

ÉTUDE STATISTIQUE DES DONNÉES COLLECTÉES PAR L'OBSERVATOIRE NATIONAL DES MALADIES DU BOIS DE LA VIGNE

F. BERTRAND *, M. MAUMY *, L. FUSSLER **,
N. KOBES **, S. SAVARY ***, J. GROSMAN ****

RÉSUMÉ

L'objectif de l'Observatoire National des Maladies du Bois de la Vigne est de dresser un état des lieux de la répartition, de la fréquence et de l'intensité de l'expression des symptômes foliaires des maladies du bois, pour répondre à la question de leur progression dans le vignoble français. En effet, suite à une interdiction de l'utilisation de l'arsénite de soude, les viticulteurs ne disposent plus d'aucune méthode de lutte chimique curative homologuée contre les maladies du bois de la vigne. Cet observatoire collecte, chaque année, depuis 2003, un ensemble de données cohérentes.

Le jeu de données est complexe : il comporte des variables quantitatives et qualitatives qui évoluent au cours du temps. La problématique de l'étude est de dégager les grandes tendances en matière d'épidémiologie végétale afin de déterminer quelles sont les mesures prophylactiques à mettre en œuvre collectivement et à grande échelle.

Le travail statistique s'est articulé en trois points. Nous avons mis en évidence des relations entre les différentes variables de l'étude, puis nous avons utilisé l'analyse des correspondances multiples, l'analyse en composantes principales et l'analyse factorielle de données mixtes. Puis, afin de préciser les relations décelées, notre choix s'est porté sur des modèles de régressions logistiques. Nous avons utilisé des techniques bootstrap pour construire des intervalles de confiance autour de leurs paramètres. Enfin, pour tenir compte du facteur temps, nous avons employé des méthodes d'analyse factorielle de tableaux multiples.

Mots-clés : Analyse des correspondances multiples, Analyse en composantes principales, Analyse factorielle de données mixtes, Régression logistique binaire, Régression logistique ordinaire, Bootstrap, Analyse factorielle de tableaux multiples.

*. Laboratoire de Statistique, IRMA ULP (UMR 7501), 67084 Strasbourg Cedex
Email : fbertran@math.u-strasbg.fr, mmaumy@math.u-strasbg.fr

** . DRAF-SRPV, Alsace Cité administrative Gaujot, 67084 Strasbourg Cedex
Email : lfussler@wanadoo.fr, nathalie.kobes@agriculture.gouv.fr

*** . UMR Santé Végétale, Centre INRA de Bordeaux, 33883 Villenave d'Ornon Cedex
Email : ssavary@bordeaux.inra.fr

**** . DRAF-SRPV Rhône, Cité administrative de la Part-Dieu, 69401 Lyon Cedex 03
Email : jacques.grosman@agriculture.gouv.fr

ABSTRACT

The Grapevine Wood Diseases National Observatory yields a cohesive, large and complex data set which may be dealt with using different approaches, as it is composed of time-varying quantitative and qualitative variables. The objective was to produce the largest possible amount of information from this data set, in order to highlight main trends. To this aim, we used both descriptive and inferential techniques. First, relationships between the different variables are identified using factorial methods, namely multiple correspondence analysis, principal component analysis and factor analysis of mixed data. Then, logistic regressions were used to specify the relationships we highlighted; confidence intervals for the parameters were constructed using bootstrap techniques. Lastly, we used factorial analysis of multi-tables in order to account for the successive three years of data.

Keywords : Multiple correspondence analysis, Principal component analysis, Factor analysis of mixed data, Binary logistic regression, Ordinal logistic regression, Bootstrap, Factor analysis of multi-tables.

1. Introduction

1.1. Problématique

Depuis l'année 2001 et l'interdiction, en France puis en Europe, de l'usage de l'arsénite de soude (ou de sodium) en agriculture, il n'existe plus de méthode de lutte curative homologuée contre les maladies du bois de la vigne, à savoir le black dead arm (BDA), l'esca et l'eutypiose. Une étude réalisée par la mutualité sociale agricole en 2000 a montré que les précautions utilisées n'étaient pas suffisantes pour assurer une bonne protection des utilisateurs¹. Suite à l'interdiction de ce produit phytosanitaire, la profession viticole a exprimé des craintes par rapport à la progression potentielle de ces trois maladies de dépérissement au sein du vignoble français, dont la conséquence ultime est bien souvent la mort des plants². De plus, le manque de références techniques sur l'épidémiologie de ces maladies ne permettait pas d'envisager une solution rapide de remplacement.

Afin de coordonner l'ensemble des dispositifs de recherche et d'acquisition de références sur ces maladies, un groupe technique national, auquel l'INRA appartient, a été constitué en 2001. C'est dans le cadre de ce groupe technique national que la création d'un observatoire pluriannuel, l'Observatoire National des Maladies du Bois de la Vigne, a été décidée en 2003 pour une durée de six ans. Cette structure est unique au monde du fait de la taille du dispositif et du nombre de partenaires impliqués. Ce dispositif a pour objectif de produire un état des lieux de la situation des trois maladies au plan national, d'évaluer l'impact économique réel de cet arrêt et de tenter d'identifier des facteurs discriminants permettant d'expliquer les variations observées de ces maladies.

1. Un rapport de cette enquête est téléchargeable à l'adresse suivante : http://www.itv-midipyrenees.com/publications/itv-colloque/documents/COLLOQUE_Effet-toxique-arsenite.pdf.

2. Un certain nombre de termes techniques propres au vocabulaire de l'épidémiologie végétale ou de la vigne ont été définis dans un lexique accessible sur Internet (Bertrand *et al.* 2008).

1.2. Protocole technique de l'observatoire

- Toutes les régions viticoles³ françaises présentent au moins l'une des trois maladies, ce qui justifie qu'elles participent à l'enquête. En pratique seules 12 zones définies soit comme des régions viticoles soit comme des régions administratives ont été enquêtées : Alsace, Aquitaine, Beaujolais (Rhône-Alpes), Bourgogne, Centre, Diois (Rhône-Alpes), Jura (Franche-Comté), Languedoc-Roussillon, Midi-Pyrénées, Pays de la Loire, Poitou-Charentes, Provence-Alpes-Côte d'Azur. Ces 12 zones représentent 95 % de la superficie totale viticole en France, voir le Tableau A de l'Annexe 6.2 à ce sujet. Notons que la Corse n'a pas été enquêtée et que la région Midi-Pyrénées n'a en fait pas communiqué de résultats avant l'année 2005 et n'a donc pas pu être incluse dans l'analyse réalisée dans cet article.
- Un minimum de 25 parcelles par vignoble et par cépage devait être observé. Tout cépage avec moins de 25 parcelles par région n'est pas retenu pour l'analyse des résultats au niveau national.
- Pour chaque parcelle, choisie au hasard à partir des réseaux d'observation, 300 ceps sont marqués et repérés. Ils sont répartis en dix placettes de 30 ceps choisies également aléatoirement. La taille de la parcelle ne rentre pas en compte dans le processus de sélection de cette dernière puisque la taille de la parcelle n'est pas un facteur pertinent en termes d'épidémiologie végétale.
- Les observations se font sur les mêmes placettes (donc les mêmes ceps) pendant les trois ans de l'enquête dans le but de mieux apprécier l'évolution des maladies.
- Les trois maladies sont observées chaque année à deux périodes particulières du stade de la vigne. À l'occasion de ces relevés est quantifiée la mortalité des plants présents sur la parcelle. Enfin, la notation des complantés, des absents ou des recépés est également réalisée.
- Enfin l'unique critère de choix de la parcelle est le cépage retenu.

Ce protocole technique a été validé par le groupe technique national.

Quant au critère de choix de la parcelle, le cépage constitue un critère d'identification simple et donc adapté à un dispositif de cette ampleur comportant de nombreux intervenants puisque plus de 40 structures différentes effectuent les relevés. D'autre part, des différences de vulnérabilité importantes ont été observées entre les cépages, et ce même au sein d'une même région viticole. De ce fait, une hypothèse ressentie par la communauté de la santé végétale est que le cépage est un des facteurs explicatifs principaux qu'il est donc nécessaire de prendre en compte lors de la planification de l'enquête.

En complément, des informations sur les caractéristiques de la parcelle et les pratiques culturales ont été recueillies par enquête auprès des exploitants viticoles dont les parcelles avaient été choisies au hasard. Une reproduction du questionnaire ayant servi à la récolte des données est disponible sur Internet (Figure B, Bertrand *et al.* 2008).

3. La Figure A de l'Annexe 6.2 situe les régions viticoles enquêtées sur la carte de France.

1.3. Mise en œuvre et plan de l'étude statistique

L'esca et le black dead arm (BDA) ayant les mêmes symptômes foliaires apparents ⁴, ils seront considérés comme une unique maladie que nous noterons désormais esca/BDA. Nous disposons de données sur trois années à savoir 2003, 2004 et 2005 ⁵. Pour expliquer les variations des intensités observées des maladies dans une population de parcelles cultivées, ainsi que celle de la mortalité des plants, les facteurs et les variables suivants ont été pris en compte : la région (facteur), le cépage (facteur), l'âge de la parcelle (variable), la densité de plantation (variable), le type de porte-greffe utilisé (facteur), mais encore le devenir des sarments (facteur), l'enlèvement des bois morts (facteur), le prétaillage ou non des vignes (facteur), le type de taille (facteur) ainsi que les dates de début et de fin de taille (deux variables) et bien entendu le nombre de traitement(s) à l'arsénite de soude effectués avant son interdiction c'est-à-dire sur la période 1999-2001 (facteur).

Il est à noter que l'incidence d'une pathologie représente le pourcentage de pieds atteints par la pathologie parmi l'ensemble des pieds d'une parcelle, et que le taux de mortalité des plants est estimé en cumulant pour chaque parcelle les ceps morts, manquants et les jeunes complants.

Au total, nous disposons d'un jeu de données représentant 701 parcelles observées dans 11 régions viticoles, correspondant à 26 cépages ⁶. (cf. Figure A de l'Annexe 6.2), et rassemblées suivant un protocole précis de saisie, et ce pour chaque année de l'étude. Nous avons retiré de la base de données quatre variétés de vins : le Alphonse Lavallée, l'Italia, le Lival et le Mourvèdre parce qu'ils ne représentaient que 1,28 % du nombre total des parcelles. Nous avons alors éliminé les parcelles présentant des données manquantes ou aberrantes et conservé les 254 parcelles pour lesquelles toutes les informations étaient valides.

Après une analyse exploratoire préliminaire, le but de la première partie de notre étude a été de dégager des grandes tendances en termes d'épidémiologie végétale. À cet effet, nous avons utilisé des techniques d'analyse factorielle sur le tableau moyen des trois années 2003 à 2005. Il s'agit de l'analyse des correspondances multiples, l'analyse en composantes principales et de l'analyse factorielle de données mixtes. Dans un second temps, nous avons eu recours à des modèles de régression logistique, aussi bien binaire qu'ordinaire combinée à l'utilisation de techniques bootstrap. Enfin, dans le but d'appréhender l'évolution des deux maladies (eutypiose, esca/BDA) et du taux de mortalité des plants au cours du temps, mais aussi pour tenir compte de la variable climatique dont il est fortement supposé qu'elle ait eu une influence sur l'expression

4. Voir la Figure C de l'Annexe 6.6 (Bertrand *et al.* 2008) pour une appréciation visuelle de cette ressemblance.

5. À l'instant où les auteurs ont rédigé l'analyse statistique, les données de 2006 n'avaient pas encore été dépouillées.

6. Le Alphonse Lavallée, le Pinot Auxerrois, le Cabernet Franc, le Cabernet Sauvignon, le Carignan, le Chardonnay, le Chenin, le Cinsault ou Cinsaut, le Gamay, le Gewurztraminer, la Grenache, l'Italia, le Lival, le Melon, le Merlot, le Mourvèdre, le Muscat De Hambourg, le Muscat Petits Grains, le Pinot Noir, le Poulsard, le Riesling, le Sauvignon, le Savagnin, le Syrah, le Trousseau et l'Ugni Blanc.

des maladies du bois de la vigne⁷, des analyses factorielles de tableaux multiples, chaque tableau représentant une année, ont été réalisées.

1.4. Analyse exploratoire et tests préliminaires

L'objectif de l'analyse exploratoire des données était de dégager les premières tendances et les liaisons entre les différentes variables. Celle-ci a essentiellement mis en évidence l'importance de l'âge de la parcelle. Ceci nous a conduits à identifier quatre principales classes d'âge (voir le Tableau B de l'Annexe 6.3) découlant de la relation entre l'âge de la parcelle et les trois variables à expliquer. Ces relations diffèrent substantiellement d'une classe à l'autre (Fussler 2006).

Ensuite, nous avons effectué des tests de Kruskal-Wallis. Les procédures de comparaisons multiples réalisées entre les variables de réponse et les facteurs explicatifs ont permis, par exemple, de définir des classes de sensibilité aux maladies pour les différents cépages. Ces résultats ont complété ceux qui étaient connus jusqu'alors.

Consécutivement, non seulement aux tests de Kruskal-Wallis mais aussi aux tests bivariés comme le test du χ^2 d'indépendance et le test exact de Fisher-Freeman-Halton d'indépendance (Freeman et Halton 1951), que nous avons réalisés, nous avons été à même de réduire de 15 à 8 le nombre de variables à considérer dans l'analyse statistique. En effet, nous n'avons sélectionné que les seules variables explicatives exprimant des liens avec les variables à expliquer, c'est-à-dire les incidences des deux maladies et le taux de mortalité des plants. Les huit variables retenues sont les suivantes : la région, le cépage, l'âge de la parcelle, le porte-greffe, le devenir des sarments, le prétaillage, le type de taille, le nombre de traitement(s) à l'arsénite de soude entre 1999 et 2001.

2. Analyses factorielles

Dans cette partie, nous utiliserons les données du tableau moyen, c'est-à-dire que les valeurs des incidences de l'eutypiose, de l'esca/BDA et les taux de mortalité des plants utilisés correspondent aux moyennes des valeurs observées sur les trois années 2003, 2004 et 2005. L'âge de la parcelle a été fixé à sa valeur en 2005. Remarquons que les valeurs des autres variables observées n'évoluent pas au cours du temps.

2.1. Analyse des correspondances multiples (ACM)

2.1.1. Application de l'ACM aux données de l'Observatoire National des Maladies du Bois de la Vigne

La raison pour laquelle nous avons choisi d'appliquer l'ACM en premier lieu est que cette méthode permet d'analyser les deux types de variables, quantitatives et qualitatives, si nous recodons les variables quantitatives, et ne nécessite pas

7. Nous rappelons que l'année 2003 a été marquée par un été caniculaire.

l'existence d'hypothèses *a priori* sur les variables comme une relation linéaire entre elles ou une homogénéité des variances.

Pour pouvoir utiliser l'analyse des correspondances multiples, il est donc nécessaire de créer des catégories pour les variables incidence de l'eutypiose (« Eutypiose »), incidence de l'esca/BDA (« Esca/BDA »), taux de mortalité des plants (« Mortalité ») et âge de la parcelle (« Âge de la parcelle »). Ces classes ont été élaborées, notamment pour l'âge de la parcelle, à partir de l'analyse exploratoire des données (*cf.* paragraphe 1.4) et de façon à obtenir des effectifs équilibrés pour chaque classe afin d'éviter un écrasement des axes par des modalités dont les effectifs seraient trop faibles. La détermination des limites inférieures et supérieures des classes des trois autres variables se font à partir du protocole recommandé par Savary (Savary et Madden 1995). Elles sont récapitulées dans le Tableau B de l'Annexe 6.3.

Dans un premier temps, une analyse des correspondances du tableau disjonctif complet a été réalisée avec comme seules variables actives les trois variables à expliquer et le facteur explicatif suivants : l'incidence de l'eutypiose, l'incidence de l'esca/BDA, le taux de mortalité des plants et l'âge de la parcelle.

Cette ACM a pour premier but de répondre à un des objectifs principaux : caractériser les fortes valeurs de mortalité des plants et les interactions entre les maladies. L'âge de la parcelle a été pris en compte, étant donné son rôle supposé par rapport au taux de mortalité et son association avec les incidences des maladies qui a été mise en évidence lors des tests préliminaires. La variable « Cépage » n'a pas été retenue à cause du grand nombre de ses modalités, 22 au total par rapport aux trois ou quatre modalités des autres variables, ce qui lui aurait conféré artificiellement une importance plus grande dans l'analyse. Il en va de même pour la variable « Région » et les 11 régions viticoles qui la constitue. Enfin, il n'était pas opportun, compte tenu des objectifs de l'observatoire en terme d'épidémiologie végétale d'opérer des regroupements entre les 22 cépages ou entre les 11 régions.

Dans un deuxième temps, le rôle des autres variables mises à notre disposition a été étudié, sous forme d'une ACM reprenant les facteurs principaux précités plus les variables « secondaires » suivantes mises en supplémentaires : la région, le cépage, le nombre de traitement(s) à l'arsénite de soude, le porte-greffe, le prétaillage, le devenir des sarments, le type de taille.

Les Tableaux 1, 2 récapitulent les résultats, valeurs propres, inerties, masses, qualités, coordonnées et cosinus carrés, associés aux variables actives de l'ACM.

Nous ne décidons de garder pour l'analyse que les quatre premiers axes de l'ACM ce qui correspond à environ 60 % de l'inertie. Rappelons que ce pourcentage a été calculé à partir de l'analyse du tableau disjonctif complet et donne de ce fait une vision pessimiste de l'information conservée. L'étude du troisième plan factoriel pourrait être envisagée. Toutefois notre choix est motivé par notre objectif de rechercher les associations les plus fortes entre les variables, par la grande difficulté d'interprétation de ce plan et par la volonté de ne pas surcharger notre étude d'un trop grand nombre de graphiques probablement peu pertinents.

ÉTUDE STATISTIQUE DES DONNÉES : MALADIES DU BOIS DE LA VIGNE

TABLEAU 1. – Valeurs propres et inerties associées aux axes factoriels de l'ACM

Axe	Valeur propre	Pourcentage de l'inertie	
		Individuel	Cumulé
1	0,425	18,904	18,904
2	0,361	16,049	34,953
3	0,301	13,372	48,326
4	0,264	11,720	60,046
5	0,237	10,536	70,582
6	0,206	9,164	79,745
7	0,189	8,411	88,156
8	0,172	7,661	95,818
9	0,094	4,182	100,000

TABLEAU 2. – Coordonnées, contributions et cosinus carrés associés aux modalités des différentes variables pour les deux premiers axes factoriels de l'ACM

	Masse	Qualité	Inertie Relative	Coord.	Inertie	Cosinus ²	Coord.	Inertie	Cosinus ²
				Dim.1	Dim.1	Dim.1	Dim.2	Dim.2	Dim.2
Euty0	0,086	0,657	0,073	-0,093	0,002	0,005	1,030	0,252	0,553
Euty1	0,103	0,587	0,065	0,551	0,074	0,214	-0,439	0,055	0,136
Euty2	0,061	0,756	0,084	-0,802	0,092	0,208	-0,703	0,083	0,159
Esca0	0,013	0,949	0,105	1,002	0,030	0,054	0,802	0,023	0,035
Esca1	0,143	0,429	0,048	0,469	0,074	0,293	-0,249	0,025	0,083
Esca2	0,094	0,622	0,069	-0,844	0,158	0,433	0,268	0,019	0,044
Mort0	0,030	0,882	0,098	1,053	0,077	0,148	1,326	0,144	0,236
Mort1	0,098	0,606	0,067	0,704	0,115	0,322	-0,222	0,013	0,032
Mort2	0,122	0,512	0,057	-0,823	0,194	0,646	-0,142	0,007	0,019
Âge0	0,033	0,866	0,096	0,458	0,016	0,032	1,472	0,201	0,335
Âge1	0,105	0,579	0,064	-0,340	0,029	0,084	0,270	0,021	0,053
Âge2	0,087	0,654	0,073	-0,192	0,007	0,020	-0,617	0,091	0,202
Âge3	0,025	0,902	0,100	1,508	0,132	0,248	-0,986	0,066	0,106

2.1.2. Bilan

Nous reportons ci-dessous quelques interprétations parmi les principales des résultats des Tableaux 1, 2, 3 et 4, (Tableaux 3 et 4 disponibles dans Bertrand *et al.* 2008), ainsi que des Figures 1, 2, 3, 4 et 5 ci-après. Les analyses des Figures 1 et 2 constituent un canevas sur lequel nous ajoutons des variables projetées en supplémentaires.

- Les niveaux de mortalité élevés sont associés à des incidences fortes des deux maladies (*cf.* Figure 1).
- Les fortes incidences d'eutypiose et d'esca/BDA sont respectivement associées à des vignes relativement âgées ($\hat{\text{Age}}_2$: 25-40 ans) et des vignes jeunes ($\hat{\text{Age}}_1$: 15-25 ans) (*cf.* Figure 1).
- Les incidences semblent évoluer en sens contraire, c'est-à-dire que lorsque l'incidence d'une des deux maladies est forte celle de l'autre est faible (*cf.* Figure 2).
- Des relations de vulnérabilité entre la région ou le cépage et les maladies sont notables. Nous avons par exemple les associations Poitou-Charentes (« PCH ») et eutypiose 2 (« Euty2 ») (*cf.* Figure 3, en bas) ou encore Poulсарd (« PLS ») et esca/BDA 2 (« Esca2 ») (*cf.* Figure 3, en haut).
- Nous remarquons la non-prépondérance de l'action de l'arsénite de soude sur l'évolution des trois maladies qui nous intéressent (*cf.* Figure 4). Cette constatation répond à l'une des principales questions que se posait l'observatoire.
- Enfin, l'association entre le brûlage et les niveaux moyens des maladies (« Euty1 » et « Esca1 »), est probablement à mettre en relation avec les régions et les cépages (*cf.* Figure 5, en bas). En effet cette technique est essentiellement utilisée en Bourgogne.

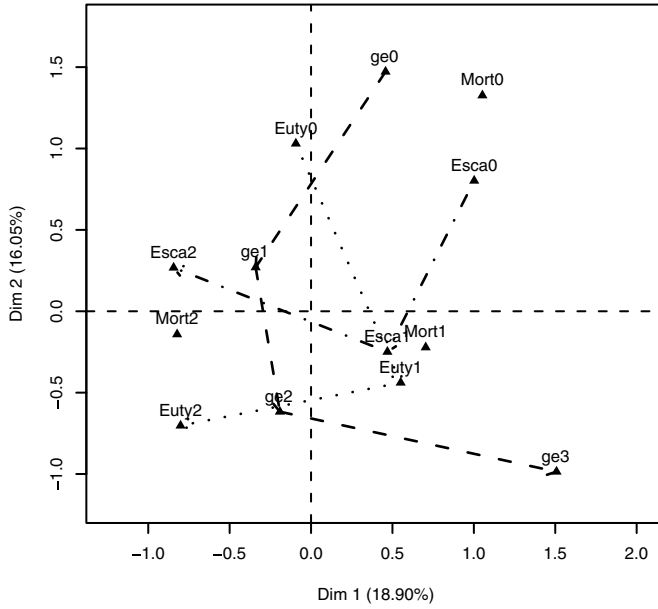


FIG 1. — Représentation graphique des deux premiers axes factoriels de l'ACM

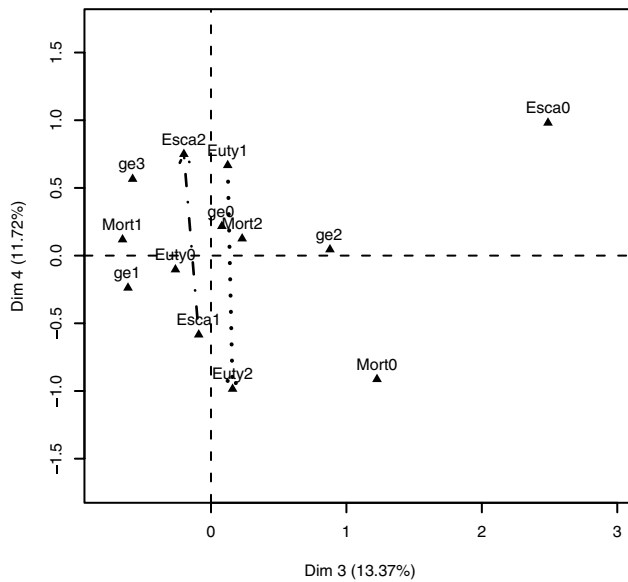


FIG 2. — Représentation graphique des troisième et quatrième axes factoriels de l'ACM

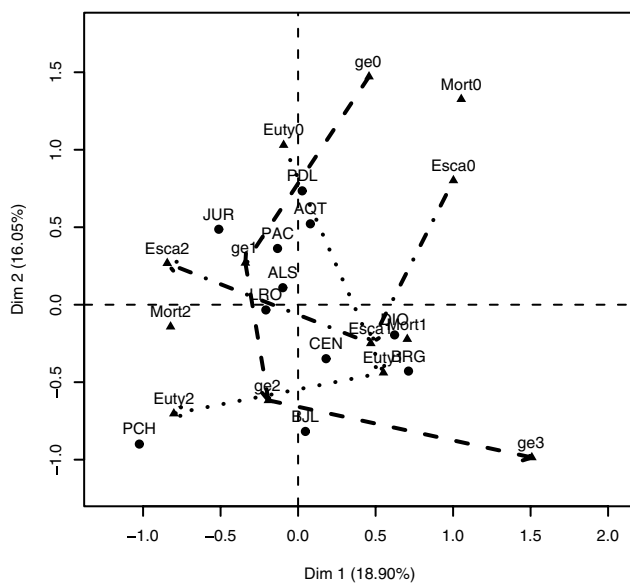
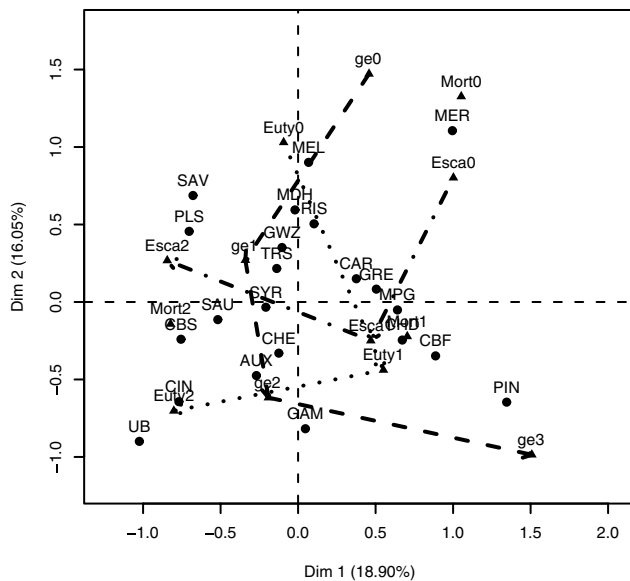


FIG 3. — Représentations graphiques des deux premiers axes factoriels de l'ACM avec respectivement le cépage puis la région mis en supplément

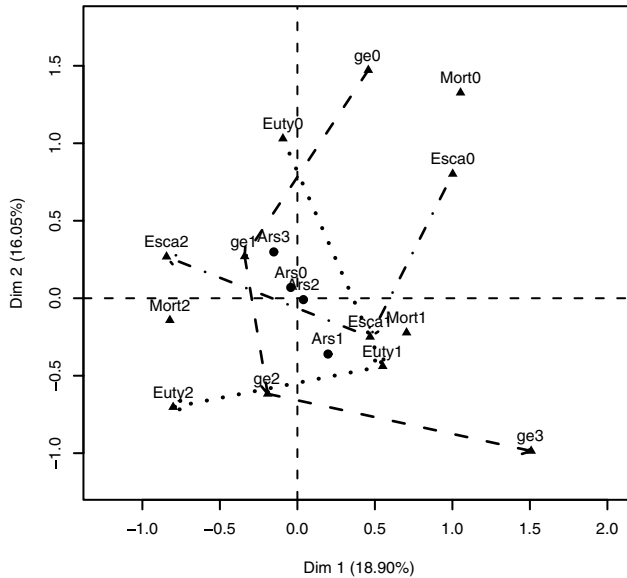


FIG 4. — Représentation graphique des deux premiers axes factoriels de l'ACM avec l'arsénite de soude mis en supplémentaire

2.2. Analyse en composantes principales (ACP)

2.2.1. Application de l'ACP aux données de l'Observatoire National des Maladies du Bois de la Vigne

Une autre stratégie d'analyse est de réaliser une analyse en composantes principales avec pour variables actives les quatre variables : l'incidence de l'eutybiose, l'incidence de l'esca/BDA, le taux de mortalité des plants et l'âge de la parcelle.

TABLEAU 5. – Valeurs propres et inerties associées aux axes factoriels de l'analyse en composantes principales

Axe	Valeur propre	Pourcentage de l'inertie	
		Individuel	Cumulé
1	1,662	41,556	41,556
2	1,210	30,260	71,817
3	0,776	19,411	91,227
4	0,351	8,773	100,000

En procédant de cette manière il n'est plus possible de mettre en évidence et de distinguer les associations existant lors de la phase d'apparition de l'une des maladies et celles correspondant à la phase d'intensification des maladies. Il s'agit bien entendu d'un désavantage par rapport aux techniques d'ACM que nous venons d'utiliser.

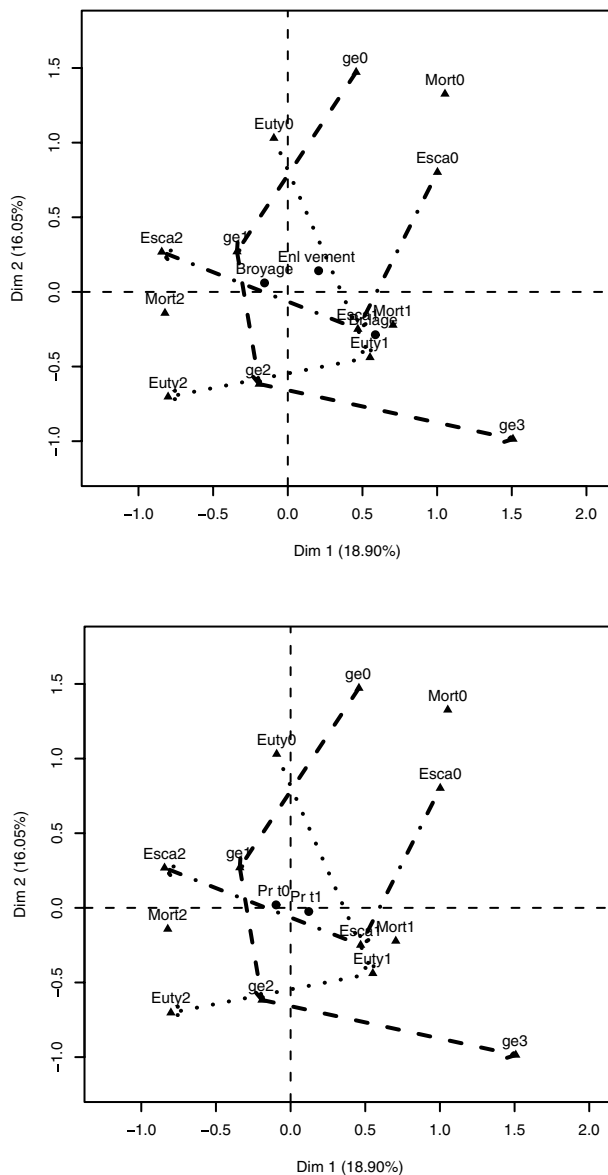


FIG 5. — Représentations graphiques des deux premiers axes factoriels de l'ACM avec respectivement le prétaillage puis le devenir des sarments mis en supplémentaire

Chacune des variables qualitatives sera mise en supplémentaire. Pour éviter de surcharger les graphiques, les variables qualitatives supplémentaires seront représentées l'une après l'autre. Nous reproduisons ici les graphiques obtenus, Figures 6, 7 et 8, afin de les comparer avec ceux produits par la technique

TABLEAU 6. – Corrélation de Pearson des variables de l’analyse en composantes principales

	Âge	Eutypiose	Esca_BDA
Âge	1,000	0,064	-0,262
Eutypiose	0,064	1,000	0,036
Esca_BDA	-0,262	0,036	1,000
Mortalité	0,042	0,426	0,469

que nous introduirons au paragraphe 2.3 qui permet de construire les axes factoriels de telle sorte que ceux-ci expliquent le mieux possible à la fois les variables quantitatives et qualitatives ce qui n’est bien sûr pas le cas de l’ACP. Des représentations graphiques plus détaillées sont disponibles (Figures 6 à 11, Bertrand *et al.* 2008). Elles comportent des ellipses de confiance au seuil de 95 %, obtenues par des techniques bootstrap, qui ont été intégrées aux représentations graphiques des individus dans le premier plan factoriel. Nous nous sommes limités au premier plan factoriel qui restitue 72 % de l’inertie.

TABLEAU 7. – Coordonnées, contributions et cosinus carrés associés aux modalités des différentes variables pour les deux premiers axes factoriels de l’analyse en composantes principales

	Qualité	Coord.	Contrib	Cosinus ²	Coord.	Contrib	Cosinus ²
		Dim.1	Dim.1	Dim.1	Dim.2	Dim.2	Dim.2
Eutypiose	0,641	0,586	20,669	0,344	0,545	24,550	0,297
Esca_BDA	0,753	0,721	31,296	0,520	-0,482	19,194	0,232
Mortalité	0,805	0,877	46,248	0,769	0,190	2,981	0,036
Âge	0,675	-0,172	1,787	0,030	0,803	53,275	0,645

2.2.2. Bilan

Nous établissons ci-dessous quelques commentaires généraux puis nous analyserons les résultats graphiques plus en détails.

- Les deux maladies semblent opposées c’est-à-dire elles n’évoluent pas de manière conjointe, conformément à ce que nous avons observé lors de l’ACM.
- Le taux de mortalité des plants peut être considéré comme étant un vecteur somme des vecteurs des incidences des deux maladies.
- L’âge de la parcelle est porté par le deuxième axe factoriel.
- Nous remarquons l’association entre la région Poitou-Charentes («PCH») et l’eutypiose («Eutypiose») (*cf.* Figure 6, en bas) puis l’association entre

le cépage Ugni Blanc (« UB ») et l'eutypiose (« Eutypiose ») (cf. Figure 7, en bas), conformément à ce que nous avons remarqué lors de l'ACM, ce qui se justifie par le fait que l'Ugni Blanc est caractéristique du Poitou-Charentes.

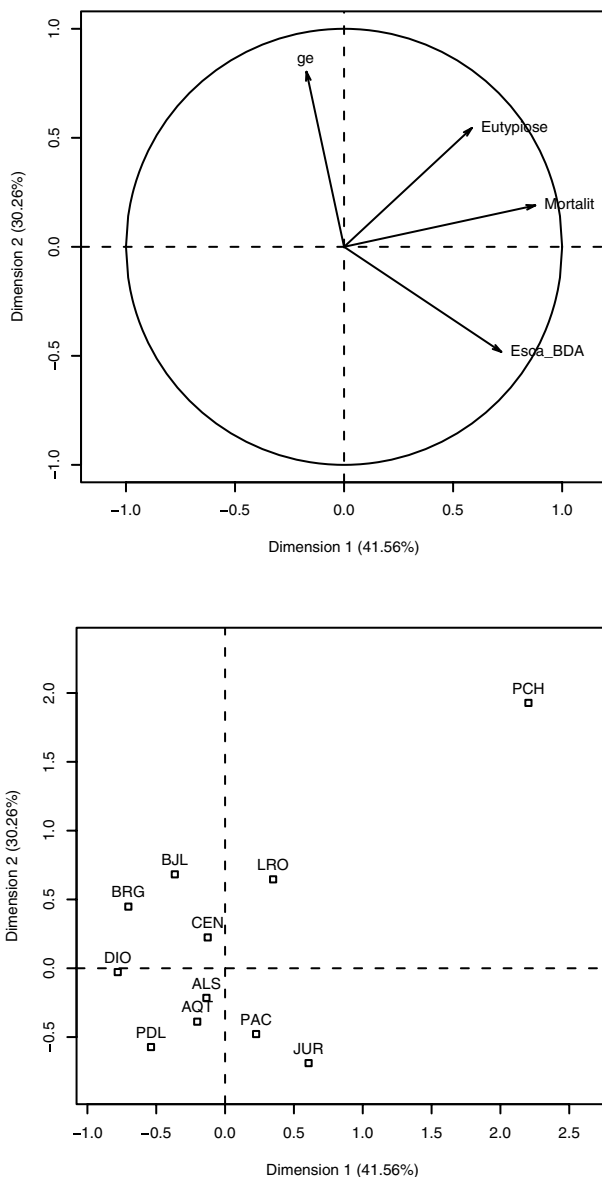


FIG 6. — Représentations graphiques du cercle des corrélations et des deux premiers axes factoriels de l'analyse en composantes principales avec pour variable qualitative la région

- De même nous notons l'association entre la région Jura («JUR») et l'esca/BDA («Esca_BDA») (cf. Figure 6, en bas) puis l'association entre les trois cépages, le Poulsard («PLS»), le Savagnin («SAV»), le Trousseau («TRS») et l'esca/BDA («Esca_BDA») (cf. Figure 7, en bas), ce qui se justifie encore par le fait que les trois cépages sont caractéristiques du Jura.

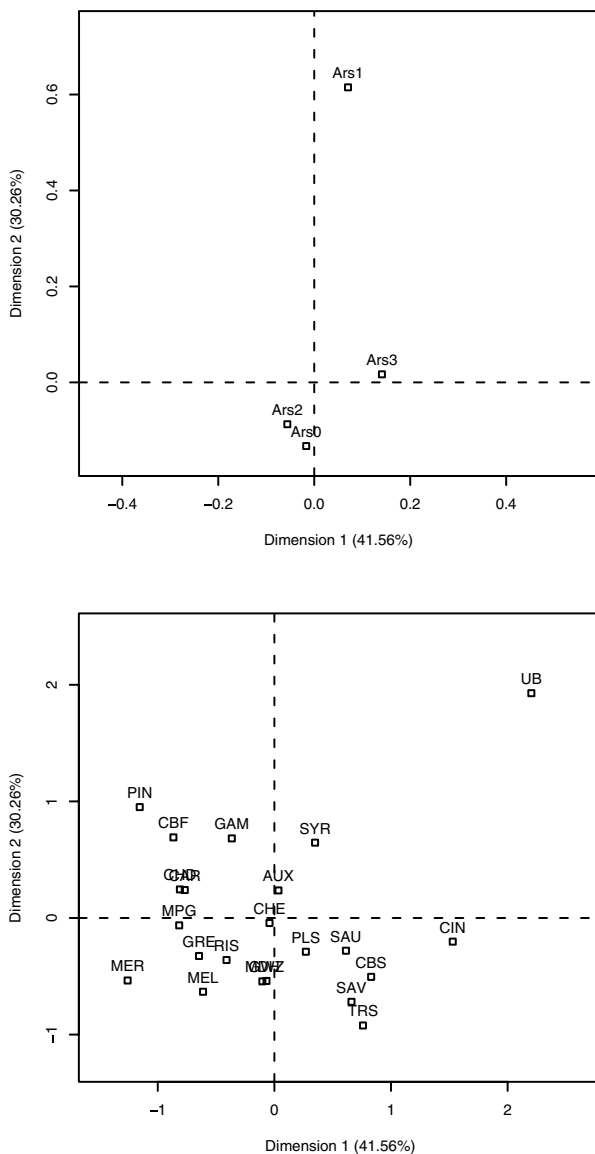


FIG 7. — Représentations graphiques des deux premiers axes factoriels de l'analyse en composantes principales avec pour variables qualitatives l'arsénite de soude et le cépage

- Il semble qu'il existe une opposition entre le taux de mortalité des plants et le fait de ne pas avoir traité («Ars0») ou d'avoir effectué deux traitements («Ars2») à l'arsénite de soude (*cf.* Figure 7, en haut).
- Le prétaillage de la vigne («Prét1») est opposé aux incidences des deux maladies ainsi qu'au taux de mortalité des plants (*cf.* Figure 8, en haut à droite). Cette observation se répète également avec le brûlage des sarments («Brûlage») (*cf.* Figure 8, en bas à gauche).
- Enfin, certains types de porte-greffes, comme le «161-49» et le «41B» apparaissent comme plus vulnérables à l'eutypiose («Eutypiose») (*cf.* Figure 8, en haut à gauche).

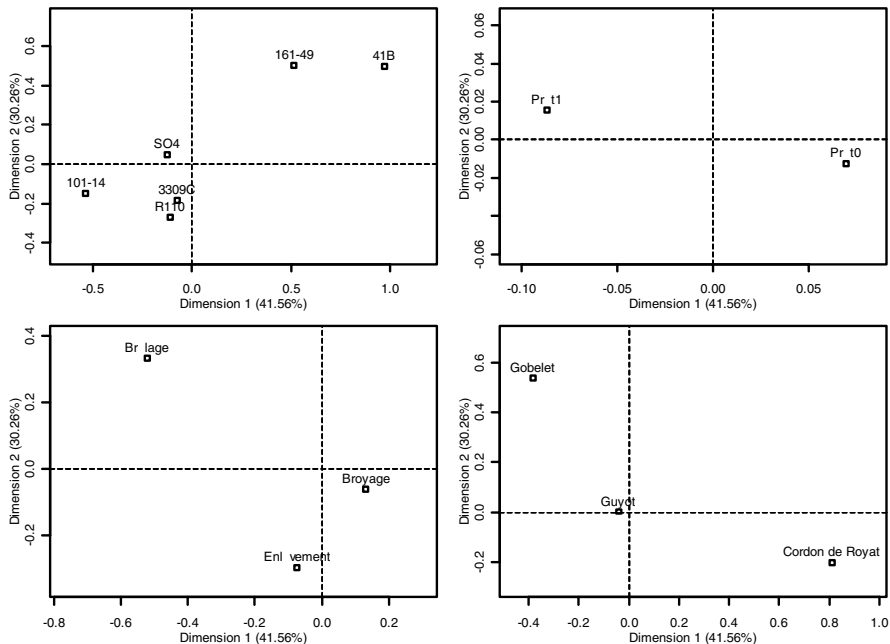


FIG 8. — Représentations graphiques des deux premiers axes factoriels de l'analyse en composantes principales avec pour variables qualitatives le porte-greffe, le prétaillage, le devenir des sarments et le type de taille.

2.3. Analyse factorielle de données mixtes (AFDM)

Au paragraphe 2.1, nous avons réalisé une analyse des correspondances multiples sur le tableau moyen. Cependant, afin de pouvoir utiliser cette technique, les données sur l'incidence de l'eutypiose, celles sur l'incidence de l'esca/BDA, celles sur le taux de mortalité des plants ainsi que celles sur l'âge de la parcelle ont dû être regroupées en classes, ce qui a pu occasionner de possibles réductions de l'information disponible. C'est pourquoi, nous introduisons à présent une technique à même d'étudier simultanément des

variables actives qualitatives et quantitatives : l'analyse factorielle de données mixtes (Pagès 2004 et Escofier 1979).

Afin de rester cohérent avec la mise en œuvre de l'ACM, les sept analyses de données mixtes qui sont réalisées suivent le même schéma d'étude des différentes variables à notre disposition : l'étude d'un tronc commun de quatre variables quantitatives actives, l'incidence de l'eutypiose, l'incidence de l'esca/BDA, le taux de mortalité des plants et l'âge de la parcelle, et d'une variable qualitative active choisie parmi les facteurs restants : région, cépage, nombre de traitement(s) à l'arsénite de soude, porte-greffe, prétaillage, devenir des sarments, type de taille (résultats non présentés dans cet article pour cette dernière variable).

Sur chacune des Figures 9 à 14, sont représentés trois graphiques (une version couleur de chacun est disponible dans Bertrand *et al.* 2008). Sur celui en haut à gauche, les carrés de corrélation des variables quantitatives avec les deux premiers facteurs et le rapport de corrélation de la variable qualitative considérée ont été reportés. Sur le graphique du bas, les modalités de la variable qualitative sont représentées par le centre de gravité des individus ayant adopté cette modalité.

2.3.1. Application aux données de l'Observatoire National des Maladies du Bois de la Vigne

Les variables qui étaient projetées en supplémentaires dans l'ACM sont maintenant prises en compte. Pour pouvoir faire un parallèle entre cette méthode, l'ACM et l'ACP nous reprendrons donc le schéma utilisé auparavant au paragraphe 2.1., c'est-à-dire que les variables «supplémentaires» se grefferont toujours au même tronc commun de variables, composé de l'incidence de l'eutypiose, de l'incidence de l'esca/BDA, du taux de mortalité des plants et de l'âge de la parcelle.

2.3.2. Bilan

Nous proposons ci-après quelques éléments d'interprétation des résultats des Figures 9, 10, 11, 12, 13 et 14.

- Les deux axes factoriels expliquent fortement la variabilité de la variable « Région » compte tenu des valeurs du rapport de corrélation, graphique en haut à gauche Figure 9.
- Un résultat remarquable est que le taux de mortalité des plants correspond approximativement à un vecteur de résultante entre les incidences des deux maladies. La même association avait été mise en évidence avec l'ACM et l'ACP.
- Les deux maladies semblent être opposées, ce qui corrobore les résultats de l'ACM et de l'ACP : les deux maladies n'évoluent pas, ou du moins sensiblement pas, de manière conjointe.
- L'âge de la parcelle se situe toujours sur le deuxième axe factoriel comme nous l'avions observé lors de l'ACM et de l'ACP.

ÉTUDE STATISTIQUE DES DONNÉES : MALADIES DU BOIS DE LA VIGNE

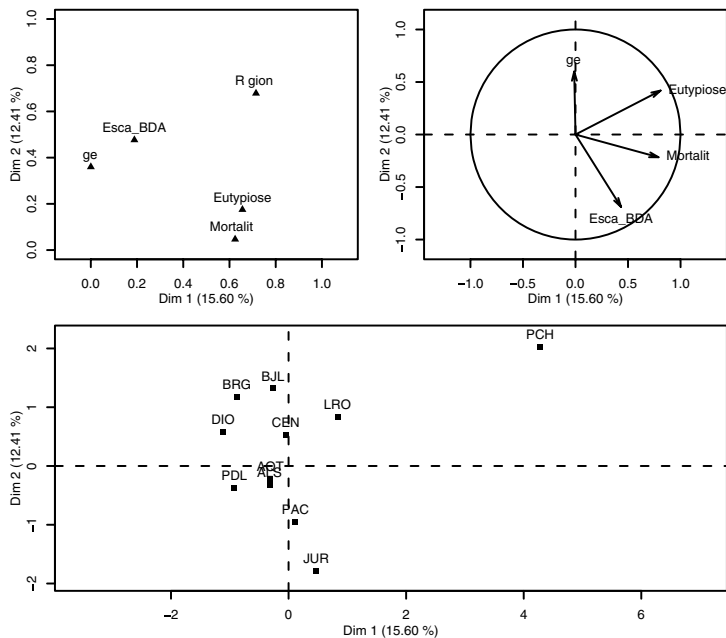


FIG 9. — Représentations graphiques des deux premiers axes factoriels des analyses factorielles de données mixtes avec la région comme variable qualitative active

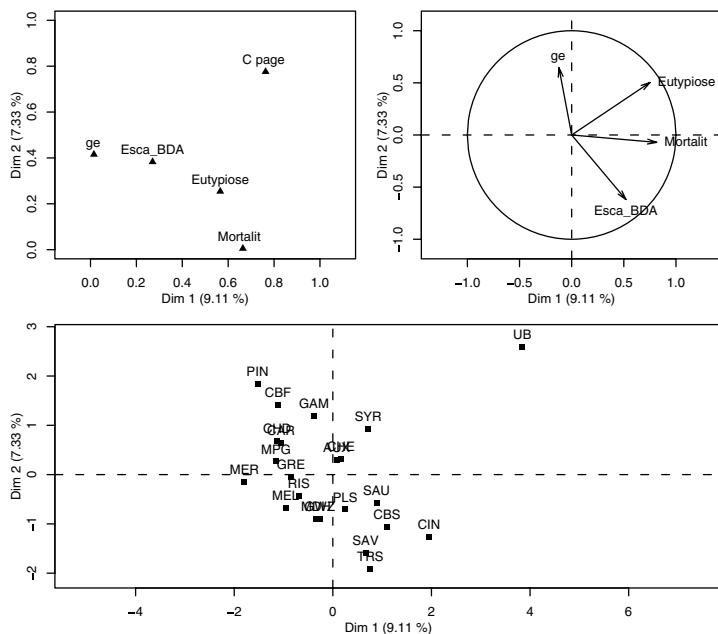


FIG 10. — Représentation graphique des deux premiers axes factoriels de l'analyse factorielle de données mixtes avec le cépage comme variable qualitative active

- Toujours en accord avec l'ACM et l'ACP, nous retrouvons la liaison entre l'esca/BDA et la région Jura. Les autres liaisons semblent être moins évidentes. Ce «non-résultat» est à rapprocher du fait que les régions de par leurs cépages sont à relier aux incidences extrêmes des maladies (c'est-à-dire des incidences faibles voire très faibles ou fortes), ce qui ne sera pas forcément décelable dans ce genre d'analyse où nous prenons en compte le centre de gravité des variables.
- Les deux nouveaux axes factoriels associés aux quatre variables actives et à la variable qualitative «Cépage» expliquent fortement la variabilité de la variable «Cépage» compte tenu des valeurs du rapport de corrélation, graphique en haut à gauche Figure 10.
- En ce qui concerne les relations entre les cépages et les maladies, le Poulsard, le Savagnin et le Trousseau sont proches de l'esca/BDA, comme lors de l'ACM et de l'ACP où ces cépages étaient associés à de fortes incidences de l'esca/BDA.
- Les autres liaisons établies lors de l'ACM sont à nouveau moins marquées et ce selon toute vraisemblance pour les mêmes raisons que celles évoquées précédemment.
- La relation maladies/mortalité reste remarquable tout comme le positionnement de l'âge de la parcelle sur le deuxième axe factoriel.
- Nous remarquons, que l'utilisation de l'arsénite de soude ne semble pas influencer la mortalité des plants (*cf.* Figure 11).

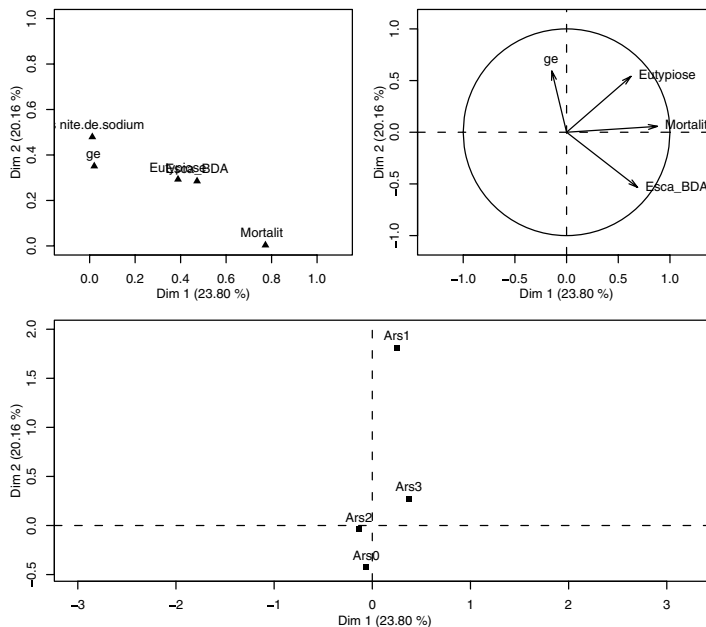


FIG 11. — Représentation graphique des deux premiers axes factoriels de l'analyse factorielle de données mixtes avec l'arsénite de soude comme variable qualitative active

Le prétaillage et le devenir des sarments, deux mesures prophylactiques, ne semblaient pas avoir d'effet sur les maladies et le taux de mortalité lors de l'ACM; qu'en est-il avec l'analyse factorielle de données mixtes?

- Nous retrouvons le même schéma pour l'incidence de l'eutypiose, pour l'incidence de l'esca/BDA, pour le taux de mortalité des plants et pour l'âge de la parcelle que lors de l'ACM, c'est-à-dire que le taux de mortalité augmente avec l'augmentation des incidences de l'eutypiose et de l'esca/BDA.
- Nous observons une opposition entre le taux de mortalité des plants et le fait d'effectuer un prétaillage de la vigne (« Prét1 »), ce qui n'était pas le cas pour l'ACM correspondante mais pour l'ACP correspondante. Toutefois les deux axes factoriels n'expliquent que très peu la variabilité de la variable prétaillage comme l'indiquent les faibles valeurs du rapport de corrélation (*cf.* Figure 12, graphique en haut à gauche).
- Le brûlage des sarments s'oppose aux maladies ainsi qu'au taux de mortalité.
- En ce qui concerne l'enlèvement des sarments (*cf.* Figure 13), il est plus délicat d'arriver à des conclusions étant donné la faiblesse des effectifs, cette pratique n'ayant été observée qu'à sept reprises sur l'échantillon étudié qui concerne 191 parcelles.

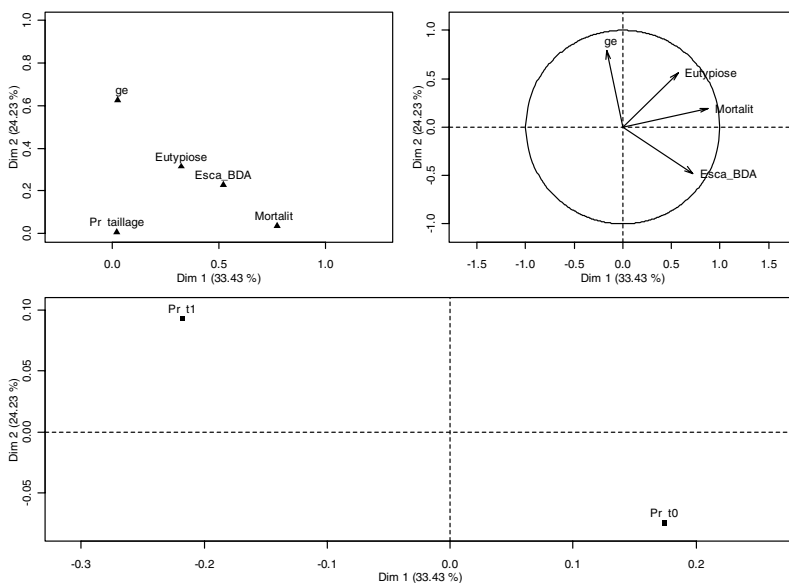


FIG 12. — Représentations graphiques des deux premiers axes factoriels des analyses factorielles de données mixtes avec le prétaillage comme variable qualitative active.

Enfin, que pouvons-nous observer avec l'analyse factorielle de données mixtes en ce qui concerne les porte-greffes? Nous notons que certains types de porte-greffes, comme le « 161-49 » et le « 41B », apparaissent comme plus vulnérables et ce uniquement pour l'eutypiose (*cf.* Figure 14), conformément à ce que nous avons pu observer lors de l'ACP.

ÉTUDE STATISTIQUE DES DONNÉES : MALADIES DU BOIS DE LA VIGNE

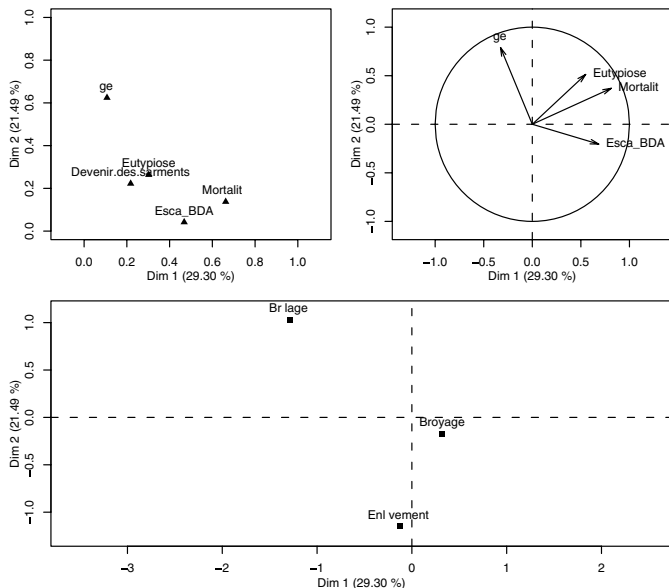


FIG 13. — Représentations graphiques des deux premiers axes factoriels des AFDM avec le devenir des sarments comme variable qualitative active

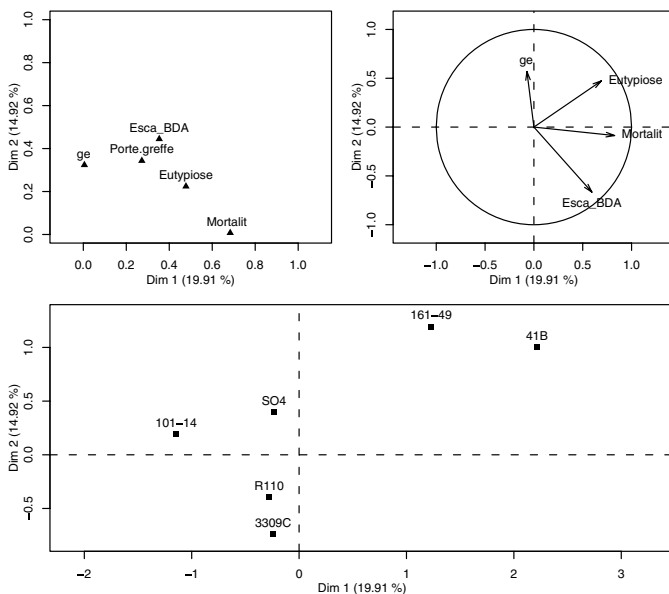


FIG 14. — Représentations graphiques des deux premiers axes factoriels des AFDM avec le devenir le porte-greffe comme variable qualitative active

3. Modèles de régressions logistiques

3.1. Présentation et rappels théoriques

Les modèles de régressions logistiques (Draper et Smith 1998, Hosmer 2000) ont été choisis pour deux raisons.

- La nature du jeu de données est adaptée à la régression logistique étant donné que les variables à expliquer sont soit de type binaire, présence ou absence, soit de type ordinal, nulle, faible ou forte.
- La recherche d'un modèle permettant de valider certaines ou toutes les constatations établies par les analyses factorielles du paragraphe 2.

Ces régressions étant de type binaire ou ordinal, les variables réponses (eutypiose, esca/BDA et mortalité des plants) ont pu être vues sous deux angles différents, que nous nommerons apparition et évolution, comprenant l'apparition des symptômes, c'est-à-dire le passage de l'état sain à l'état malade, et l'évolution des symptômes, c'est-à-dire la transition entre trois états : «sain», «malade et incidence faible», et «malade et incidence forte». Des classes différentes ont été créées ; le Tableau 8 en fait état.

TABLEAU 8. – Descriptif des classes des maladies et du taux de mortalité

Incidences ou taux de mortalité		Eutypiose	Esca/BDA	Mortalité des plants
Apparition	Non	0 %	0 %	< 3 %
	Oui	> 0 %	> 0 %	≥ 3 %
Intensification/Évolution	Nulle	0 %	0 %	< 3 %
	Faible	≤ 2 %	≤ 3 %	< 10 %
	Forte	> 2 %	> 3 %	≥ 10 %

Différents modèles de régressions logistiques ont été testés afin d'expliquer séparément la mortalité et l'incidence de chacune des maladies. Pour prendre en compte les autres variables ainsi que leurs associations et éviter certains problèmes de confusions d'effets des modèles emboîtés ont été étudiés en particulier lorsque le facteur région de la parcelle ou/et le facteur cépage intervenait.

3.1.1. Régression logistique binaire

Si la variable Y représente la présence de la maladie ou le fait que le plant soit considéré comme mort et ne prend de ce fait que les valeurs 0 ou 1, absence ou présence de la maladie, productivité ou improductivité, la probabilité de l'incidence de la maladie ou du taux de mortalité peut être modélisée comme suit :

$$P([Y = 1]) = \frac{\exp(\alpha + \sum \beta_i X_i)}{1 + \exp(\alpha + \sum \beta_i X_i)},$$

soit encore

$$\text{logit}(P([Y = 1])) = \alpha + \sum \beta_i X_i,$$

où α est une constante liée au modèle, β_i les paramètres à estimer, X_i les covariables ou cofacteurs, dont la définition est rappelée au paragraphe 3.1.3, et la fonction suivante :

$$\text{pour tout } x \text{ dans }]0; 1[, \text{logit}(x) = \ln\left(\frac{x}{1-x}\right).$$

3.1.2. Régression logistique ordinale

Si la variable Y prend k valeurs C_1, \dots, C_k , correspondant à des classes ordonnées par exemple, nous définissons les cotes du succès d'un des évènements $[Y \in C_1], \dots, [Y \in C_1 \cup \dots \cup C_{k-1}]$:

$$\begin{aligned} \theta_1 &= \frac{P([Y \in C_1])}{1 - P([Y \in C_1])}, \\ &\vdots \\ \theta_{k-1} &= \frac{P([Y \in C_1 \cup \dots \cup C_{k-1}])}{1 - P([Y \in C_1 \cup \dots \cup C_{k-1}])}. \end{aligned}$$

Alors le modèle s'écrit :

$$\ln(\theta_j) = \alpha_j - \sum \beta_i X_i \quad \text{pour } j = 1, \dots, k - 1,$$

où α_j est une constante associée au modèle, β_i les paramètres à estimer et X_i les covariables ou cofacteurs.

3.1.3. Modèle emboîté

Nous entendons par covariable une variable explicative impliquée dans un modèle de type quantitatif tandis qu'un cofacteur est une variable explicative impliquée dans un modèle de type qualitatif. Un modèle emboîté est un modèle qui fait intervenir des cofacteurs imbriqués les uns dans les autres suivant une hiérarchisation répondant à des contraintes expérimentales : certains cépages sont spécifiques de certaines régions : par exemple, le Riesling ou le Gewurztraminer sont des cépages typiquement alsaciens. Pour dénoter les emboitements le cofacteur imbriqué est mis entre parenthèses : en écrivant $ALS(GWZ)$, nous emboîtons le cépage Gewurztraminer dans la région Alsace.

3.1.4. Critère de bon classement

Afin de pouvoir apprécier la qualité des modèles que nous allons introduire, nous avons utilisé différents critères de bon classement : le D de Somers (Somers 1962), le γ de Goodman et Kruskal (Goodman et Kruskal 1954) et le τ_a de Kendall (Kendall 1938).

3.2. Application aux données de l'Observatoire National des Maladies du Bois de la Vigne

3.2.1. Modèles sélectionnés

La qualité des modèles de régressions logistiques est évaluée de plusieurs manières. Tout d'abord l'ajustement du modèle en termes de tests du χ^2 de Pearson, du χ^2 de déviance et du χ^2 de Hosmer-Lemeshow, doit être vérifié au seuil de 5 %. Le test de la nullité de toutes les pentes doit être significatif au seuil de 5 %. Les modèles retenus ici sont ceux qui sont les plus complets et pour lesquels les variables explicatives restent significatives au seuil de 5 %.

Parmi tous ces modèles, seuls ceux qui sont les plus informatifs en termes d'épidémiologie végétale ont été reproduits dans le Tableau 9. La définition de ces modèles ainsi que l'évaluation de la qualité de leur ajustement, constituent le Tableau 10.

Les résultats sont remarquables en termes de critères de bon classement pour expliquer l'incidence de l'eutypiose ou le taux de mortalité des plants. Par contre, les modèles pour l'esca/BDA sont moins satisfaisants.

3.2.2. Bootstrap et régressions logistiques

Nous pouvons nous demander si les estimations par intervalle des paramètres des différents modèles obtenues dans le cadre de l'approximation asymptotique classique sont précises. Une façon de répondre à cette question est l'utilisation de techniques bootstrap (Davison et Hinkley 1997) pour construire des intervalles de confiance pour ces paramètres.

La méthode des percentiles a été appliquée aux paramètres des régressions logistiques présentées dans le Tableau 9. Les résultats obtenus, sauf pour ceux concernant le modèle 2, ont été reproduits dans le Tableau 11.

3.2.3. Bilan

- Nous pouvons distinguer, sur la Figure 15 associée au Modèle 1, Tableau 11, deux phases dans l'évolution de l'apparition des symptômes par rapport à l'âge de la parcelle. La première s'étale de 0 à 35 ans qui dénote une croissance rapide : la probabilité d'être malade passe de 0,60 à 0,96 sur cette période, ce qui nous donne un taux d'accroissement annuel de l'ordre de 1 %. La deuxième concerne les vignes de plus de 35 ans où là l'accroissement est beaucoup plus lent.
- Le Modèle 3, apparition de l'esca/BDA, est le seul modèle à traduire une influence significative au seuil de 5 % du traitement à l'arsénite de soude.
- Pour le Modèle 1 nous retrouvons quatre cépages parmi les moins sensibles à l'eutypiose. Les signes de leurs coefficients sont négatifs ce qui confirme les analyses précédentes en termes de relations maladie-cépage.
- Concernant la mortalité des plants, nous disposons de deux modèles différents et adéquats : les Modèles 5I et 5II. Leur point commun est l'incidence de l'esca/BDA, ce qui prouve toute l'importance de cette maladie sur la mortalité de la vigne.

ÉTUDE STATISTIQUE DES DONNÉES : MALADIES DU BOIS DE LA VIGNE

TABLEAU 9. – Modèles obtenus par régressions logistiques : estimations des coefficients

Modèle	Incidence de l'eutypiose : Apparition
1	$\text{logit}(P(\ll \text{Malade} \gg)) = 0,446 + 0,082 \hat{\text{Age}} - 1,527I_{\{GWZ\}} - 3,031I_{\{PLS\}} - 1,935I_{\{RIS\}} - 3,870I_{\{SAV\}}$
Modèle	Incidence de l'eutypiose : Intensification
2	$\ln(\theta_1) = 1,423 + 0,051 \hat{\text{Age}} - 1,819I_{\{BRG\}} - 2,630I_{\{CEN\}} - 1,758I_{\{DIO\}} - 3,962I_{\{JUR\}} - 2,951I_{\{PAC\}} - 2,596I_{\{GWZ(ALS)\}} - 2,944I_{\{RIS(ALS)\}} - 4,170I_{\{MER(AQT)\}} + 4,420I_{\{PIN(BRG)\}} + 4,559I_{\{CHE(CEN)\}} + 2,565I_{\{SYR(DIO)\}},$ $\ln(\theta_2) = -2,116 + 0,051 \hat{\text{Age}} - 1,819I_{\{BRG\}} - 2,630I_{\{CEN\}} - 1,758I_{\{DIO\}} - 3,962I_{\{JUR\}} - 2,951I_{\{PAC\}} - 2,596I_{\{GWZ(ALS)\}} - 2,944I_{\{RIS(ALS)\}} - 4,170I_{\{MER(AQT)\}} + 4,420I_{\{PIN(BRG)\}} + 4,559I_{\{CHE(CEN)\}} + 2,565I_{\{SYR(DIO)\}}.$
Modèle	Incidence de l'esca/BDA : Apparition
3	$\text{logit}(P(\ll \text{Malade} \gg)) = 3,384 - 1,133I_{\{ARS1\}} - 1,998I_{\{ARS3\}}$
Modèle	Incidence de l'esca/BDA : Intensification
4	$\ln(\theta_1) = 3,922 - 3,938I_{\{CAR\}} + 1,603I_{\{GWZ\}} - 2,904I_{\{MER\}} - 2,253I_{\{MPG\}} - 2,211I_{\{PIN\}} + 1,667I_{\{PLS\}} + 2,001I_{\{SAV\}} - 4,620I_{\{SYR\}}$ $\ln(\theta_2) = 0,968 - 3,938I_{\{CAR\}} + 1,603I_{\{GWZ\}} - 2,904I_{\{MER\}} - 2,253I_{\{MPG\}} - 2,211I_{\{PIN\}} + 1,667I_{\{PLS\}} + 2,001I_{\{SAV\}} - 4,620I_{\{SYR\}}$
Modèle	Taux de mortalité des plants : Apparition
5I	$\text{logit}(P(\ll \text{Mort} \gg)) = -2,865 + 276,535 \text{Esca/BDA} + 0,161 \hat{\text{Age}}$
5II	$\text{logit}(P(\ll \text{Mort} \gg)) = -0,084 + 413,693 \text{Eutypiose} + 240,379 \text{Esca/BDA}$
Modèle	Taux de mortalité des plants : Évolution
6	$\ln(\theta_1) = 2,112 + 22,520 \text{Eutypiose} + 0,030 \hat{\text{Age}} - 1,977I_{\{CBF\}} - 2,083I_{\{CHD\}} - 3,867I_{\{MER\}} - 1,539I_{\{MPG\}} - 2,180I_{\{PEN\}}$ $\ln(\theta_2) = -0,565 + 22,520 \text{Eutypiose} + 0,030 \hat{\text{Age}} - 1,977I_{\{CBF\}} - 2,083I_{\{CHD\}} - 3,867I_{\{MER\}} - 1,539I_{\{MPG\}} - 2,180I_{\{PEN\}}$

- Par exemple, le Modèle **5I** implique à la fois l'esca/BDA et l'âge de la parcelle avec des résultats fort probants : 96,4 % de concordants. La mortalité augmente donc simultanément avec l'incidence de l'esca/BDA et l'âge de la parcelle. Plus particulièrement, trois phases successives se distinguent. Une première de 0 à 15 ans où la croissance est moyenne et la probabilité de mortalité est inférieure à 0,5 soulignant le fait que la mortalité est faible pour les vignes «jeunes», une deuxième jusqu'à 35-40 ans où la croissance est forte (allant jusqu'à 0,90 en probabilité) signifiant quant à elle la hausse de la mortalité durant cette période, et enfin une dernière phase où l'accroissement de mortalité est marginal pour les vignes âgées.
- Le Modèle **6** correspond à l'évolution de la mortalité des plants en fonction de l'âge de la parcelle et de l'eutypiose, passage d'une mortalité faible à une mortalité moyenne et d'une mortalité moyenne à une mortalité élevée.

TABLEAU 10. – Modèles obtenus par régressions logistiques : critères de bon classement

Variable à expliquer	Phase	Modèle	Concordants (%)	D	γ	τ_a
Incidence de l'eutypiose	Apparition	1	89,3	0,79	0,79	0,36
	Intensification	2	89,0	0,78	0,79	0,51
Incidence de l'esca/BDA	Apparition	3	45,4	0,30	0,50	0,03
	Intensification	4	78,1	0,62	0,66	0,33
Taux de mortalité des plants	Apparition	5I	96,4	0,93	0,93	0,10
		5II	94,6	0,90	0,90	0,09
	Évolution	6	79,8	0,60	0,60	0,36

TABLEAU 11. – Paramètres estimés pour chaque modalité incluse dans les Modèles **1**, **3**, **4**, **5I**, **5II** et **6** et leurs intervalles de confiance pour la méthode des percentiles

Modèle	Paramètre	Estimation	Écart-type	Intervalle de confiance
Apparition				
Modèle 1 logit(<i>Eutypiose</i>) ~ Âge + Cépage	Constante	0,446	0,834	[-2, 252; 17, 020]
	Âge	0,082	0,021	[0, 055; 0, 111]
	GWZ	-1,527	0,768	[-2, 779; -0, 194]
	PLS	-3,031	1,034	[-5, 019; -0, 850]
	RIS	-1,935	0,790	[-3, 261; -0, 621]
	SAV	-3,870	1,001	[-5, 279; -2, 474]
Modèle 3 logit(<i>Esca/BDA</i>) ~ <i>Ars</i>	Constante	3,384	0,415	[2, 936; 3, 666]
	Ars1	-1,133	1,106	[-2, 597; 0, 687]
	Ars3	-1,998	0,893	[-3, 770; 6, 217]
Modèle 5I logit(<i>Mortalité</i>)~ <i>Esca/BDA</i> +Âge	Constante	-2,865	1,075	[-6, 761; -1, 036]
	<i>Esca/BDA</i>	276,535	95,019	[178, 359; 533, 5164]
	Âge	0,161	0,048	[0, 100; 0, 329]
Modèle 5II logit(<i>Mortalité</i>)~ <i>Eutypiose</i> + <i>Esca/BDA</i>	Constante	-0,084	0,498	[-1, 497; 0, 852]
	<i>Eutypiose</i>	413,693	187,216	[182, 988; 1689, 516]
	<i>Esca/BDA</i>	240,379	90,363	[151, 406; 496, 307]
Intensification				
Modèle 4 logit(<i>Esca/BDA</i>) ~ <i>Cépage</i>	Constante1	3,922	0,656	[2, 934; 6, 027]
	Constante2	0,968	0,104	[0, 191; 1, 990]
	CAR	-3,938	1,548	[-20, 236; 0, 962]
	GWZ	1,603	0,629	[0, 494; 3, 121]
	MER	-2,904	1,260	[-15, 816; -0, 860]
	MPG	-2,253	0,850	[-4, 870; -0, 148]
	PIN	-2,211	0,837	[-4, 325; -1, 048]
	PLS	1,667	0,838	[0, 080; 3, 354]
	SAV	2,001	0,690	[0, 811; 3, 953]
	SYR	-4,620	1,386	[-20, 327; -1, 278]

TABLEAU 11. – (Suite)

Modèle	Paramètre	Estimation	Écart-type	Intervalle de confiance
Évolution				
Modèle 6 $\text{logit}(\text{Mortalité}) \sim$ $\text{Eutypiose} + \hat{\text{Age}}$ $+ \text{Cépage}$	Constante1	2,112	0,610	[0, 989; 3, 204]
	Constante2	-0,565	0,092	[-1, 280; 0, 954]
	Eutypiose	22,520	8,350	[12, 830; 32, 010]
	Âge	0,030	0,013	[0, 005; 0, 054]
	CBF	-1,977	0,759	[-3, 620; -0, 508]
	CHD	-2,083	0,674	[-3, 282; -0, 799]
	MER	-3,867	1,265	[-19, 734; -1, 013]
	MPG	-1,539	0,639	[-2, 874; -0, 309]
PIN	-2,180	0,680	[-3, 532; -0, 799]	

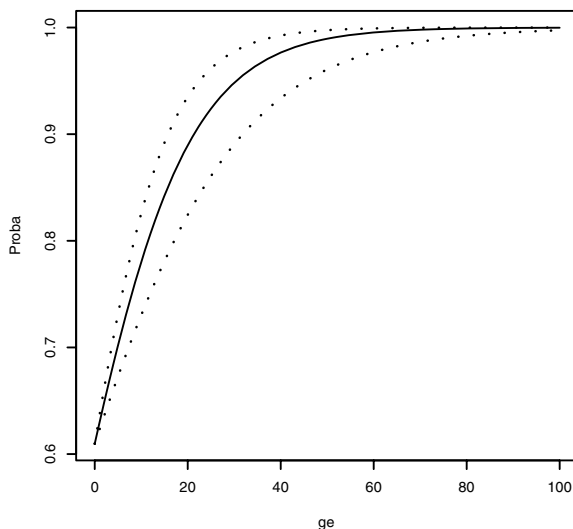


FIG 15. — Probabilité d’avoir au moins un pied symptomatique (eutypiose, phase d’apparition, Modèle 1) dans une parcelle du réseau en fonction de l’âge de la parcelle avec les limites de confiance obtenues par la technique bootstrap de la méthode des percentiles

4. Analyse factorielle de tableaux multiples

À cette étape de l’étude, nous nous sommes posés la question du rôle que peut jouer la variable temporelle dans l’analyse statistique. En effet, bien que satisfaisants, les résultats précédents sont tous basés sur un tableau moyen des trois années 2003, 2004, 2005. Or décrire l’évolution des maladies est aussi l’un des objectifs de l’observatoire national. De plus l’été 2003 présente

une particularité climatique importante du fait de la canicule. Ces raisons légitiment donc de donner toute son importance à la variable « Année ».

Une variante des techniques d'analyse factorielle des tableaux multiples, valide lorsque les tableaux comportent les mêmes individus et les mêmes variables et rappelée dans l'article de Cazes (2004), est utilisée ici afin de pouvoir mieux appréhender l'aspect chronologique de l'évolution des incidences de l'eutypiose et de l'esca/BDA et du taux de mortalité des plants. Cette méthode va permettre, par exemple, d'observer si une année occupe une place prépondérante dans l'explication de l'évolution globale des maladies et du taux de mortalité, ou encore de déceler si une des trois années induit des comportements ou des associations radicalement différents des deux autres.

TABLEAU 12. – Stratégie d'analyse par AFC

	Colonnes Actives	Colonnes supplémentaires		
Lignes actives	D_{Tot}	D_{2003}	D_{2004}	D_{2005}
Lignes supplémentaires	D_{2003}			
	D_{2004}			
	D_{2005}			

4.1. Présentation

Nous avons réalisé l'analyse des correspondances du tableau disjonctif complet, D_{Tot} , introduit au paragraphe 2.1 avec les tableaux disjonctifs complets de l'année 2003, D_{2003} , de l'année 2004, D_{2004} et de l'année 2005, D_{2005} mis en lignes et en colonnes supplémentaires.

4.2. Application aux données de l'Observatoire National des Maladies du Bois de Vigne

Nous sommes plus particulièrement intéressés par la stabilité temporelle des associations entre les modalités des différentes variables que nous jugerons comme satisfaisante si elle s'observe sur les quatre premiers axes factoriels.

De manière générale, nous retrouvons les résultats déjà constatés lors de l'ACM du paragraphe 2.1. De plus, l'ensemble des années sont proches et ce pour l'ensemble des modalités des variables. Cela correspond donc à une stabilité de l'expression des symptômes pour les années 2003, 2004 et 2005, ce qui avait déjà été remarqué lors de l'analyse préliminaire. Par exemple pour l'eutypiose, l'incidence moyenne est de 3,5 %, 3,5 % et 3,6 % respectivement pour les années 2003, 2004 et 2005.

Remarque. — D'autres méthodes factorielles d'analyse de tableaux multiples ont été effectuées lors de cette étude, notamment l'analyse factorielle multiple (Escofier et Pagès 1998) et la méthode STATIS (Lavit *et al.* 1994). Elles conduisent aux mêmes résultats que l'analyse des correspondances réalisée au paragraphe 4.2.

5. Conclusion

Cette étude montre la complémentarité de différentes méthodes statistiques. Son objectif était d'élaborer une méthodologie statistique performante pour traiter des données quantitatives, qualitatives et dynamiques. En effet, les relations entre les variables étudiées ont fait l'objet de peu d'études expérimentales permettant d'affirmer leur association. Il s'agissait donc d'étudier en profondeur le jeu de données, en portant une attention plus particulière aux effets de l'arsénite de soude sur les trois maladies, le black dead arm, l'esca et l'eutypiose, et à l'évolution de celles-ci au cours du temps.

Ainsi, l'analyse des correspondances multiples a établi un panorama des associations entre les variables les plus importantes. L'analyse en composantes principales avec les variables qualitatives en supplémentaires ne permet pas de retrouver le même niveau de complexité dans les associations entre les variables que celui de l'analyse des correspondances multiples. En effet seules des liaisons de type linéaire seront pleinement mises en évidence par l'analyse en composantes principales. L'analyse factorielle de données mixtes, contrairement à l'analyse en composantes principales, permet de prendre en compte dans la construction des axes factoriels le souci de la bonne représentation des variables qualitatives. L'analyse factorielle de données mixtes a confirmé les résultats de l'analyse des correspondances multiples et a dégagé d'autres relations entre les variables. La régression logistique a permis de modéliser ces relations, nos variables à expliquer étant de type binaire ou ordinal. Enfin, l'analyse factorielle de tableaux multiples autorise la prise en compte du temps et ainsi de juger de l'évolution des maladies et du taux de mortalité et de la stabilité des relations entre ces variables au cours du temps. Avec l'apport des données des années à venir, de nouvelles techniques pourront être mises en œuvre, notamment pour mieux appréhender la question de l'évolution des maladies.

Un premier résultat essentiel dégagé par l'ensemble de ces analyses est que ces trois maladies présentent des variations importantes, significatives, selon les cépages et donc selon certaines régions viticoles. Par la suite, nous avons pu noter que, de manière individuelle l'eutypiose évolue essentiellement et parallèlement avec l'âge de la parcelle, alors que le black dead arm et l'esca touchent particulièrement les plants jeunes et sont liés aux traitements à l'arsénite de soude dans sa phase d'intensification, c'est-à-dire le passage d'incidences du black dead arm et de l'esca faibles à fortes, ce qui n'est pas le cas pour l'eutypiose. En ce qui concerne le taux de mortalité des plants, les facteurs prépondérants sont l'eutypiose, le black dead arm, l'esca et l'âge de la parcelle. Quant aux traitements à l'arsénite de soude, ils n'apportent pas de contribution significative à la description de l'évolution du taux de mortalité. Enfin, une perspective de travail sera l'application de techniques (Dray-Petorelli-Chessel 2003) basées sur l'algorithme NIPALS qui permettra de considérer l'ensemble des données récoltées même si elles sont incomplètes.

Remerciements : Les auteurs tiennent à exprimer leur sincère reconnaissance aux relecteurs anonymes et aux éditeurs en chef de la revue pour leurs conseils avisés qui ont permis d'améliorer grandement la qualité du fond et de la forme de cet article.

6. Annexes

6.1. Logiciel et programmes

Tous les calculs réalisés dans cet article ont été menés à bien avec le logiciel **R** (Team 2007). Les analyses factorielles des sections 2 et 4 proviennent de l'utilisation du package FactoMineR (Husson, Lê et Mazet 2007) adapté compte tenu de la spécificité de certains aspects de cette étude. Les résultats de l'application de techniques bootstrap proviennent de l'utilisation du package boot (Canty et Ripley 2006) du logiciel **R** (Team 2007).

L'intégralité du code des programmes ainsi que le jeu de données est disponible sur simple demande auprès des auteurs mais ne peut être intégré à cet article compte tenu de sa longueur.

6.2. Localisation géographique des principales régions viticoles et des cépages qui y sont principalement cultivés

TABLEAU A. – Tableau des surfaces viticoles par régions administratives françaises

Régions administratives	Superficie recouverte par des plants de vignes (ha)	Pourcentage de la superficie française totale recouverte par des plants de vignes (%)
Alsace	15 160,14	1,7
Aquitaine	150 727,61	17,3
Auvergne	1 096,32	0,1
Bourgogne	29 973,03	3,4
Centre	22 244,65	2,5
Champagne-Ardenne ¹	28 181,57	3,2
Corse	7 089,85	0,8
Franche-Comté	2 017,70	0,2
Île de France	47,51	0,0
Languedoc-Roussillon	297 227,77	34,0
Limousin	49,52	0,0
Lorraine	181,51	0,0
Midi-Pyrénées	41 000,29	4,7
Pays de la Loire	37 877,60	4,3
Picardie	2 810,02	0,3
Poitou-Charentes	80 794,61	9,2
Provence-Alpes-Côte d'Azur	99 850,91	11,4
Rhône-Alpes	57 371,59	6,6
Total	873 702,20	100,0
Total enquête	834 245,90	95,5

1. La région Champagne-Ardenne dispose de son propre observatoire des maladies du bois de la vigne.

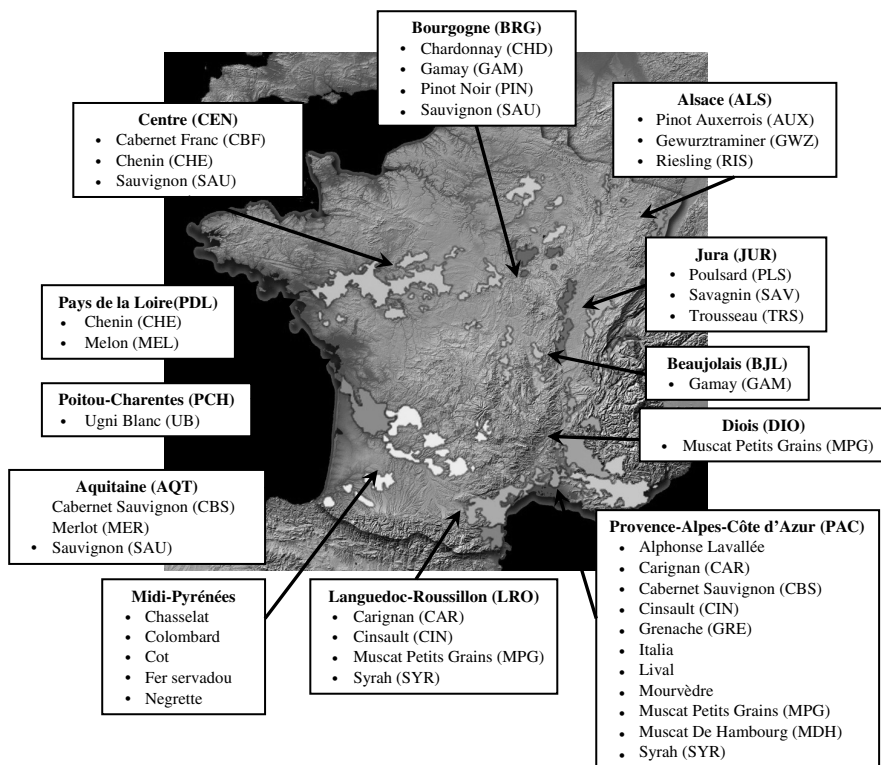


FIG A. — Principales régions viticoles françaises et cépages majoritairement cultivés dans ces régions. Les cépages et les régions présents dans le jeu de données analysé sont accompagnés du code qui a été introduit pour les identifier au Tableau B

6.3. Codage des variables

(Voir tableau ci-après)

TABLEAU B
Tableau descriptif des variables et de leurs modalités

Variable	Symbole	Catégories	Définition des catégories	Unité
<i>Maladies et mortalité</i>				
Eutyopiose	Euty	Euty0 ; Euty1 ; Euty2	Euty0 : Euty = 0 ; Euty1 : 0 < Euty ≤ 2 ; Euty2 : 2 < Euty	%
Esca/BDA	Esca	Esca0 ; Esca1 ; Esca2	Esca0 : Esca= 0 ; Esca1 : 0 < Esca ≤ 3 ; Esca2 : 3 < Esca	%
Mortalité	Mort	Mort0 ; Mort1 ; Mort2	Mort0 : 0 ≤ Mort < 3; Mort1 : 3 ≤ Mort < 10; Mort2 : 10 ≤ Mort	%
<i>Variable explicative</i>				
Âge de la parcelle	Âge	Âge0 ; Âge1 ; Âge2 ; Âge3	Âge0 : 0 ≤ Âge < 15; Âge1 : 15 ≤ Âge < 25; Âge2 : 25 ≤ Âge < 40 ; Âge3 : 40 ≤ Âge	Années
<i>Supplémentaires</i>				
Nombre de traitements à l'arsénite de soude	Ars	Ars0 ; Ars1 ; Ars2 ; Ars3	Ars0 : 0 traitement ; Ars1 : 1 traitement ; Ars2 : 2 traitements ; Ars3 : 3 traitements	Aucune
Région	-	ALS ; AQT ; BJL ; BRG ; CEN ; DIO ; JUR ; LRO ; PAC ; PCH ; PDL	ALS : Alsace ; AQT : Aquitaine ; BJL : Beaujolais ; BRG : Bourgogne ; CEN : Centre ; DIO : Diois ; JUR : Jura ; LRO : Languedoc-Roussillon ; PAC : PACA ; PCH : Poitou-Charentes ; PDL : Pays de la Loire	Aucune
Cépage	-	AUX ; CAR ; CBF ; CBS ; CHD ; CHE ; CIN ; GAM ; GRE ; GWZ ; MDH ; MEL ; MER ; MPG ; PIN ; PLS ; RIS ; SAU ; SAV ; SYR ; TRS ; UB	AUX : Pinot Auxerrois ; CAR : Carignan ; CBF : Cabernet Franc ; CBS : Cabernet Sauvignon ; CHD : Chardonnay ; CHE : Chenin ; CIN : Cinsault ; GAM : Gamay ; GRE : Grenache ; GWZ : Gewurztraminer ; MDH : Muscat De Hambourg ; MEL : Melon ; MER : Merlot ; MPG : Muscat Pet its Grains ; PIN : Pinot Noir ; PLS : Poulсарd ; RIS : Riesling ; SAU : Sauvignon ; SAV : Savagnin ; SYR : Syrah ; TRS : Trousseau ; UB : Ugni Blanc	Aucune
Porte -greffe	-	101-14; 161 -49; 3309C; 41B; R110; SO4	Porte -greffes correspondants aux catégories	Aucune
Prétaillage	Prét	Prét0 ; Prét1	Prét0 : non, on n'a pas prétaillé ; Prét1 : oui, on a prétaillé	Aucune
Devenir des sarments	-	Broyage ; Brûlage ; Enlèvement	Devenirs des sarments correspondants aux catégories	Aucune
Type de taille	-	Cordon de Royat ; Gobelet ; Guyot	Types de taille correspondants aux catégories	Aucune

Références

- BERTRAND F., MAUMY M., FUSSLER L., KOBES N., SAVARY S. et GROSMAN J. (2008), Étude statistique des données collectées par l'Observatoire National des Maladies du Bois de la Vigne : Annexe de tables et graphiques, disponible en version électronique.
http://www-irma.u-strasbg.fr/~fbertran/recherche/Annexe_MdBVigne.pdf.
- CANTY A. (version originale pour S) et RIPLEY B. (adaptation en R) (2006), *boot* : *Bootstrap R (S-Plus) Functions (Canty)*, R package version 1.2-27.
- CAZES P. (2004), Quelques méthodes d'analyse factorielle d'une série de tableaux de données. *Revue MODULAD*, **31**, 1-31.
- DAVISON A.C. et HINKLEY D.V. (1997), *Bootstrap Methods and their Applications*. Cambridge University Press.
- DRAPER N.R. et SMITH H. (1998), *Applied regression analysis*. 3^{ème} édition, J. Wiley, New York.
- DRAY S., PETTORELLI N. et CHESSEL D. (2003), Multivariate Analysis of Incomplete Mapped Data, *Transactions in GIS*, **7**, 411-422.
- ESCOFIER B. (1979), Traitement simultané de variables quantitatives et qualitatives en analyse factorielle. *Les cahiers de l'analyse des données*, **4** (2), 132-146.
- ESCOFIER B. et PAGÈS J. (1998), *Analyses factorielles simples et multiples*. 3^{ème} édition, Dunod, Paris.
- FUSSLER L. (2006), *Analyse statistique de données de l'Observatoire National des Maladies du Bois*, Université Louis Pasteur, Strasbourg, Mémoire de deuxième année de Master.
- FREEMAN G.H. et HALTON J.H. (1951), Note on an exact treatment of contingency, goodness of fit and other problems of significance. *Biometrika*, **38**, 141-149.
- GOODMAN L.A. et KRUSKAL W.H. (1954), Measures of Association for Cross Classifications. *Journal of the American Statistical Association*, **49**, 732-764.
- HOSMER D.W. (2000), *Applied logistic regression*. 2^{ème} édition, J. Wiley, New York.
- HUSSON F., LÊ S. et MAZET J. (2007), *FactoMineR : Factor Analysis and Data Mining with R*.
<http://factominer.free.fr>, <http://www.agrocampus-rennes.fr/math/>.
- KENDALL M.G. (1938), A new measure of rank correlation. *Biometrika*, **30**, 81-93.
- LAVIT Ch., ESCOUFIER Y., SABATIER R. et TRAISSAC P. (1994), The ACT (STATIS Method). *Computational Statistics and Data Analysis*, **18**, 97-119.
- PAGÈS J. (2004), Analyse factorielle de données mixtes. *Rev. Statistique Appliquée*, **LII** (4), 93-111.
- SAVARY S. et MADDEN L.V. (1995), Use of Categorical Information and Correspondence Analysis in Plant Disease Epidemiology, *Advances in Botanical Research*, **21**, 213-240.
- SOMERS R.H. (1962), A new asymmetric measure of association for ordinal variables. *American Sociological Review*, **27**, 799-811.
- R Development Core Team (2007), *R : A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna.
<http://www.R-project.org>.