec une analyse ROC, on choisit un compromis entre une forte sensibilité et une forte spécificité.

- Avec la marge moyenne ou le MSEP, on choisit la règle conduisant à un niveau satisfaisant du critère.

- Dans tous les cas, on doit tenir compte du coût d’utilisation.

Bonne solution: fournir un maximum d’informations à l’utilisateur et le laisser choisir.