Plan

STRATÉGIE DE CHOIX DE MODÈLE

- non orthogonalité
- instabilité des paramètres

DEUX EXEMPLES :

- Armillaire
- folle avoine

Exemple armillaire, objectif

Influence du pH du sol
sur
l'abondance de l'Armillaire

Champignon décomposeur de bois
et
parasite opportuniste

$\Rightarrow$ colonisation du bois mort par l'Armillaire

Étude de la colonisation du bois mort par l'Armillaire (1996)

19 bois morts examinés

échantillonnés les plus proches de 19 points déjà matérialisés sur le terrain.

Variable binaire :
→ colonisé par l'Armillaire
→ non colonisé par l'Armillaire

1				13
	2	8	14	
	3	9	15	
	4	10	16	
	5	11	17	
	6	12	18	
7				19

Échantillonnage sur une parcelle

Exemple Armillaire, dispositif

1985 : dispositif

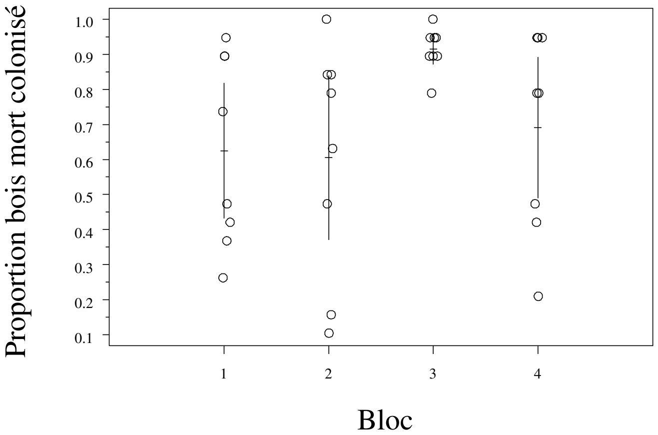
4 blocs (Bl)
X
3 facteurs :

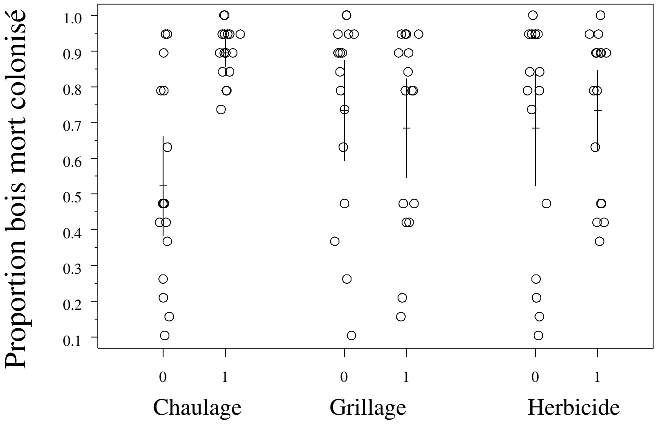
- Chaulage (Ch)
- Engrillagement (Gr)
- Herbicide (He)

+ leurs combinaisons (8 par bloc)
⇒ **32 parcelles en tout**

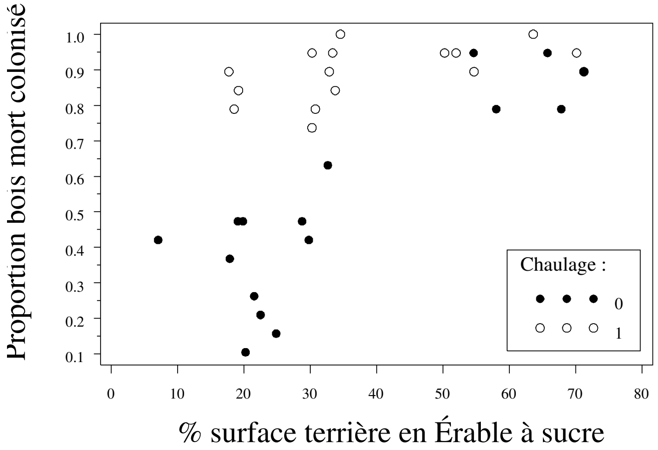
Coupe de régénération :
50 % de la surface terrière enlevée.

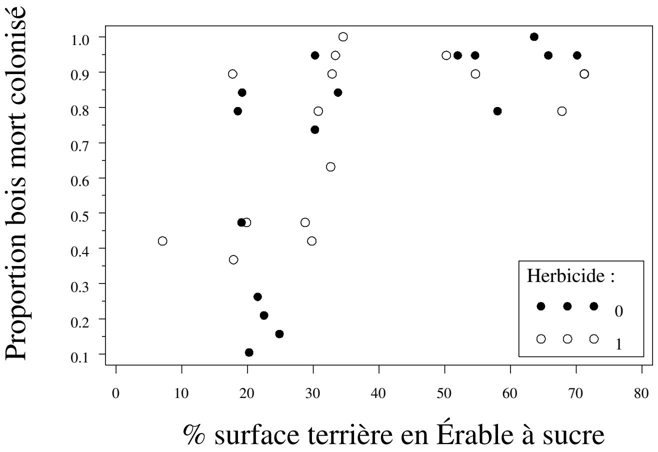
Impact des facteurs sur la colonisation du bois mort





Impact des facteurs sur la colonisation du bois mort





Type et quantité de bois mort présent

Forte influence du type de bois mort

	Origine du bois mort	
	Erable à sucre	Autres essences
Colonisation par l'Armillaire	83 %	55 %

Prise en compte?

- Proportion d'érable à sucre dans le peuplement (PropES).
- Surface terrière d'érable à sucre enlevée en 85 (EScoupe).
- Surface terrière totale enlevée en 85 (Totcoupe).

Type et quantité de bois mort présent

Choix du régresseur :

	Déviance résiduelle	Chi2	DDl	Pr>Chi
Modèle nul	229.95			
Totcoupe	216.34	13.61	1	0.0002
Modèle nul	229.95			
EScoupe	174.39	55.56	1	0.0000
Modèle nul	229.95			
PropES	138.25	**91.70**	1	0.0000

Modèle complet

Class Level Information

Class	Levels	Values
GR	2	0 1
CH	2	0 1
HE	2	0 1
BL	4	1 2 3 4

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	12	16.6245	1.3854
Pearson Chi-Square	12	17.3018	1.4418
Log Likelihood	.	-260.0528	

LR Statistics For Type 3 Analysis

Source	DF	ChiSquare	Pr>Chi
PROPES	1	4.5236	**0.0334**
BL	3	1.1781	0.7583
CH	1	59.4951	**0.0000**
CH*BL	3	5.1135	0.1637
HE	1	0.2064	0.6496
HE*BL	3	13.5797	**0.0035**
GR	1	0.0748	0.7844
GR*BL	3	3.0955	0.3771
GR*HE	1	0.9237	0.3365
CH*HE	1	0.2634	0.6078
GR*CH	1	0.4574	0.4988

Corrélation entre paramètres

Modèle avec PropES et bloc uniquement :

Estimated Correlation Matrix

	PropES	Bl1	Bl2	Bl3
PropES	1.000			
Bl1	-0.099	1.000		
Bl2	-0.044	0.561	1.000	
Bl3	**-0.756**	0.349	0.240	1.000

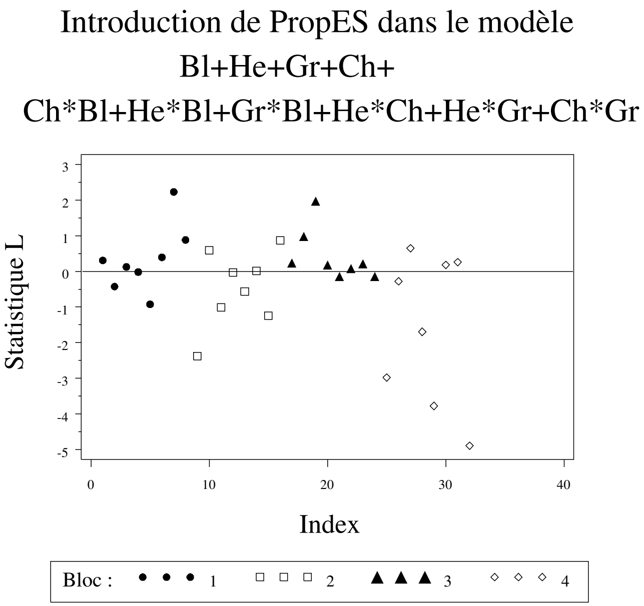
$\Rightarrow$ introduction séquentielle des 2 variables :

LR Statistics For Type 1 Analysis

Source	Déviance	DF	ChiSquare	Pr>Chi
Intercept	229.9			
PropES	138.2	1	91.7	**0.0001**
Bl	135.2	3	3.0	**0.39**
Intercept	229.9			
Bl	178.3	3	51.6	**0.0001**
PropEs	135.2	1	43.1	**0.001**

Statistique L

Contribution des observations à la déviance associée à PropES



Second modèle proposé

Class Level Information

Class	Levels	Values
CH	2	0 1
HE	2	0 1

Criteria For Assessing Goodness Of Fit

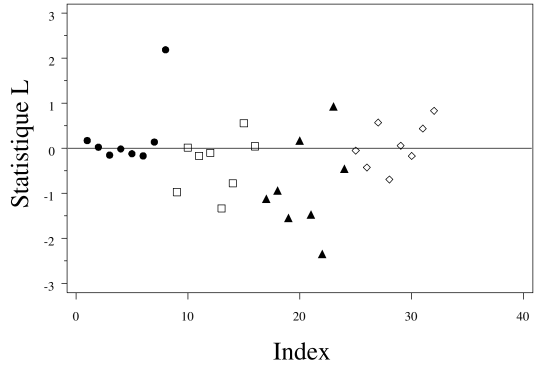
Criterion	DF	Value	Value/DF
Deviance	26	30.1926	1.1613
Pearson Chi-Square	26	29.0665	1.1179
Log Likelihood	.	-266.8369	

LR Statistics For Type 3 Analysis

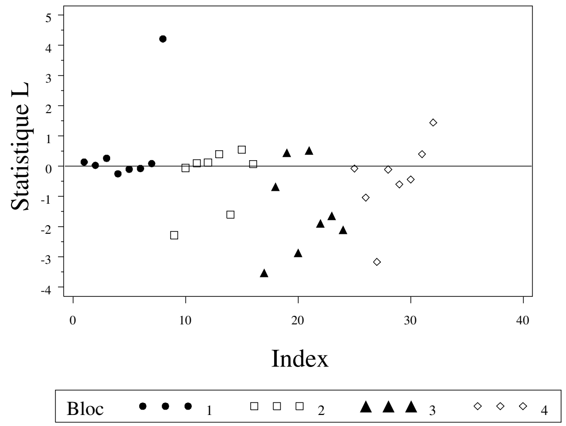
Source	DF	ChiSquare	Pr>Chi
PROPES	1	37.5531	0.0000
HE	1	15.9413	0.0001
HE*PROPES	1	12.1970	0.0005
CH	1	35.8300	0.0000
CH*PROPES	1	6.2152	0.0127

Validation du modèle

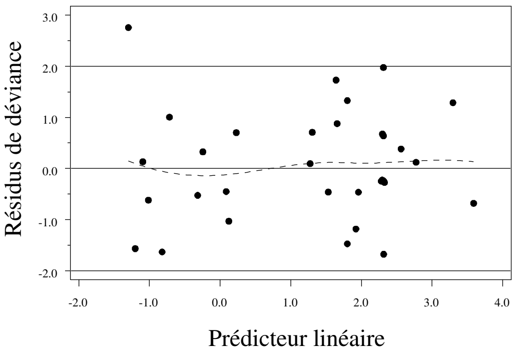
Introduction de PropES*Ch dans le modèle comprenant
PropES, Ch, He, et PropES*He

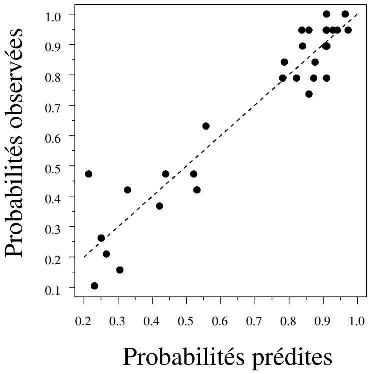


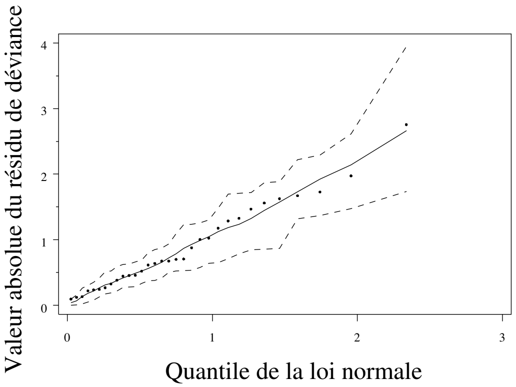
Introduction de Propes*He dans le modèle comprenant
PropES, Ch, He et PropES*Ch



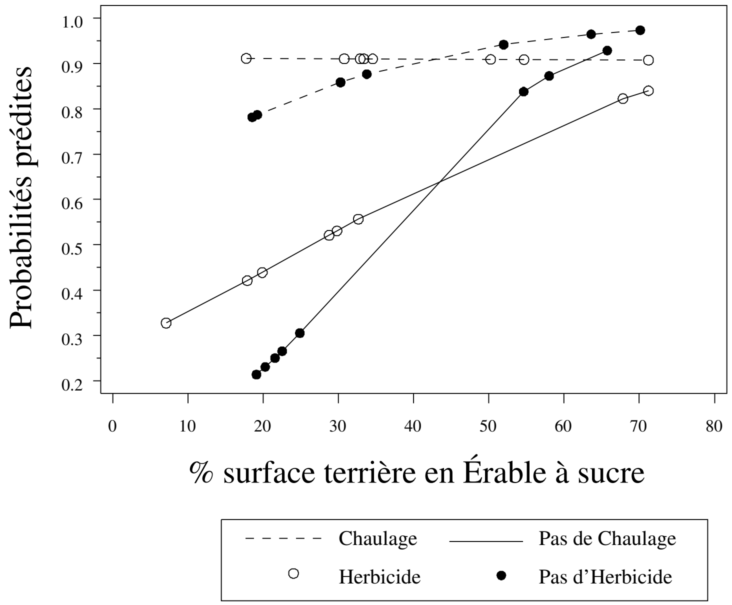
Validation du modèle







Résultats



Probabilités prédites
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0
10
20
30
40
50
60
70
80
% surface terrière en Érable à sucre
Chaulage
Pas de Chaulage
Herbicide
Pas d'Herbicide

Surparamétrisation et instabilité des paramètres

facteur A	Proportion
1	$\boxed{0/4}$
2	$5/6$

$$\begin{aligned} logit(p) &= \beta_0, & \text{si A=1} \\ &= \beta_0 + \alpha_2, & \text{si A=2} \end{aligned}$$

	estimation	écart-type
β_0	$\boxed{-12.37}$	$\boxed{146.9}$
α_2	$\boxed{13.97}$	$\boxed{146.9}$

Pourquoi?

$$\left.\begin{array}{c} \text{jeu de données particulières} \\ \text{surparamétrisation} \end{array}\right\} \Rightarrow \text{instabilité}$$

Surparamétrisation

Quoi ?

- trop grand nombre de paramètres ...

Quand ?

- données creuses
- modèle trop complexe

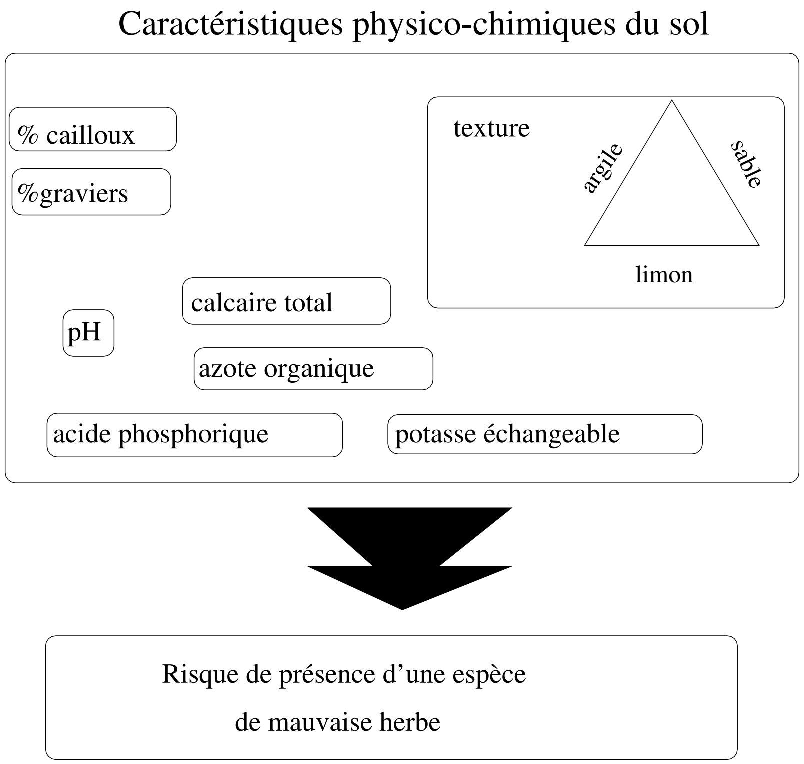
Diagnostic

- message de non convergence (éventuellement).
- $\begin{cases} \text{certains} \quad \widehat{\text{paramètres}} \quad \text{très grands.} \\ \text{certains} \quad \widehat{\text{écarts-type}} \quad \text{très grands.} \end{cases}$

Conséquences

- paramètres instables
- estimations peu précises

Exemple malherbologie, objectif



Exemple malherbologie, protocole

Côte d'Or 1969 $\longrightarrow$ 1976
245 parcelles

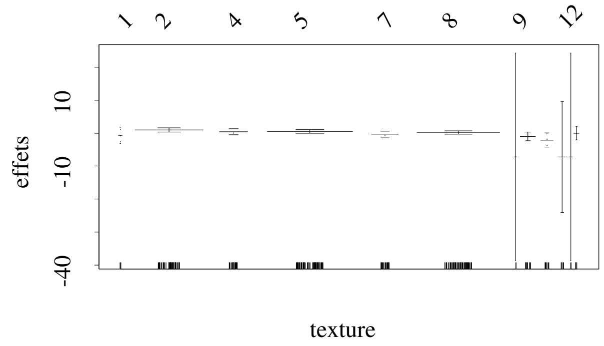
218 espèces inventoriées

parcelle	$\rightarrow$ 3 années	culture automne + culture printemps $\downarrow$ relevés botaniques notation abondance 0 à 5

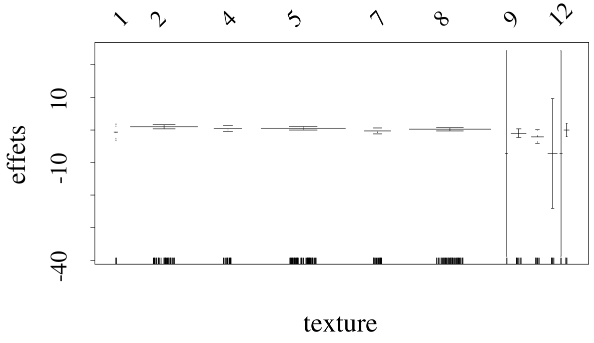
0/1 :
0 : absence de l'espèce
1 : présence de l'espèce pour un relevé au moins

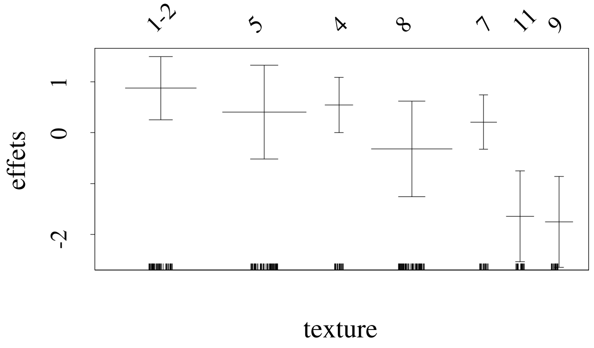
Modèle : texture

```
Coefficients:
                    Value Std. Error
 texture1 -6.931472e-01  1.2247449
 texture2  9.903987e-01  0.3247992
 texture4  4.054651e-01  0.4564355
 texture5  5.465437e-01  0.2678999
 texture7 -3.184537e-01  0.4646602
 texture8  2.076394e-01  0.2640294
 texture9 -7.201716e+00 15.7257175
texture10 -9.808293e-01  0.6770032
texture11 -2.079442e+00  1.0606602
texture12 -7.201716e+00  8.4057496
texture14 -7.201716e+00 15.7257175
texture15  1.110223e-16  1.0000000
```



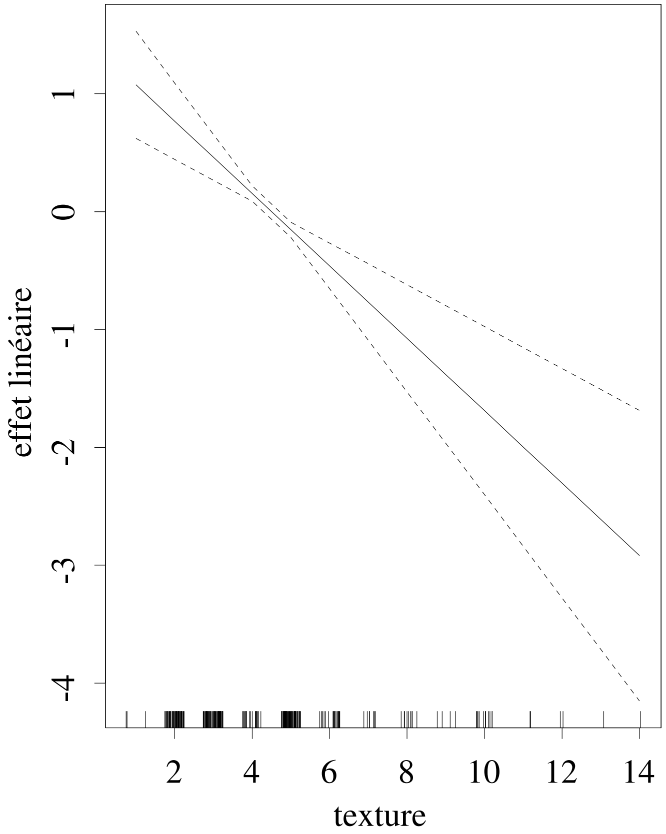
Modèle : texture





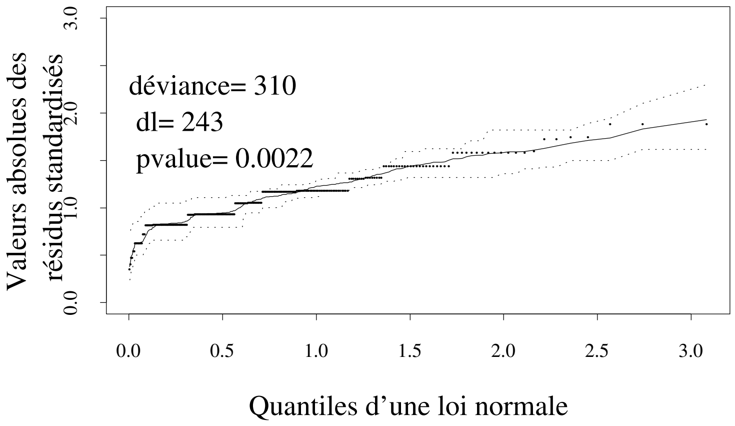
effectifs des textures regroupées						
1-2	5	4	8	7	11-12 15	9-10-13 14-17
51	60	20	58	19	20	17

Modèle : effet linéaire de la texture

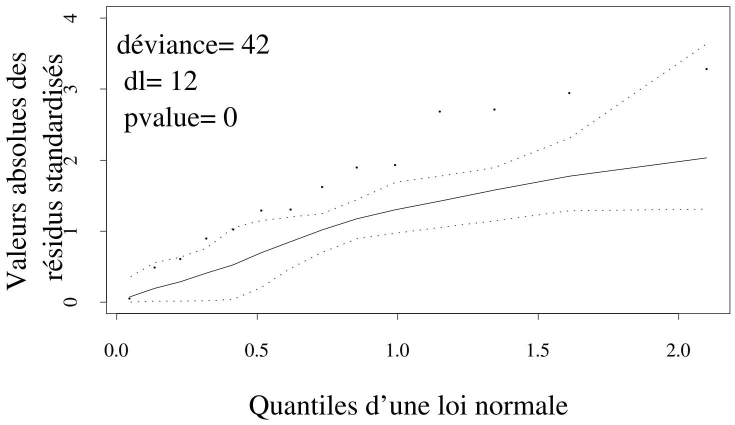


Modèle : effet linéaire de la texture

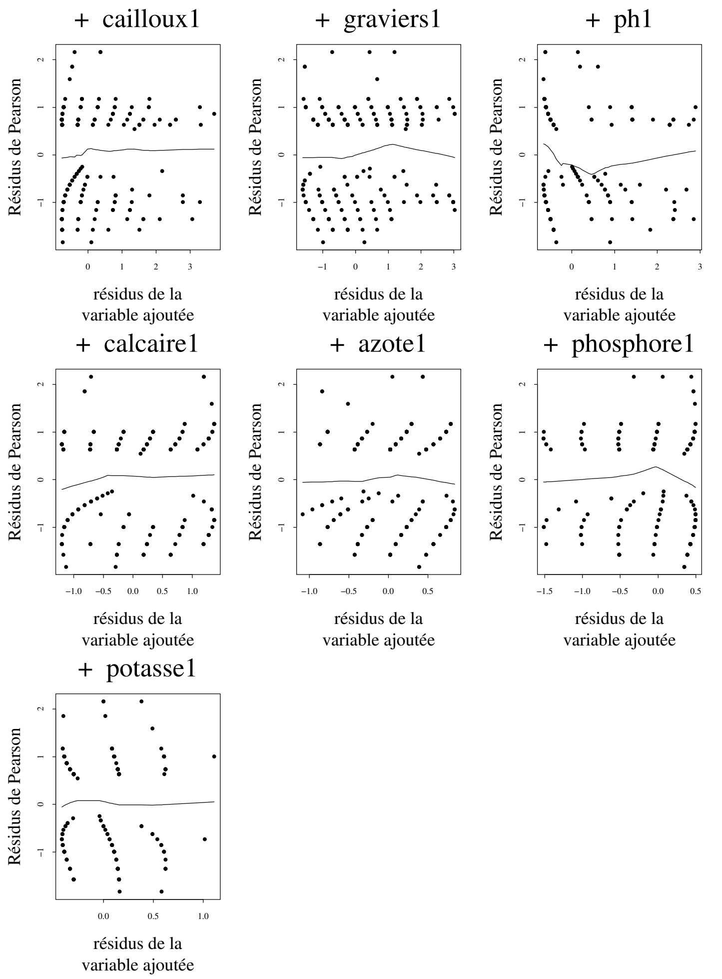
données non groupées



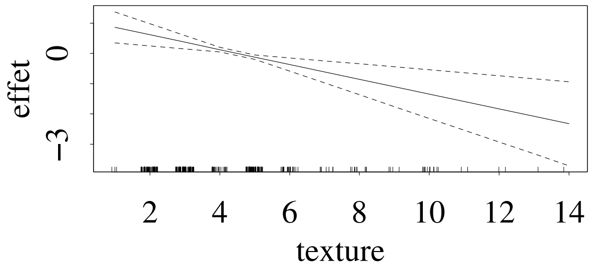
données groupées

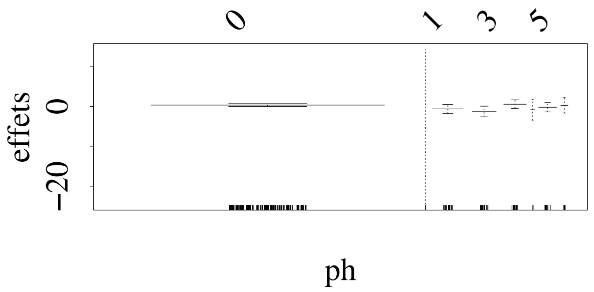


Graphiques des variables ajoutées



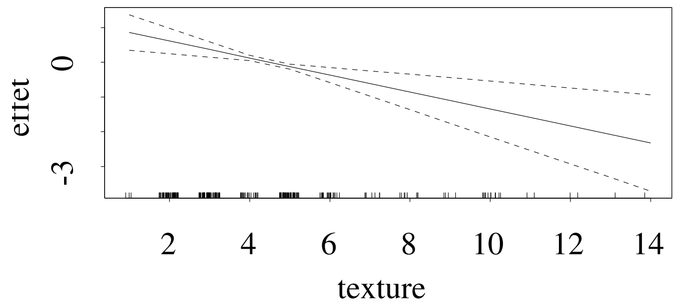
Modèle : texture.lin + ph

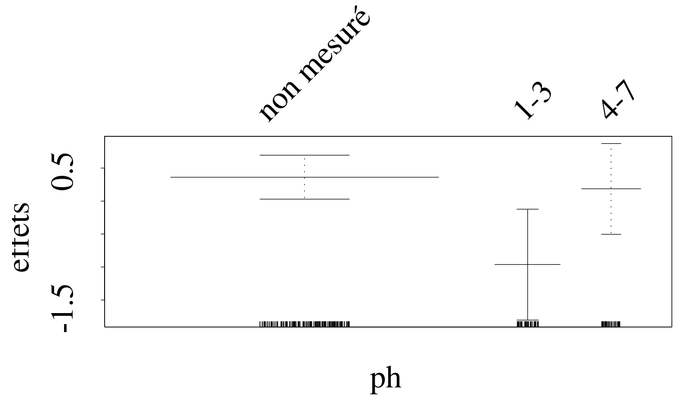




	pH							
	0	1	2	3	4	5	6	7
n	167	2	22	17	16	3	13	5

Modèle texture.lin + ph



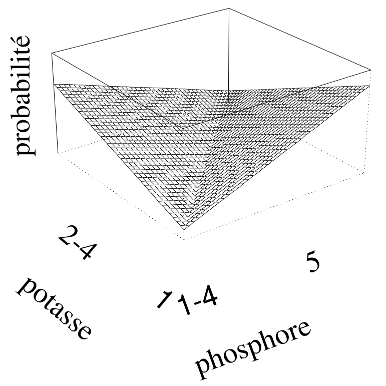


Recherche d'interactions

modèle additif : toutes les var. expl.

+ interactions d'ordre 2

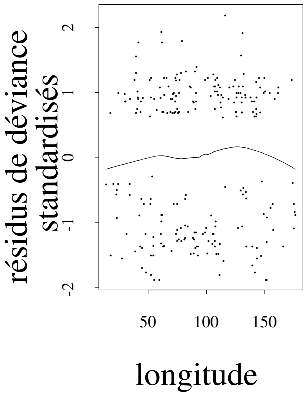
→ phosphore : potasse (p=1 %)

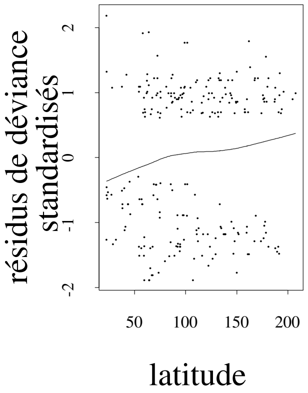


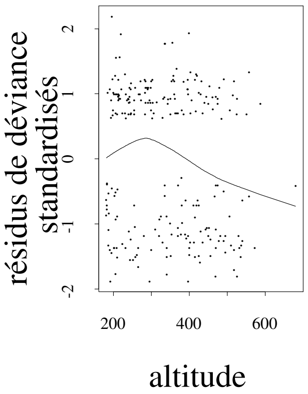
$\left\{ \begin{array}{c} \text{à valider avec de nouvelles données} \\ \text{difficile à mesurer} \end{array} \right.$

→ retour au modèle additif.

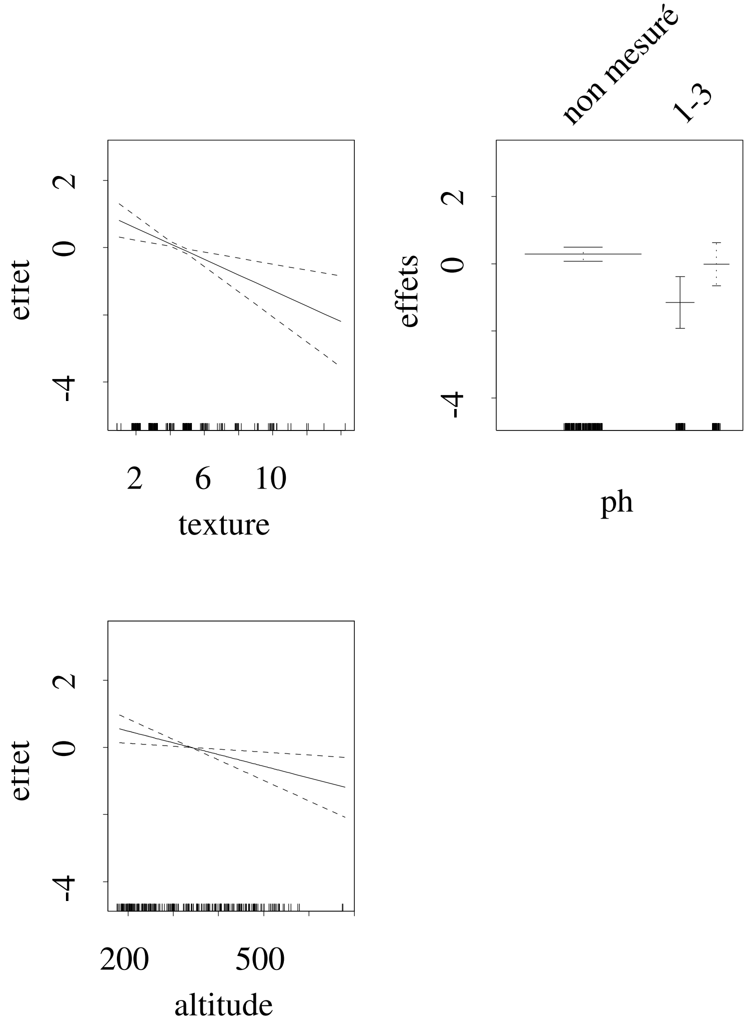
Structure des résidus?



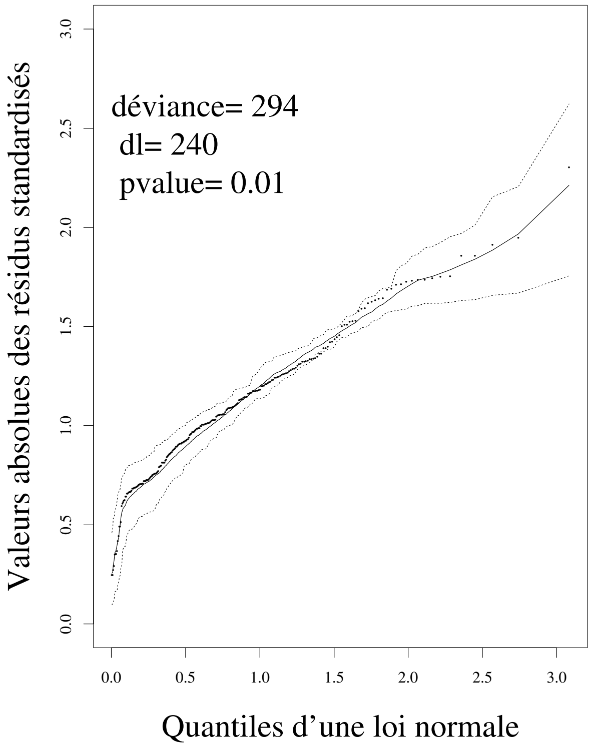




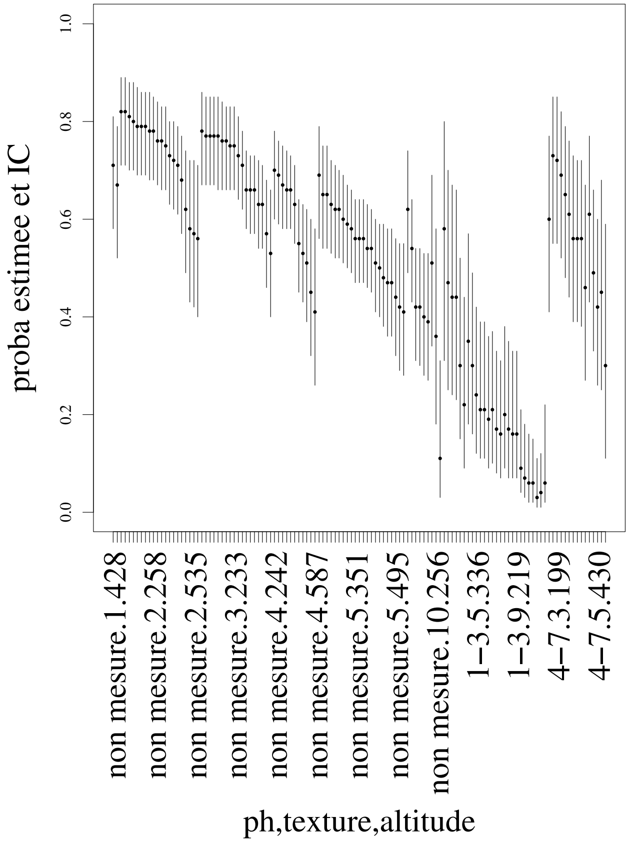
Modèle final : texture, ph, altitude



Modèle final : texture, ph, altitude



Probabilités estimées et IC



Stratégie de sélection de modèle

stratégie idéale : descendante

Données

$\downarrow$

Modèle le plus complet possible

$\downarrow$

Validation

$\downarrow$

Simplification du modèle

$\rightarrow$ interprétation + facile
$\rightarrow$ + grande précision

Sélection de modèle

Comparaison de modèles emboîtés
Tests du rapport de vraisemblance

sélection

$\swarrow \qquad \searrow$

descendante ascendante

mixte

utiliser la connaissance
Réduire le nombre de tests

Comparer des modèles non emboîtés

Minimiser le critère d'Akaïke :
$AIC = \text{Dev} + 2.\ \text{nb param}$

Sélection de modèle

Ne pas interpréter indépendamment les uns des autres des paramètres fortement corrélés

→ polynômes orthogonaux
→ choix de var. expl. peu corrélées
→ régression sur composantes principales

Attention aux relations causales :

$$\begin{tikzcd} \mathbf{X} & & \mathbf{Y} \\ \mathbf{X'} \arrow[u] \arrow[urr] & & \end{tikzcd}$$

X : var. expl. d'intérêt
X' : var. expl.

RÉSUMÉ

difficultés	comment diagnostiquer ?
non orthogonalité	matrice de corrélation des param
instabilité	écart-type des paramètres

Choix de modèle :

par ordre décroissant de préférence,

stratégie descendante
stratégie mixte
stratégie ascendante