



HAL
open science

Étude statistique des données collectées par l'Observatoire National des Maladies du Bois de la Vigne.

Frédéric Bertrand, Myriam Maumy, Lionel Fussler, Nathalie Kobes, Serge
Savary, Jacques Grosman

► **To cite this version:**

Frédéric Bertrand, Myriam Maumy, Lionel Fussler, Nathalie Kobes, Serge Savary, et al.. Étude statistique des données collectées par l'Observatoire National des Maladies du Bois de la Vigne.. 2007. hal-00166929

HAL Id: hal-00166929

<https://hal.science/hal-00166929>

Preprint submitted on 13 Aug 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Étude statistique des données collectées par l'Observatoire National des Maladies du Bois de la Vigne : utilisation du logiciel R.

F. BERTRAND⁽¹⁾, M. MAUMY⁽¹⁾, L. FUSSLER⁽²⁾, N. KOBES⁽²⁾, S. SAVARY⁽³⁾,
J. GROSMAN⁽⁴⁾

⁽¹⁾Laboratoire de Statistique
IRMA ULP (UMR 7501), 67084 Strasbourg Cedex
Email : fbertran@math.u-strasbg.fr
mmaumy@math.u-strasbg.fr

⁽²⁾DRAF-SRPV Alsace
Cité administrative Gaujot, 67084 Strasbourg Cedex
Email : lfussler@wanadoo.fr
nathalie.kobes@agriculture.gouv.fr

⁽³⁾UMR Santé Végétale
Centre INRA de Bordeaux, 33883 Villenave d'Ornon Cedex
Email : ssavary@bordeaux.inra.fr

⁽⁴⁾DRAF-SRPV Rhône
Cité administrative de la Part-Dieu, 69401 Lyon Cedex 03
Email : jacques.grosman@agriculture.gouv.fr

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 objectivement à la question de leur progression éventuelle dans le vignoble français. En effet, suite à l'interdiction de l'utilisation de l'arsénite de soude en traitement d'hiver de la vigne, le 8 novembre 2001, les viticulteurs ne disposent plus d'aucune méthode de lutte chimique « curative » contre les principales maladies du bois de la vigne que sont le black dead arm, l'esca et l'eutypiose. Cet observatoire collecte ainsi, chaque année, depuis sa création en 2003 par le Ministère chargé de l'Agriculture, un ensemble de données cohérentes susceptibles d'être abordées par différentes méthodes d'analyse statistique.

Le jeu de données pour l'étude à réaliser est complexe. En effet, il comporte des variables quantitatives et qualitatives qui évoluent au cours du temps, étant donné que les questionnaires ont été récoltés de 2003 à 2005. La problématique de l'étude est de dégager les grandes tendances en matière d'épidémiologie végétale présentes dans le jeu de données afin de déterminer quelles sont les mesures prophylactiques adaptées à mettre en œuvre collectivement et à grande échelle. Pour aborder le problème, nous avons utilisé des techniques d'analyse de données et des outils de modélisation.

Le travail statistique s'est donc articulé en trois points. D'abord, nous nous sommes efforcés de mettre en évidence des relations entre les différentes variables de l'étude, puis

nous avons utilisé des méthodes factorielles, à savoir l'analyse des correspondances multiples, l'analyse des composantes principales et l'analyse factorielle de données mixtes. Puis, afin de préciser les relations que nous avons décelées, notre choix s'est porté sur la construction et l'évaluation de modèles de régressions logistiques binaires et ordinales pour l'étude desquels nous avons utilisé des techniques bootstrap. Enfin, pour tenir compte du facteur temps, nous avons employé des méthodes d'analyse factorielle de tableaux multiples, chaque tableau représentant une année de données. Les résultats des différentes analyses réalisées seront présentés dans le corps de l'article.

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

ABSTRACT

The Grapevine Wood Diseases National Observatory yields a cohesive and large data set which may be dealt with using different approaches. In our study, we have to deal with complex data, as it is composed of quantitative and qualitative variables which evolve with time, since data for three successive years are available. The objective was to produce of 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. Our study thus developed in three points. 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, binary and ordinal 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, each table representing a year, in order to account for the successive years of data. Those different techniques will be explained here, as well as our results.

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 Europe¹, de l'usage de l'arsénite de soude² 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. L'arsénite de soude est un composé chimique à base d'arsenic dont il n'est plus possible de se servir pour des raisons de santé publique puisque les risques cancérogènes pour les viticulteurs sont importants et avérés. En effet, une étude réalisée par la mutualité sociale agricole (MSA) en 2000 avait montré que les protections utilisées n'étaient pas suffisantes pour assurer une bonne protection des utilisateurs³. 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

¹ Le règlement 2076/2002/CE du 20 novembre 2002 impose aux États membres de l'Union Européenne d'interdire l'utilisation des préparations à base d'arsénite de soude au plus tard le 31 décembre 2003.

² Cette substance chimique est également appelée arsénite de sodium.

³ 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

conséquence ultime est bien souvent la mort des plants⁴. 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 a été constitué en 2001. C'est dans le cadre de ce groupe de travail national que la création d'un observatoire pluriannuel, l'Observatoire National des Maladies du Bois de la Vigne a été décidée. Cette structure est unique au monde du fait de la taille du dispositif et du nombre de partenaires impliqués. Ce genre d'étude n'est en effet réalisé généralement que sur une région ou sur un cépage, alors qu'ici l'ensemble des régions viticoles françaises est considéré. D'une durée de six années, de 2003 à 2009, ce dispositif a pour objectif⁵ 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.

1.2 Protocole technique de l'observatoire

Le protocole technique a été défini par la note de service DGAL/SDQPV/N2003-8085 du 19 mai 2003⁶. Nous en rappelons ici les principales caractéristiques :

- Toutes les régions viticoles⁷ françaises présentent au moins l'une des trois maladies, ce qui justifie qu'elles participent à l'enquête. En fait, 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 les régions viticoles dont la superficie recouverte par des plants de vigne est la plus grande. En effet à elles seules, elles représentent 95% de la superficie totale viticole en France, voir le Tableau A de l'Annexe 6.4 à ce sujet. Il est à noter que la Corse n'a pas été enquêtée. En effet, par son caractère insulaire, elle a sa propre spécificité.
- Un minimum de 25 parcelles par vignoble et par cépage devait être observé, pour faire un état des lieux satisfaisant de la répartition, de la fréquence et de l'intensité de ces maladies pour un cépage donné, dans un vignoble donné. 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, répartis en dix placettes de 30 ceps choisies également aléatoirement. Ce processus est préférable à une observation sur le rang entier pour une question de représentativité au sein de la parcelle. 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.

⁴ Un certain nombre de termes techniques propres au vocabulaire de l'épidémiologie végétale ou de la vigne ont été définis dans le lexique se trouvant à l'Annexe 6.2.

⁵ La note technique définissant ces objectifs est disponible à l'adresse électronique suivante : <http://www.agriculture.gouv.fr/spip/IMG/pdf/dgaln20048126z.pdf>.

⁶ Cette note technique est disponible à l'adresse électronique suivante : <http://www.agriculture.gouv.fr/spip/IMG/pdf/dgaln20038085.pdf>.

⁷ La Figure B de l'Annexe 6.4 situe les régions viticoles enquêtées sur la carte de France.

- Les trois maladies sont observées chaque année à deux périodes particulières du stade de la vigne (floraison pour l'eutypiose, fin mai-début juin, véraison pour le black dead arm et l'esca, fin août). 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 proposé par le Ministère de l'Agriculture puis validé par le groupe technique national auquel l'ITV⁸ et l'INRA⁹ appartiennent. Il reprend les ordres de grandeur des nombres d'observations qui sont réalisées habituellement lors des expérimentations du même type.

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. Le lecteur trouvera une reproduction du questionnaire ayant servi à la récolte des données à la Figure C de l'Annexe 6.5.

1.3 Mise en œuvre et plan de l'étude statistique

L'esca et le black dead arm (BDA) ayant les mêmes symptômes foliaires apparents¹⁰, 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 2005¹¹, concernant les deux maladies du bois de la vigne qui sont donc l'eutypiose et l'esca/BDA. Pour expliquer les variations des intensités observées de ces 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). D'autres données concernant la vigueur de la vigne, le type de sol ou encore la superficie de la parcelle ont été également rassemblées. Cependant le manque de réponses ou de fiabilité de celles-ci nous a amené à ne pas les prendre en compte dans le cadre de cette étude.

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.

⁸ Institut Technique de la Vigne et du Vin, www.itvfrance.com.

⁹ Institut National de la Recherche Agronomique, www.inra.fr.

¹⁰ Voir la Figure A de l'Annexe 6.3 pour une appréciation visuelle de cette ressemblance.

¹¹ À l'instant où les auteurs ont rédigé l'analyse statistique, les données de 2006 n'avaient pas encore été dépouillées.

Au total, nous disposons d'un jeu de données représentant 701 parcelles observées dans les 12 régions viticoles citées précédemment, correspondant à 26 cépages^{12,13}, 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 à eux seuls que 1,28% du nombre total des parcelles. Nous avons alors éliminé les parcelles présentant des données manquantes ou aberrantes¹⁴ et conservé les 254 parcelles pour lesquelles toutes les informations étaient valides. 1 La région Midi-Pyrénées devait être enquêtée en 2003 et 2004 pourtant elle n'a renvoyé des résultats satisfaisants qu'à partir de 2005 et de ce fait ne pouvait être intégrée à l'analyse nationale que nous avons réalisée pour la période de 2003 à 2005.

La nature des questions soulevées par la problématique qui a abouti à la création de l'Observatoire des Maladies du Bois de la Vigne a nécessité l'utilisation de plusieurs méthodes statistiques. Ainsi, 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'analyses factorielles sur le tableau moyen, défini par la suite, des trois années 2003 à 2005. Il s'agit de l'analyse des correspondances multiples, l'analyse des composantes principales et de l'analyse factorielle de données mixtes. Dans un second temps nous avons eu recours à des outils de modélisation. Notre choix, compte tenu de la nature des réponses observées, s'est porté sur la 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 des maladies du bois de la vigne¹⁵, 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 qui a été effectuée 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 notamment dans l'expression des symptômes de l'esca/BDA. Ceci nous a conduits à identifier quatre principales classes (voir le Tableau B de l'Annexe 6.6) découlant de la relation entre l'âge de la parcelle et les trois variables que nous voulons expliquer. Ces relations diffèrent substantiellement d'une classe à l'autre avec même la présence de signes opposés dans l'expression des valeurs des coefficients de corrélation entre l'âge de la parcelle et les incidences des deux maladies ainsi que le taux de mortalité¹⁶.

Ensuite, nous avons effectué de nombreux tests non paramétriques de Kruskal-Wallis, exposé dans le livre de Siegel et de Castellan (1988). Les procédures de comparaisons multiples réalisées après ces tests entre les variables de réponse et les facteurs explicatifs

¹² 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 Poulard, le Riesling, le Sauvignon, le Savagnin, le Syrah, le Trousseau et l'Ugni Blanc.

¹³ La Figure B de l'Annexe 6.4 situe ces cépages sur la carte de France.

¹⁴ C'est l'une des deux procédures recommandée lorsqu'il y a des données manquantes. À ce sujet, nous renvoyons à l'article de Dray, Pettoelli et Chessel intitulé « Multivariate Analysis of Incomplete Mapped Data » dans *Transactions in GIS* en 2003, 7, p : 411-422. L'application des techniques exposées dans cet article qui se basent sur l'algorithme NIPALS fera prochainement l'objet d'un travail de la part des auteurs.

¹⁵ Nous rappelons que l'année 2003 a été marquée par un été caniculaire.

¹⁶ Ceci a été l'objet du rapport de Lionel Fussler dans le cadre de sa deuxième année de Master de Statistique.

comme le cépage ou la région nous ont permis, par exemple, de définir des classes de sensibilité aux maladies pour les différents cépages. Ces résultats ont confirmé ceux qui étaient connus jusqu'alors tout en apportant un important complément d'information.

Consécutivement, non seulement aux tests non paramétriques de Kruskal-Wallis mais aussi aux multiples tests bivariés comme le test du χ^2 d'indépendance détaillé dans le livre d'Agresti (1990) et le test exact de Fisher-Freeman-Halton d'indépendance, présenté dans l'article de Freeman et de Halton (1951), que nous avons réalisés, nous avons été à même de réduire le nombre de variables, de 15 à 8, 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 ont été retenus pour la suite de l'étude. 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 Références

Nous renvoyons, par exemple, aux livres de Bry (1995, 1996) et de Escofier et Pagès (1998) pour un exposé de l'analyse des correspondances multiples.

2.1.2 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 même qualitatives, si nous recodons les variables quantitatives, et ne nécessite pas 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 ne sont pas réalisées sans raison mais suivent le protocole recommandé par Savary (1995). Elles sont récapitulées dans le Tableau B de l'Annexe 6.6.

Dans un premier temps, une analyse des correspondances multiples 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,
- l'âge de la parcelle.

Cette analyse des correspondances multiples a pour premier but de répondre à un des objectifs principaux : caractériser l'évolution de la 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, et de la diminution importante du pourcentage de l'inertie, passage de 34,953% à 13,180%, associé au premier plan factoriel qui résultait de son introduction comme variable active. De même la variable « Région » n'a pas pu être conservée car elle entraînait une réduction non négligeable du pourcentage de l'inertie du premier plan factoriel, baisse de 34,953% à 20,601%. Notons aussi que le nombre de modalités est important, 11 régions viticoles, par rapport aux nombres des autres variables retenues. 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.

Axe	Valeur propre singulière	Pourcentage de l'inertie	
		Individuel	Cumulé
1	0,652	18,904	18,904
2	0,601	16,049	34,953
3	0,549	13,372	48,326
4	0,514	11,720	60,046
5	0,487	10,536	70,582
6	0,454	9,164	79,745
7	0,435	8,411	88,156
8	0,415	7,661	95,818
9	0,307	4,182	100,000

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

Les Tableaux 1, 2 et 3 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. 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.

	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

TABLEAU 2 (Haut) et 3 (Bas)

Coordonnées, contributions et cosinus carrés associés aux modalités des différentes variables pour les quatre premiers axes factoriels de l'ACM

	Masse	Qualité	Inertie Relative	Coord.	Inertie	Cosinus ²	Coord.	Inertie	Cosinus ²
				Dim. 3	Dim. 3	Dim. 3	Dim. 4	Dim. 4	Dim. 4
Euty0	0,086	0,657	0,073	-0,263	0,020	0,036	-0,103	0,003	0,006
Euty1	0,103	0,587	0,065	0,124	0,005	0,011	0,667	0,174	0,314
Euty2	0,061	0,756	0,084	0,159	0,005	0,008	-0,985	0,225	0,313
Esca0	0,013	0,949	0,105	2,489	0,263	0,334	0,981	0,047	0,052
Esca1	0,143	0,429	0,048	-0,091	0,004	0,011	-0,584	0,185	0,454
Esca2	0,094	0,622	0,069	-0,200	0,013	0,024	0,749	0,201	0,341
Mort0	0,030	0,882	0,098	1,226	0,147	0,201	-0,913	0,093	0,112
Mort1	0,098	0,606	0,067	-0,653	0,140	0,277	0,119	0,005	0,009
Mort2	0,122	0,512	0,057	0,230	0,021	0,051	0,125	0,007	0,015
Âge0	0,033	0,866	0,096	0,080	0,001	0,001	0,217	0,006	0,007
Âge1	0,105	0,579	0,064	-0,613	0,131	0,273	-0,238	0,023	0,041
Âge2	0,087	0,654	0,073	0,879	0,222	0,409	0,045	0,001	0,001
Âge3	0,025	0,902	0,100	-0,579	0,027	0,037	0,566	0,030	0,035

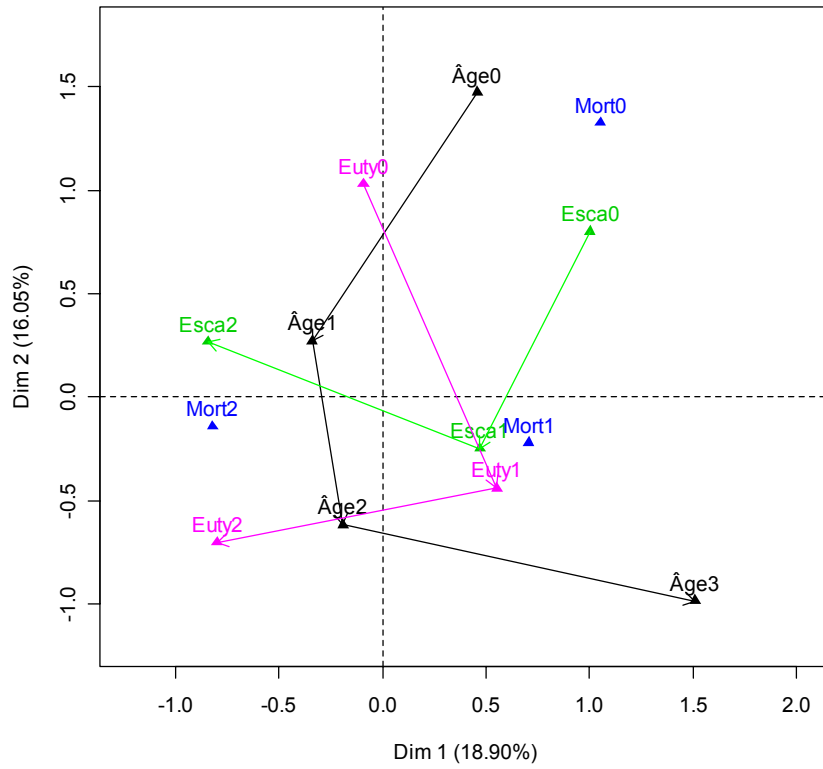


FIGURE 1
Représentation graphique des deux premiers axes factoriels de l'ACM

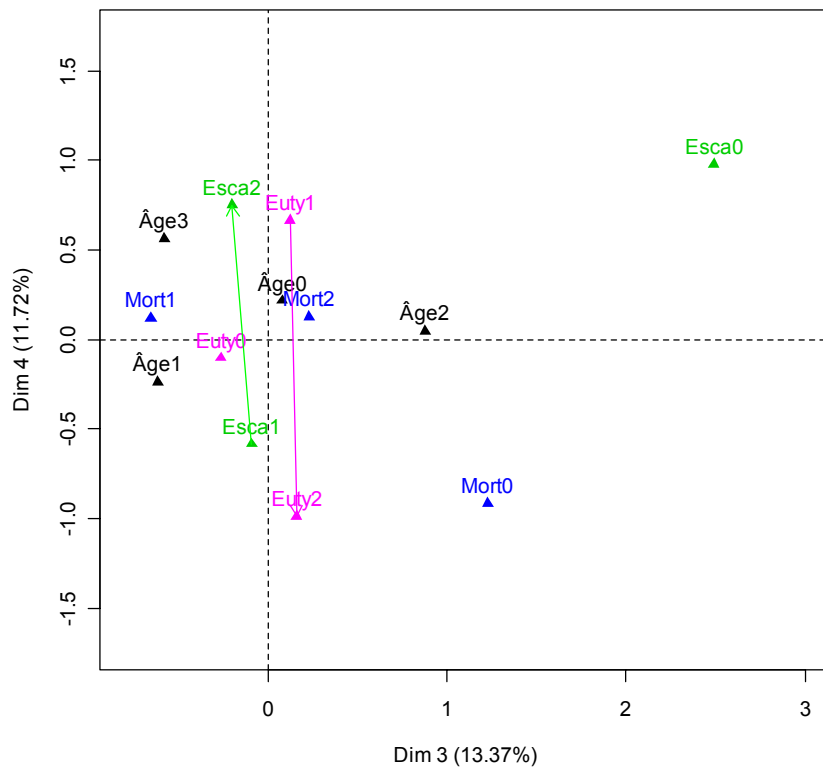


FIGURE 2
Représentation graphique des troisième et quatrième axes factoriels de l'ACM

Les analyses ci-dessous constituent un canevas sur lequel nous ajoutons des variables projetées en supplémentaires. Celles-ci ne contribuent donc pas à l'élaboration des axes, nous ne présenterons donc dans le Tableau 4 que leurs coordonnées et les valeurs des cosinus carrés dans le premier plan factoriel. Seuls les graphiques les plus informatifs sont reproduits ci-après.

Variables et leurs modalités	1 ^{er} axe factoriel		2 nd axe factoriel	
	Coordonnées	Cosinus ²	Coordonnées	Cosinus ²
Région				
ALS	-0,098	0,007	0,110	0,009
AQT	0,080	0,002	0,522	0,089
BJL	0,047	0,000	-0,818	0,136
BRG	0,712	0,251	-0,429	0,091
CEN	0,180	0,017	-0,348	0,062
DIO	0,622	0,148	-0,194	0,014
JUR	-0,512	0,127	0,487	0,115
LRO	-0,208	0,004	-0,034	0,000
PAC	-0,132	0,014	0,363	0,106
PCH	-1,024	0,231	-0,899	0,178
PDL	0,027	0,000	0,735	0,192
Cépage				
AUX	-0,269	0,039	-0,474	0,123
CAR	0,376	0,018	0,150	0,003
CBF	0,887	0,207	-0,347	0,032
CBS	-0,755	0,133	-0,241	0,013
CHD	0,672	0,200	-0,245	0,027
CHE	-0,124	0,002	-0,330	0,015
CIN	-0,769	0,075	-0,644	0,052
GAM	0,047	0,000	-0,818	0,136
GRE	0,505	0,038	0,083	0,001
GWZ	-0,103	0,006	0,351	0,068
MDH	-0,020	0,000	0,595	0,129
MEL	0,067	0,001	0,901	0,221
MER	0,995	0,075	1,106	0,093
MPG	0,641	0,167	-0,052	0,001
PIN	1,345	0,454	-0,646	0,105
PLS	-0,702	0,170	0,455	0,071
RIS	0,103	0,005	0,505	0,114
SAU	-0,518	0,084	-0,114	0,004
SAV	-0,678	0,146	0,687	0,150
SYR	-0,208	0,004	-0,034	0,000
TRS	-0,138	0,008	0,216	0,021
UB	-1,024	0,231	-0,899	0,178

Variables et leurs modalités	1 ^{er} axe factoriel		2 nd axe factoriel	
	Coordonnées	Cosinus ²	Coordonnées	Cosinus ²
Arsénite de soude				
Ars0	-0,042	0,002	0,069	0,005
Ars1	0,197	0,030	-0,360	0,101
Ars2	0,038	0,001	-0,009	0,000
Ars3	-0,153	0,007	0,299	0,027
Porte-greffe				
101-14	0,006	0,000	-0,252	0,027
161-49	0,027	0,001	-0,052	0,002
3309C	0,086	0,006	0,127	0,014
41B	-0,465	0,121	-0,111	0,007
R110	0,119	0,003	0,580	0,074
SO4	-0,037	0,001	-0,152	0,022
Prétaillage				
Prét0	-0,097	0,010	0,020	0,000
Prét1	0,121	0,015	-0,025	0,001
Devenir des sarments				
Broyage	-0,156	0,026	0,060	0,004
Brûlage	0,586	0,211	-0,287	0,050
Enlèvement	0,207	0,014	0,141	0,007
Type de taille				
Cordon de Royat	-0,400	0,104	0,106	0,007
Gobelet	-0,002	0,000	-0,274	0,012
Guyot	0,024	0,001	-0,002	0,000

TABLEAU 4
Coordonnées et cosinus carrés des modalités des variables mises en supplémentaires dans l'ACM

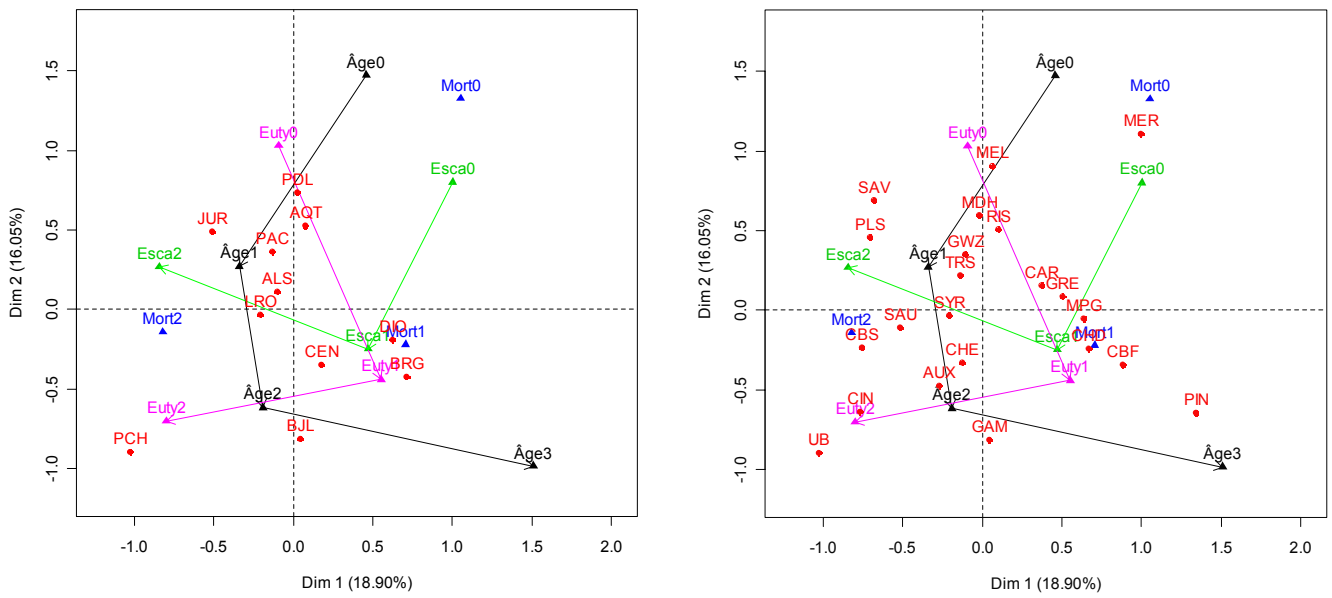


FIGURE 3
Représentations graphiques des deux premiers axes factoriels des ACM avec respectivement la région puis le cépage mis en supplémentaire

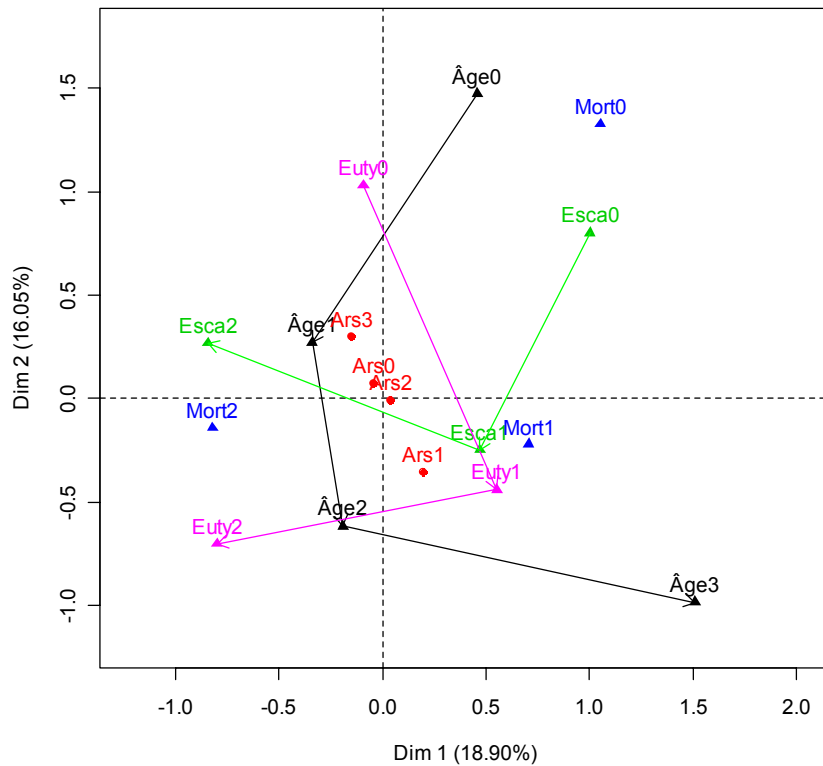


FIGURE 4
Représentation graphique des deux premiers axes factoriels de l'ACM avec l'arsénite de soude mis en supplémentaire

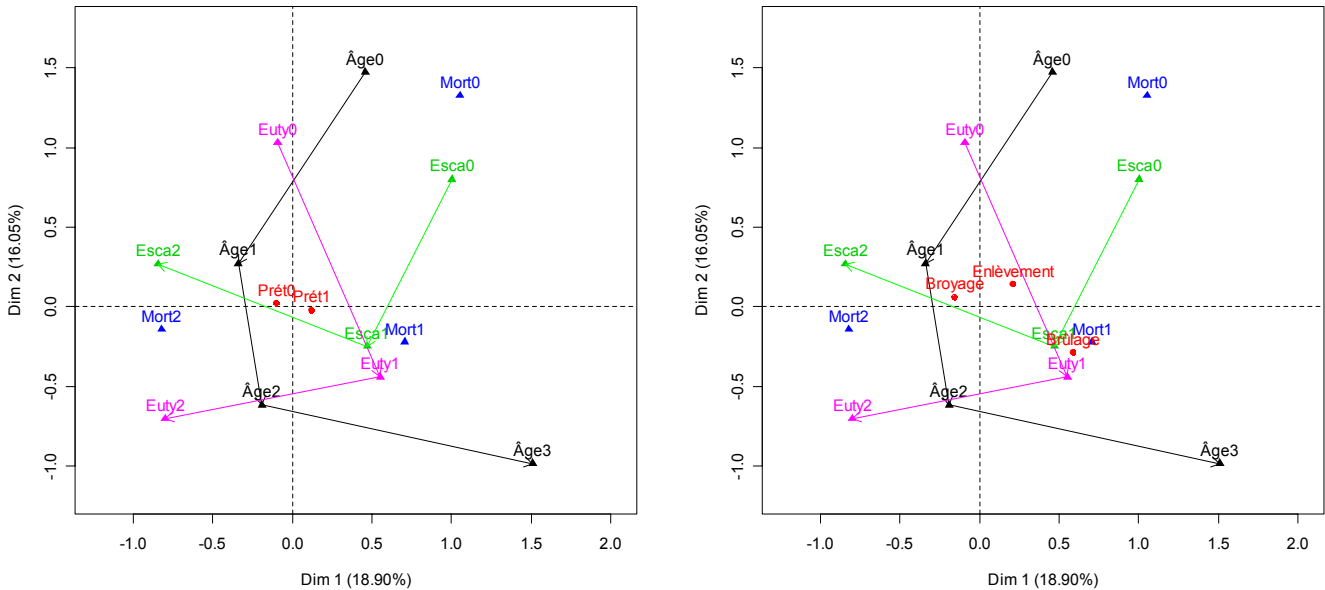


FIGURE 5

Représentations graphiques des deux premiers axes factoriels des ACM avec respectivement le prétaillage puis le devenir des sarments mis en supplémentaire

2.1.3 Bilan

Nous reportons ci-dessous quelques interprétations parmi les principales des résultats des Tableaux 1, 2, 3 et 4 ainsi que des Figures 1, 2, 3, 4 et 5.

- Les niveaux de mortalité élevés sont associés à des incidences fortes des deux maladies (cf. Figure1).
- Les fortes incidences d'eutypiose et d'esca/BDA sont respectivement associées à des vignes relativement âgées (25-40 ans) et des vignes jeunes (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, à gauche) ou encore Poulsard (« PLS ») et esca/BDA 2 (« Esca2 ») (cf. Figure 3, à droite).
- 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, à droite). En effet cette technique est essentiellement utilisée en Bourgogne.

2.2 Analyse en composantes principales (ACP)

2.2.1 Références

Nous renvoyons, par exemple, aux livres de Bry (1995, 1996) et de Escofier et Pagès (1998) pour un exposé de l'analyse en composantes principales.

2.2.2 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'eutypiose, l'incidence de l'esca/BDA, le taux de mortalité des plants et l'âge de la parcelle.

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.

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, 8, 9, 10 et 11, afin de les comparer avec ceux produits par la technique 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 ellipses de confiance au seuil de 95%, obtenues par des techniques bootstrap, ont été intégrées aux deux représentations graphiques des individus dans le premier plan factoriel. Sur le premier de ces deux graphiques, situé en bas à gauche, les ellipses de confiance représentent des zones associées aux individus présentant la même modalité d'une variable qualitative mise en supplémentaire. Les secondes concernent les barycentres de ces groupes d'individus c'est-à-dire les modalités des variables qualitatives mises en supplémentaires. Enfin, il est à noter que nous nous sommes limités au premier plan factoriel car nous obtenons ainsi une inertie environ égale à 72% pour conserver environ le même ordre de grandeur que celui qui avait été considéré lors des ACM.

Axe	Valeur propre singulière	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

TABLEAU 5
Valeurs propres et inerties associées aux axes factoriels de l'ACP

	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

TABLEAU 6
Coordonnées, contributions et cosinus carrés associés aux modalités des différentes variables pour les deux premiers axes factoriels de l'ACP

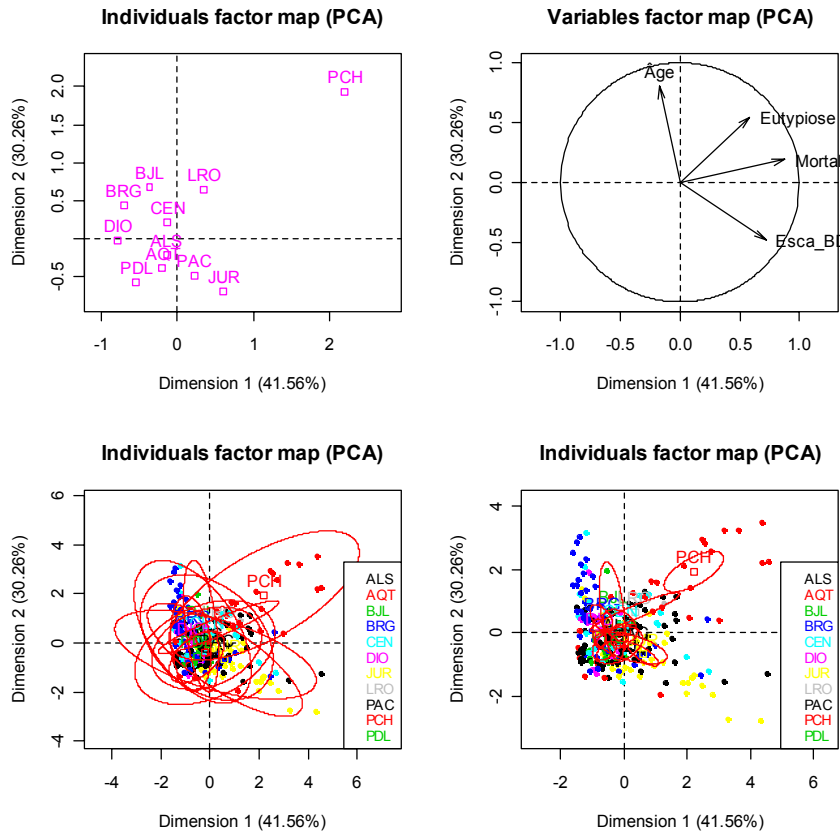
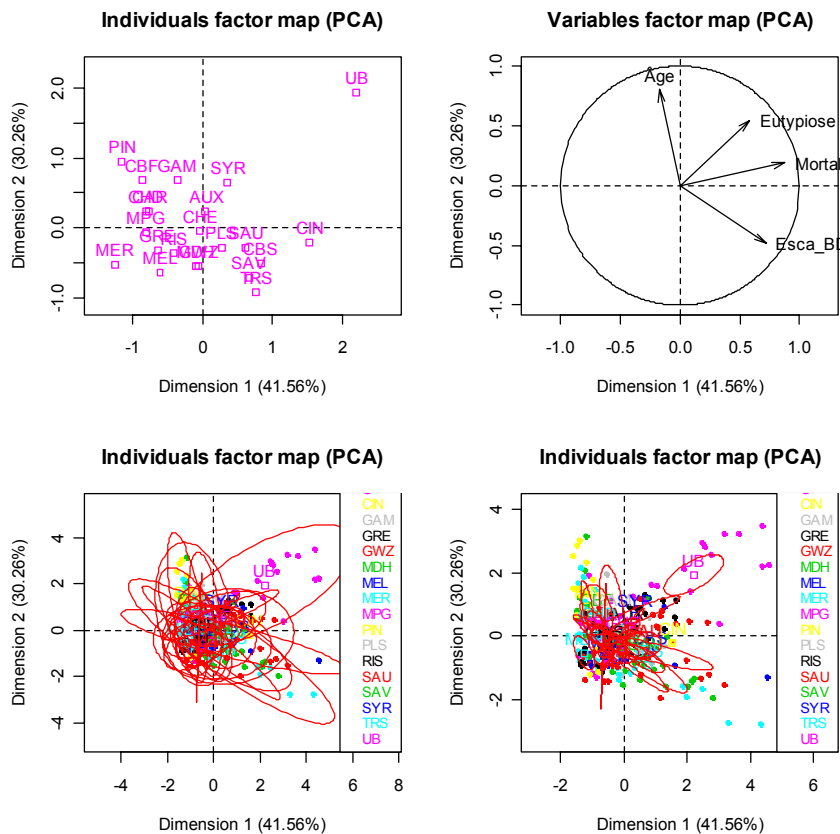


FIGURE 6 (Haut) et 7 (Bas)

Représentations graphiques des deux premiers axes factoriels des analyses en composantes principales avec pour variables qualitatives la région et le cépage



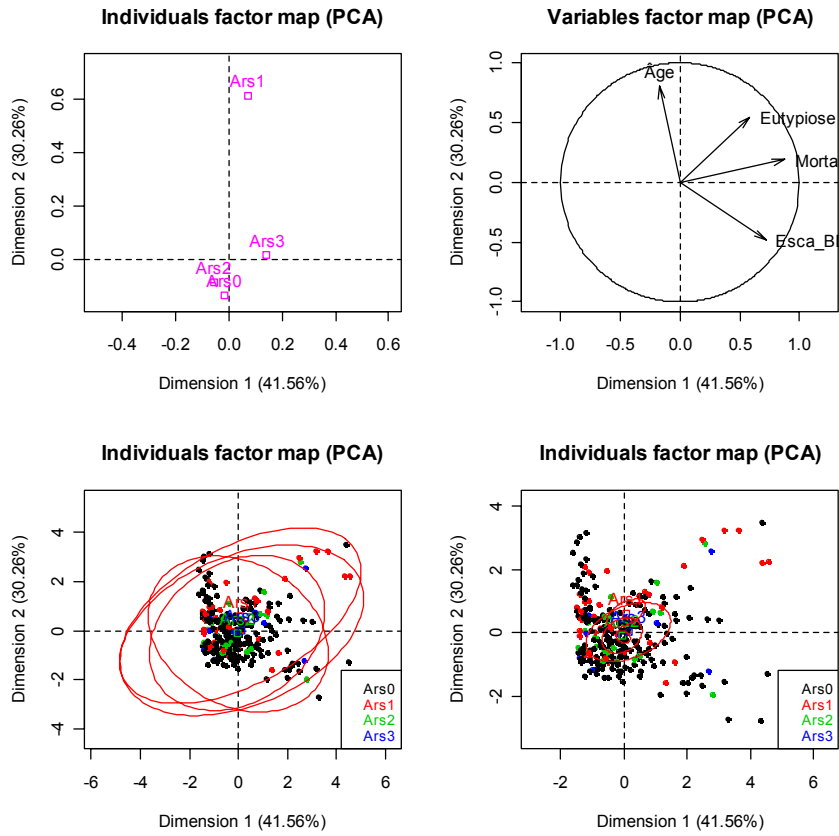
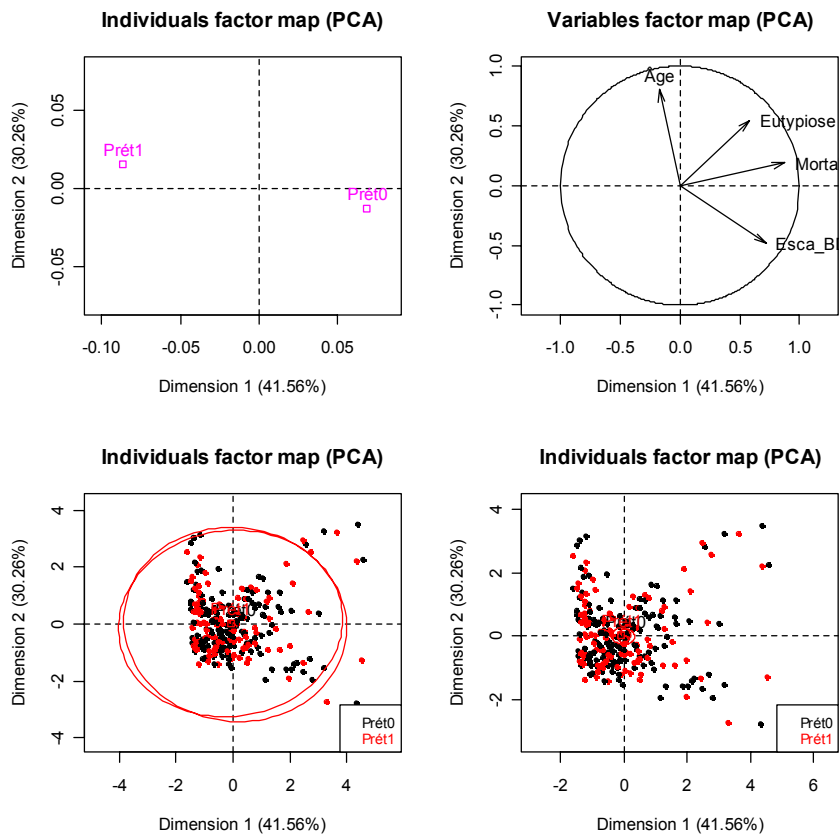


FIGURE 8 (Haut) et 9 (Bas)
Représentations graphiques des deux premiers axes factoriels des analyses en composantes principales avec pour variables qualitatives l'arsénite de soude et le prétaillage



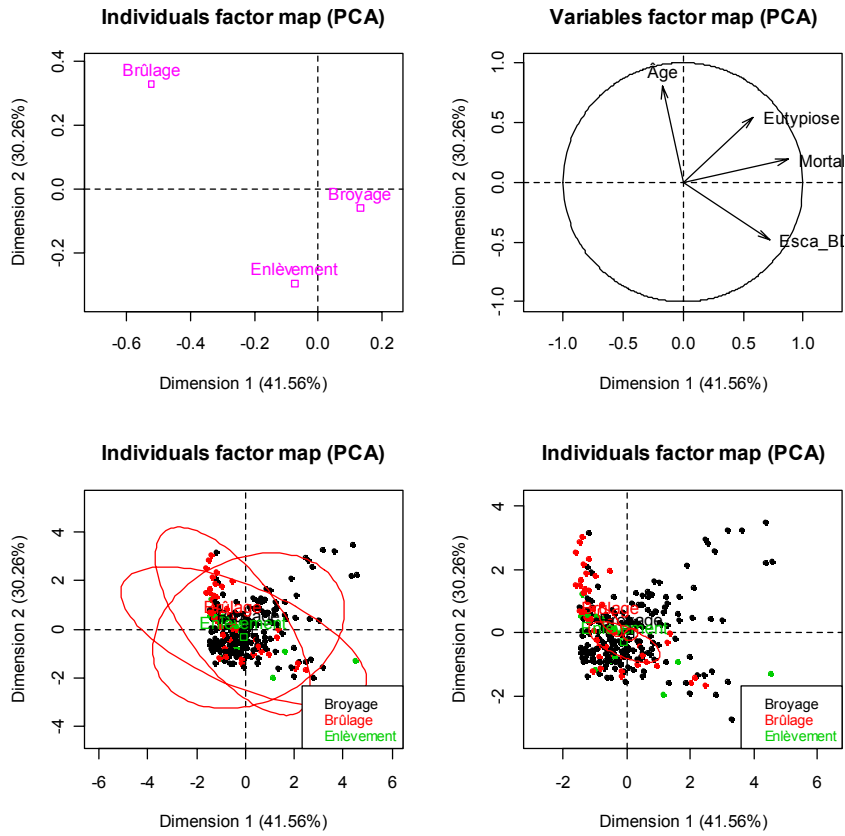
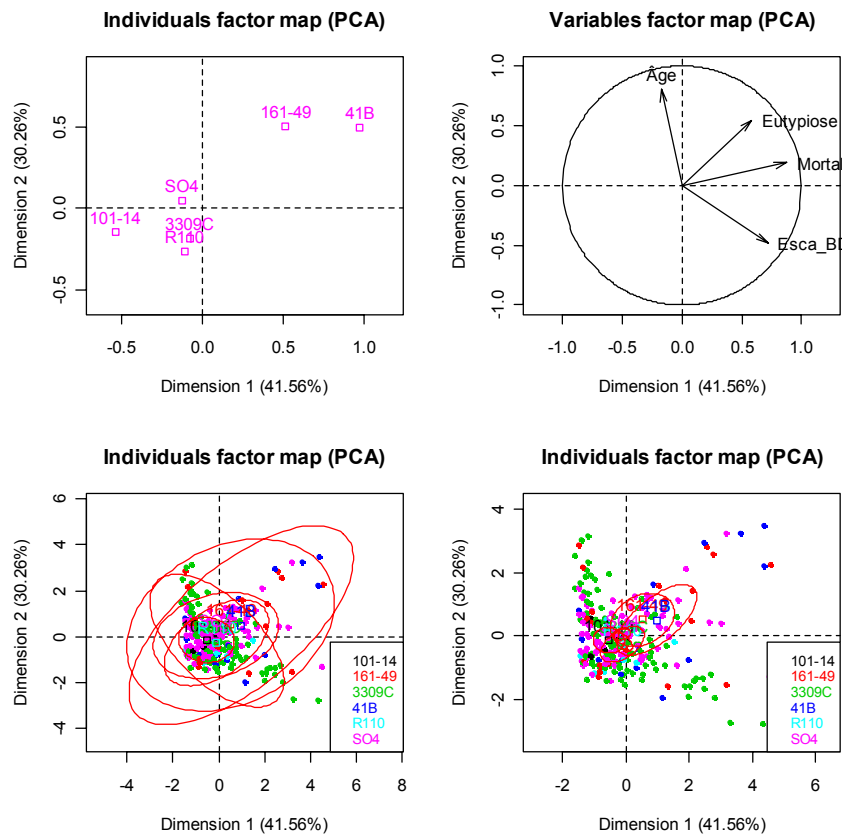


FIGURE 10 (Haut) et 11 (Bas)

Représentations graphiques des deux premiers axes factoriels des analyses en composantes principales avec pour variables qualitatives le devenir des sarments et le porte-greffe



2.2.3 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 des 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 toujours porté par le deuxième axe factoriel.
- Nous remarquons l'association entre la région Poitou-Charentes («PCH») et l'eutypiose («Eutypiose») (cf. Figure 6, les deux en haut) puis l'association entre le cépage Ugni Blanc («UB») et l'eutypiose («Eutypiose») (cf. Figure 7, les deux en haut), conformément à ce que nous avons remarqué lors des ACM. Ce qui se justifie par le fait que l'Ugni Blanc est caractéristique du Poitou-Charentes.
- De même nous notons l'association entre la région Jura («JUR») et l'esca/BDA («Esca_BDA») (cf. Figure 6, les deux en haut) 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, les deux en haut). Ce qui se justifie encore par le fait que les trois cépages sont caractéristiques du Jura.
- Il semble qu'il existe une opposition entre le taux de mortalité des plants et le fait de ne pas avoir traité ou d'avoir effectué deux traitements à l'arsénite de soude (cf. Figure 8, les deux 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 9, les deux en haut). Cette observation se répète également avec le brûlage des sarments («Brûlage») (cf. Figure 10, les deux en haut).
- Enfin, certains types de porte-greffes, comme le «161-49» et le «41B» apparaissent comme plus vulnérables à l'eutypiose («Eutypiose») (cf. Figure 11, les deux en haut).

2.3 Analyse factorielle de données mixtes (AFDM)

Au paragraphe 2.1.2, 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.

Afin de rester cohérent avec la mise en œuvre des ACM, les analyses de données mixtes 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.

2.3.1 Références

Nous renvoyons, par exemple, à l'article de Pagès (2004) pour un exposé de l'analyse factorielle de données mixtes qui développe une idée exposée dans l'article d'Escofier (1979).

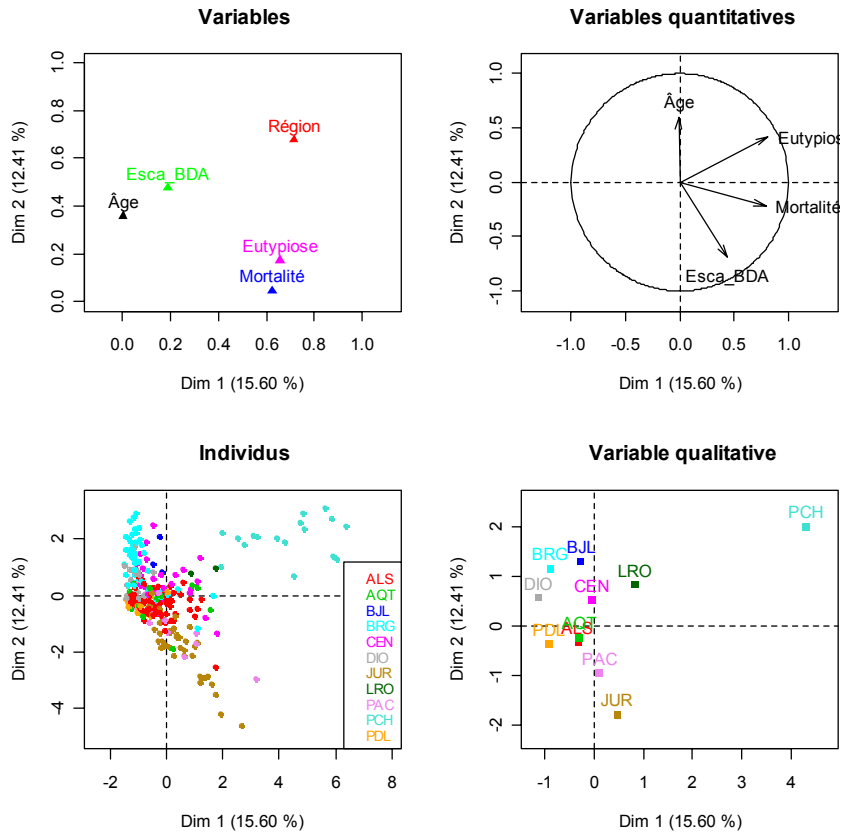
2.3.2 Application aux données de l'Observatoire National des Maladies du Bois de la Vigne

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.

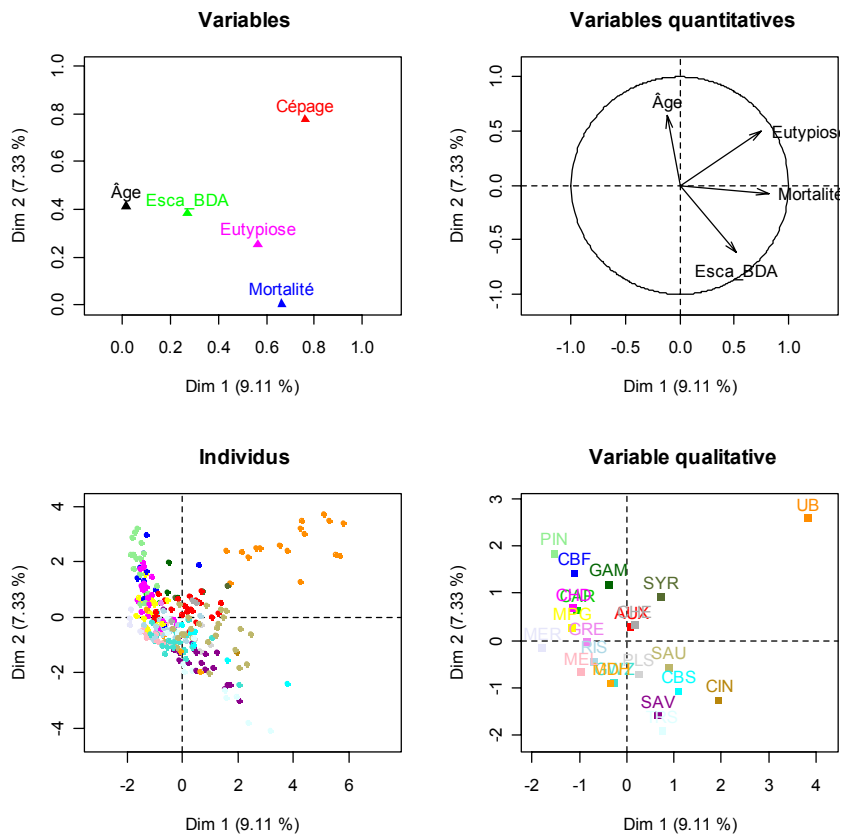
2.3.3 Bilan

Nous proposons ci-après quelques éléments d'interprétation des résultats des Figures 12, 13, 14, 15, 16 et 17.

- Les deux axes factoriels expliquent fortement la variabilité de la variable « Région » compte tenu des valeurs du rapport corrélation, graphique en haut à gauche Figure 12.
- 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 les ACM et les ACP.
- Les deux maladies semblent être opposées, ce qui corrobore les résultats des ACM et des 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 des ACM et des ACP.
- Toujours en accord avec les ACM et les 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.



FIGURES 12 (Haut) et 13 (Bas)
 Représentations graphiques des deux premiers axes factoriels des analyses factorielles de données mixtes avec la région ou le cépage comme variable qualitative



- 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 corrélation, graphique en haut à gauche Figure 13.
- En ce qui concerne les relations entre les cépages et les maladies, le Poulсарd, le Savagnin et le Trousseau sont proches de l'esca/BDA, comme lors des ACM et des ACP où ces cépages étaient associés à de fortes incidences de l'esca/BDA.
- Les autres liaisons établies lors des ACM sont à nouveau moins marquées et ce selon toute vraisemblance pour les mêmes raisons que celles évoquées précédemment.

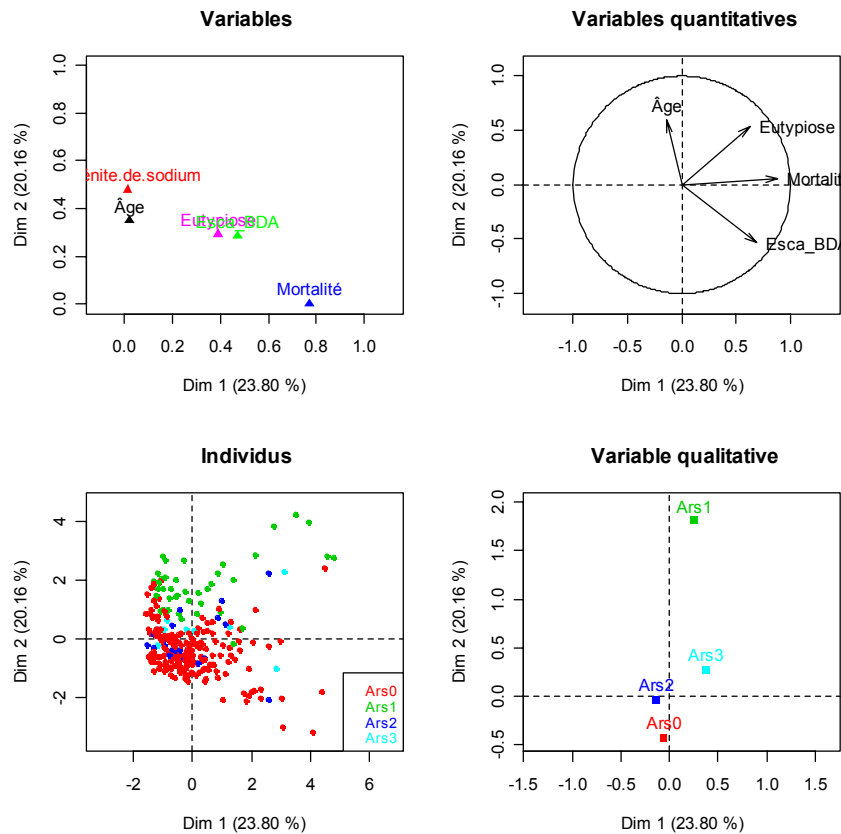


FIGURE 14

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

- La relation maladies/mortalité reste remarquable tout comme le positionnement de l'âge de la parcelle sur le deuxième axe factoriel.
- Nous remarquons, comme lors des ACP, une opposition entre l'augmentation du taux de mortalité des plants et le fait de ne pas avoir traité ou d'avoir effectué deux traitements à l'arsénite de soude, ainsi que la proximité entre les incidences fortes de l'eutypiose et le fait d'avoir traité les vignes à trois reprises avec de l'arsénite de soude.
- Nous notons néanmoins que les parcelles traitées à trois reprises sur cette période sont peu nombreuses et que l'efficacité de l'arsénite de soude vis-à-vis de l'eutypiose n'a jamais été prouvée. Ce résultat est donc à confirmer.

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 des 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 des 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'indique les faibles valeurs du rapport de corrélation du graphique en haut à gauche de la Figure 15.

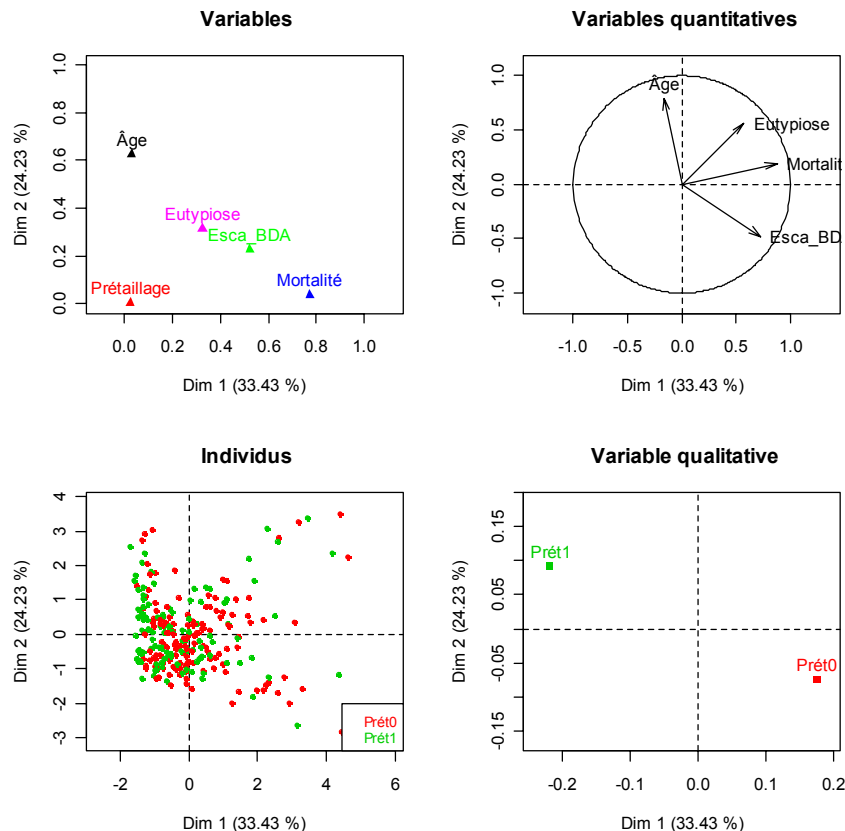


FIGURE 15

Représentations graphiques des deux premiers axes factoriels des analyses factorielles de données mixtes avec le prétaillage comme variable qualitative

- Le brûlage des sarments s'oppose aux maladies ainsi qu'au taux de mortalité.
- En ce qui concerne l'enlèvement des sarments, il est plus délicat d'arriver à des conclusions étant donné la faiblesse des effectifs (cf. Figure 16). En effet cette pratique n'a été observée qu'à sept reprises sur l'échantillon étudié qui concerne 191 parcelles.

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 17), conformément à ce que nous avons pu observer lors des ACP.

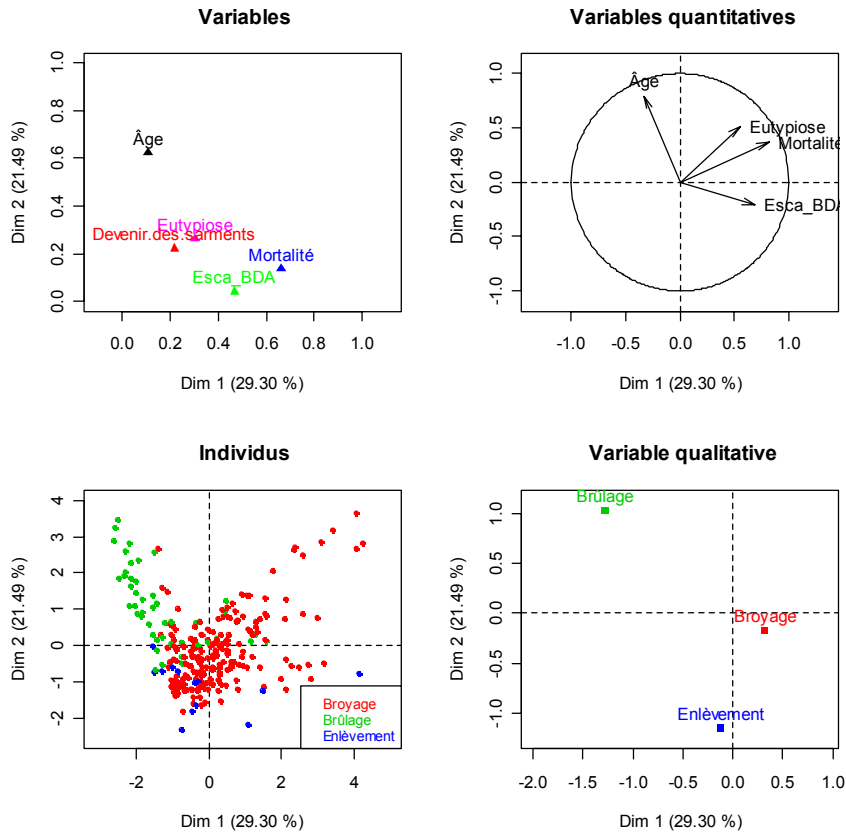
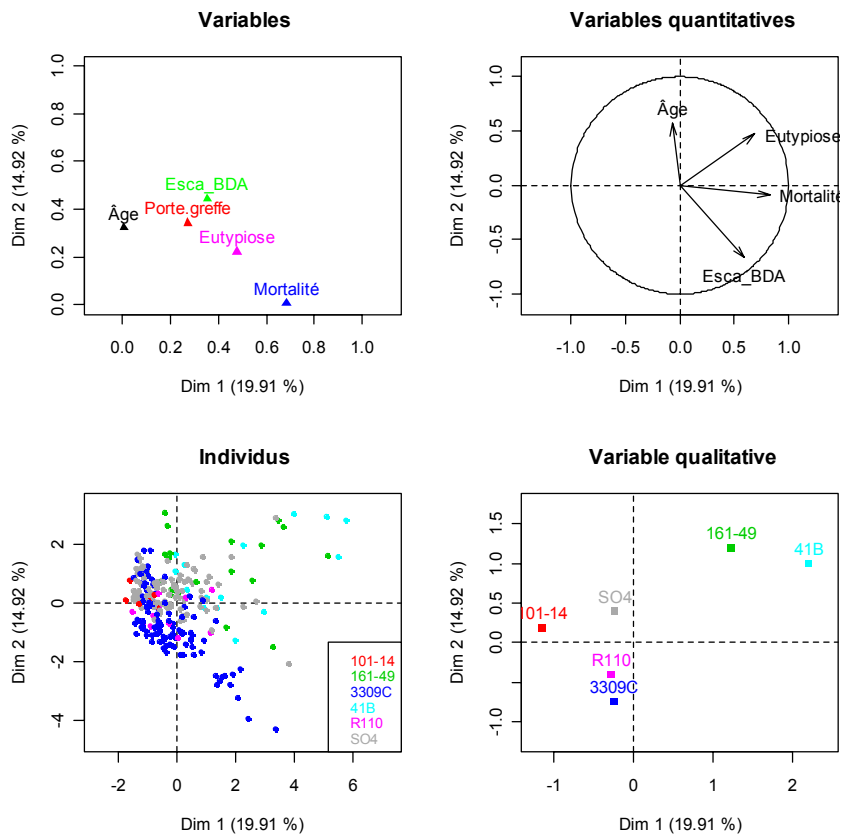


FIGURE 16 (Haut) et 17 (Bas)

Représentations graphiques des deux premiers axes factoriels des analyses factorielles de données mixtes avec le devenir des sarments ou le porte-greffe comme variables qualitative



3 Modèles de régressions logistiques et bootstrap

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 7 en fait état.

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%

TABLEAU 7

Descriptif des classes des maladies et du taux de mortalité

Ainsi, là aussi plusieurs modèles de régressions logistiques ont été testés. Tout d'abord, ont été réalisées les régressions portant sur les liens mortalité-maladies-âge de la parcelle-arsénite de soude, eutypiose-âge de la parcelle-arsénite de soude et esca/BDA-âge de la parcelle-arsénite de soude indépendamment du taux de mortalité des plants et des incidences des autres maladies à savoir respectivement esca/BDA et eutypiose, et ce que ce soit pour le stade d'apparition et d'intensification. Ensuite, pour prendre en compte les autres variables ainsi que leurs associations et d'éviter les problèmes de confusions d'effets, ces mêmes modèles ont été repris mais en utilisant la technique de régression emboîtée avec comme facteurs principaux la région de la parcelle et/ou le cépage.

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 :

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

où α est une constante associée au modèle, β_i les paramètres à estimer, X_i les covariables ou cofacteurs, définies au paragraphe 3.1.3, et logit la fonction définie par :

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

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 :

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

où α_j est une constante associée au modèle, β_i les paramètres à estimer et X_i les covariables ou cofacteurs qui seront précisés au paragraphe 3.1.3.

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.

Ainsi, un modèle emboîté est un modèle qui fait intervenir des cofacteurs imbriqués les uns dans les autres suivant une hiérarchisation. Dans notre étude, un exemple d'emboîtement est d'imbriquer le cépage dans la région ou inversement. En effet, nous savons que certains cépages sont spécifiques de certaines régions : par exemple, le Riesling ou le Gewurztraminer sont des cépages typiquement alsaciens. Il convient donc de ne pas les dissocier. Pour dénoter les emboîtements nous mettrons le cofacteur imbriqué 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 considérons différents critères de bon classement.

Toutes les paires d'individus observés, ici les parcelles, dont les variables réponses (Y_i) sont différentes sont considérées. Alors, trois options se présentent

- **Une paire est concordante** si la valeur prédite pour l'individu dont la probabilité est la plus grande est supérieure à la valeur prédite pour l'individu dont la probabilité est la plus petite.

- **Une paire est discordante**, dans le cas contraire, c'est-à-dire si la valeur prédite pour l'individu dont la probabilité est la plus grande est inférieure à la valeur prédite pour l'individu dont la probabilité est la plus petite.
- **Une paire est déclarée ex æquo** si elle n'est ni concordante ni discordante.

Les indices de corrélation que sont le D de Somers (Somers 1962), le γ de Goodman et Kruskal (Goodman et Kruskal 1954) et le τ_a de Kendall (Kendall 1938) sont calculés à partir du nombre de paires concordantes et discordantes. Ils sont utilisés pour confirmer les prédictions d'un modèle. Leurs définitions respectives sont :

$$D = \frac{n_c - n_d}{N}, \quad \gamma = \frac{n_c - n_d}{n_c + n_d} \quad \text{et} \quad \tau_a = \frac{n_c - n_d}{\frac{1}{2}N(N-1)}$$

où n_c est le nombre de paires concordantes, n_d le nombre de paires discordantes, n_t le nombre de paires ex æquo et $N = n_c + n_d + n_t$ le nombre total d'observations.

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%. Enfin 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 8. La définition de ces modèles ainsi que leurs valeurs des critères de bon classement, introduits au paragraphe 3.1.4, constituent le Tableau 9.

Modèle	Incidence de l'eutypiose : Apparition
1	$\text{logit}(P(\text{"Malade"})) = 0,082\hat{Age} - 1,527I_{\{GWZ\}} - 3,031I_{\{PLS\}} - 1,935I_{\{RIS\}} - 3,870I_{\{SAV\}}$
Modèle	Incidence de l'eutypiose : Intensification
2	$\ln(\theta_1) = 1,423 + 0,051\hat{Age} - 1,819I_{\{BRG\}} - 2,630I_{\{CEN\}} - 1,758I_{\{DIO\}} - 3,962I_{\{JUR\}}$ $- 2,951I_{\{PAC\}} - 2,596I_{\{GWZ(ALS)\}} - 2,944I_{\{RIS(ALS)\}} - 4,170I_{\{MER(AQT)\}}$ $+ 4,420I_{\{PIN(BRG)\}} + 4,559I_{\{CHE(CEN)\}} + 2,565I_{\{SYR(DIO)\}}$ $\ln(\theta_2) = -2,116 + 0,051\hat{Age} - 1,819I_{\{BRG\}} - 2,630I_{\{CEN\}} - 1,758I_{\{DIO\}} - 3,962I_{\{JUR\}}$ $- 2,951I_{\{PAC\}} - 2,596I_{\{GWZ(ALS)\}} - 2,944I_{\{RIS(ALS)\}} - 4,170I_{\{MER(AQT)\}}$ $+ 4,420I_{\{PIN(BRG)\}} + 4,559I_{\{CHE(CEN)\}} + 2,565I_{\{SYR(DIO)\}}$
Modèle	Incidence de l'esca/BDA : Apparition
3	$\text{logit}(P(\text{"Malade"})) = 3,384 - 1,133I_{\{Ars1\}} - 1,998I_{\{Ars3\}}$

Modèle	Incidence de l'esca/BDA : Intensification
4	$\ln(\theta_1) = 3,912 - 3,927I_{\{CAR\}} + 1,799I_{\{GWZ\}} - 2,895I_{\{MER\}} - 2,246I_{\{MPG\}}$ $- 2,203I_{\{PIN\}} + 2,223I_{\{PLS\}} + 2,292I_{\{SAV\}} - 4,611I_{\{SYR\}}$ $\ln(\theta_2) = -0,968 - 3,927I_{\{CAR\}} + 1,799I_{\{GWZ\}} - 2,895I_{\{MER\}} - 2,246I_{\{MPG\}}$ $- 2,203I_{\{PIN\}} + 2,223I_{\{PLS\}} + 2,292I_{\{SAV\}} - 4,611I_{\{SYR\}}$
Modèle	Taux de mortalité des plants : Apparition
5I	$\text{logit}(P("Mort")) = -2,868 + 276,470Esca / BDA + 0,161\hat{Age}$
5II	$\text{logit}(P("Mort")) = 411,400Eutypiose + 240,430Esca / BDA$
Modèle	Taux de mortalité des plants : Évolution
6	$\ln(\theta_1) = 2,140 + 26,862Eutypiose + 0,027\hat{Age} - 1,942I_{\{CBF\}} - 2,056I_{\{CHD\}}$ $- 3,857I_{\{MER\}} - 1,515I_{\{MPG\}} - 2,132I_{\{PIN\}}$ $\ln(\theta_2) = -0,573 + 26,862Eutypiose + 0,027\hat{Age} - 1,942I_{\{CBF\}} - 2,056I_{\{CHD\}}$ $- 3,857I_{\{MER\}} - 1,515I_{\{MPG\}} - 2,132I_{\{PIN\}}$

TABLEAU 8

Modèles obtenus par régression logistique : estimations des coefficients

Variable à expliquer	Phase	Modèle	Concordants (%)	D	γ	τ_a
<i>Incidence de l'eutypiose</i>	Apparition	1	89,3	0,79	0,79	0,36
	Intensification	2	89,0	0,78	0,79	0,51
<i>Incidence de l'esca/BDA</i>	Apparition	3	45,4	0,30	0,50	0,03
	Intensification	4	78,1	0,62	0,66	0,33
<i>Taux de mortalité des plants</i>	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 9

Modèles obtenus par régression logistique : critères de bon classement

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 Représentations graphiques

Nous présentons ici quelques représentations graphiques qu'il est possible de construire en utilisant les modèles des Tableaux 8 et 9.

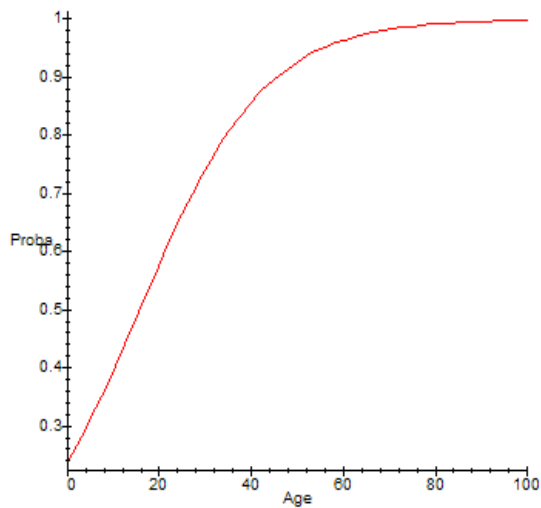


FIGURE 18
*Probabilité d'avoir au moins un pied
 symptomatique (eutyposie, phase
 d'apparition, Modèle 1) dans une parcelle du
 réseau en fonction de l'âge de la parcelle*

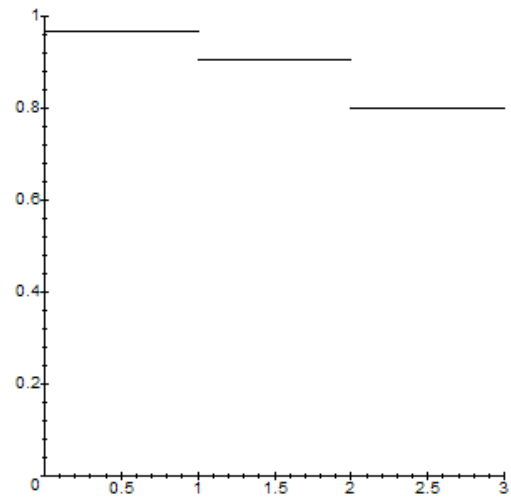


FIGURE 19
*Probabilité d'avoir au moins un pied
 symptomatique (esca/BDA, phase
 d'apparition, Modèle 3) dans une parcelle du
 réseau en fonction du nombre de
 traitement(s) à l'arsénite de soude*

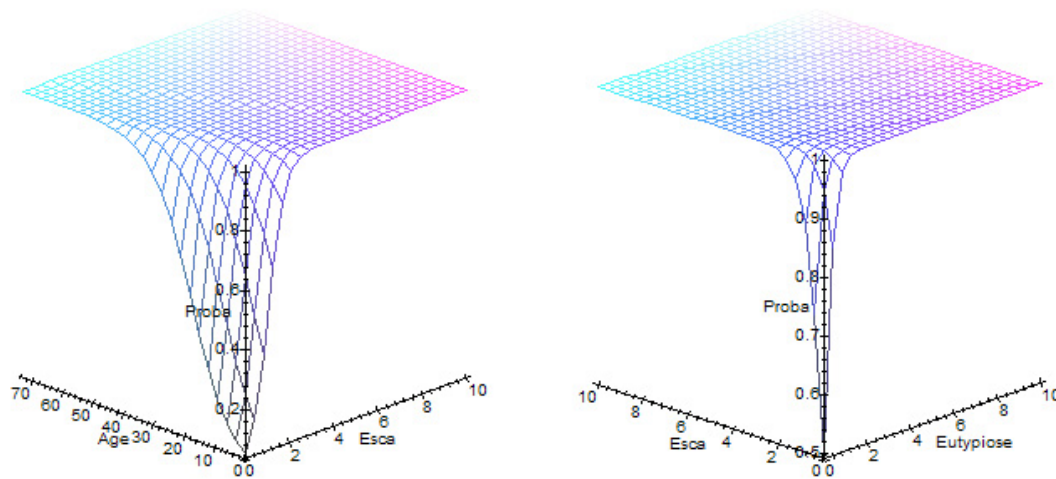


FIGURE 20
*Probabilité d'avoir au moins un pied mort, phase d'apparition du taux de mortalité, dans une
 parcelle du réseau selon respectivement le Modèle 5I, et le Modèle 5II, définis dans les
 Tableaux 8 et 9*

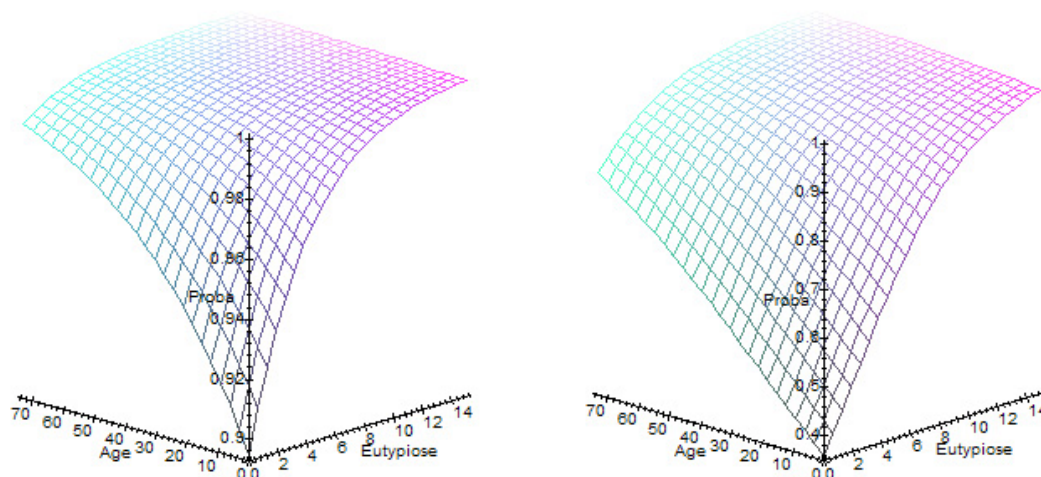


FIGURE 21

Évolution du taux de mortalité des plants d'une parcelle du réseau en fonction de l'âge de la parcelle et de l'incidence de l'eutypiose pour le Modèle 6. Sur le graphique de gauche, il s'agit de la probabilité du passage d'une mortalité faible à une mortalité moyenne et sur le graphique de droite d'une mortalité moyenne à une mortalité élevée

3.2.3 Bilan

- Nous pouvons distinguer, sur la Figure 18 associée au Modèle 1, 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 0,24 à 0,80 sur cette période, ce qui nous donne un taux d'accroissement annuel de l'ordre de 1,6%. La deuxième concerne les vignes de plus de 35 ans où là l'accroissement est beaucoup plus lent.
- La Figure 19 est issue du Modèle 3, apparition de l'esca/BDA, 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, voir la Figure 20 pour une représentation graphique. 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.
- Par exemple, le Modèle 5I implique à la fois l'esca/BDA et l'âge de la parcelle avec des résultats forts 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.
- Les deux représentations de la Figure 21 correspondent à l'évolution de la mortalité des plants en fonction de l'âge de la parcelle et de l'eutypiose.

3.3 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.

3.3.1 Quelques rappels théoriques

Notons S la statistique d'intérêt. S a une distribution d'échantillonnage, notée F_S , qui dépend de la fonction de répartition F de la variable aléatoire X . Comme la fonction de répartition F est inconnue, nous travaillons avec un estimateur de F , noté \hat{F} . Il est :

- soit construit à partir d'un modèle paramétrique qui ajuste bien les observations,
- soit est égal à la fonction de répartition empirique, notée F_n .

Approximation de la distribution d'échantillonnage

Le fait de remplacer la fonction de répartition F par l'un des deux estimateurs \hat{F} évoqués ci-dessus donnera une distribution d'échantillonnage F_S également modifiée puisque F_S dépend de F . Il y a deux approches pour calculer un estimateur de F_S , noté \hat{F}_S :

- *l'approche mathématique* qui consiste à déterminer \hat{F}_S grâce au calcul statistique
- *l'approche computationnelle* ou bootstrap qui consiste à déterminer \hat{F}_S par la simulation de pseudo-données.

Il y a deux formes de bootstrap : le bootstrap paramétrique et le bootstrap non-paramétrique.

Dans le bootstrap paramétrique, nous remplaçons d'abord F par une estimation \hat{F} construite à partir d'un modèle paramétrique qui ajuste bien les observations. Nous simulons ensuite r échantillons indépendants de taille n qui proviennent de la distribution \hat{F} . Enfin, nous calculons la statistique S pour chacun des r échantillons et nous obtenons les r valeurs simulées $s_1^*, s_2^*, \dots, s_r^*$.

Si nous ne pouvons pas attribuer un modèle paramétrique aux observations, nous utilisons la fonction de répartition empirique F_n comme approximation de F et obtiendrons ainsi le bootstrap non-paramétrique. Nous générons alors r échantillons indépendants de taille n provenant de F_n . Nous calculons à nouveau les r valeurs simulées $s_1^*, s_2^*, \dots, s_r^*$.

Dans les deux cas, la distribution empirique des valeurs simulées $s_1^*, s_2^*, \dots, s_r^*$ fournit une estimation de la distribution d'échantillonnage de S .

Intervalles de confiance, méthode des percentiles :

Supposons que nous recherchons un intervalle de confiance pour le paramètre θ , de risque α . Notons $\hat{\theta}_0$ une estimation de θ . Désignons par $\hat{\theta}_1^*, \hat{\theta}_2^*, \dots, \hat{\theta}_r^*$ les valeurs obtenues par technique bootstrap pour les estimations de θ . Alors, les limites inférieure et supérieure de l'intervalle de confiance pour la méthode des percentiles sont respectivement égales à :

$$L = \hat{\theta}_0 - (q_{1-\alpha/2} - \bar{\theta}^*)$$

$$U = \hat{\theta}_0 + (q_{\alpha/2} - \bar{\theta}^*)$$

où $\bar{\theta}^*$ est la moyenne des $\hat{\theta}_1^*, \dots, \hat{\theta}_r^*$, $q_{1-\alpha/2}$ et $q_{\alpha/2}$ sont respectivement les $100(1-\alpha/2)$ unième et $100\alpha/2$ unième percentiles de l'approximation de la distribution d'échantillonnage de θ . Il faut noter que la méthode des percentiles est efficace pour les estimateurs sans biais et lorsque l'erreur standard pour $\hat{\theta}_0$ est relativement constante. Une alternative, pour des cas plus généraux, est la méthode *BCa*, bootstrap avec correction accélérée pour le biais.

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

La méthode des percentiles a été appliquée aux paramètres des régressions logistiques présentées dans les Tableaux 8 et 9. Les résultats constituent le Tableau 10 et la Figure 22.

Modèle	Paramètre	Estimation	Intervalle de confiance
Apparition			
Modèle 1 <i>logit(Eutypiose)~Âge+Cépage</i>	Âge	0,082	[0,055 ; 0,111]
	GWZ	-1,527	[-2,779 ; -0,194]
	PLS	-3,031	[-5,019 ; -0,850]
	RIS	-1,935	[-3,261 ; -0,621]
	SAV	-3,870	[-5,279 ; -2,474]
Modèle 3 <i>logit(Esca/BDA)~Ars</i>	Constante	3,384	[2,936 ; 3,666]
	Ars1	-1,133	[-11,532 ; 5,811]
	Ars3	-1,998	[-5,279 ; -2,474]
Modèle 5I <i>logit(Mortalité)~Esca/BDA+Âge</i>	Constante	-2,868	[-4,270 ; -1,314]
	Esca/BDA	276,470	[254,440 ; 289,890]
	Âge	0,161	[0,108 ; 0,208]
Modèle 5II <i>logit(Mortalité)~Eutypiose + Esca/BDA</i>	Eutypiose	411,400	[377,360 ; 426,460]
	Esca/BDA	240,430	[216,800 ; 252,530]
Intensification			
Modèle 4 <i>logit(Esca/BDA)~Cépage</i>	Constante1	3,912	[3,015 ; 4,752]
	Constante2	-0,968	[-1,792 ; -0,050]
	CAR	-3,927	[-20,264 ; 1,428]
	GWZ	1,799	[0,387 ; 2,843]
	MER	-2,895	[-4,785 ; -0,231]
	MPG	-2,246	[-3,600 ; -0,902]
	PIN	-2,203	[-3,671 ; -0,843]
	PLS	2,223	[0,080 ; 3,354]
	SAV	2,292	[0,708 ; 3,292]
	SYR	-4,611	[-21,234 ; -1,079]

Modèle	Paramètre	Estimation	Intervalle de confiance
Évolution			
Modèle 6 $\text{logit}(\text{Mortalité}) \sim \text{Eutypiose} + \hat{\text{Age}}$ $+ \text{Cépage}$	Constante1	2,140	[1,021 ; 3,169]
	Constante2	-0,573	[-1,707 ; 0,559]
	Eutypiose	26,862	[12,830 ; 32,010]
	Âge	0,027	[0,005 ; 0,054]
	CBF	-1,942	[-3,620 ; -0,508]
	CHD	-2,056	[-3,282 ; -0,799]
	MER	-3,857	[-19,734 ; -1,013]
	MPG	-1,515	[-2,874 ; -0,309]
	PIN	-2,132	[-3,532 ; -0,799]

TABLEAU 10

Paramètres estimés pour chaque modalité inclus dans le modèle et leurs intervalles de confiance pour la méthode des percentiles

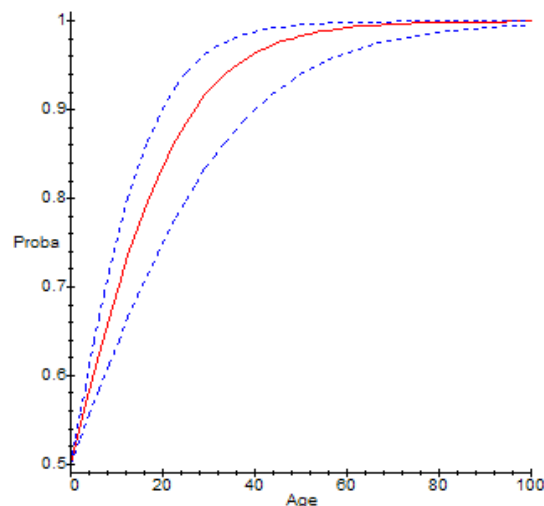


FIGURE 22

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 Analyses factorielles de tableaux multiples

À cette étape de l'étude, nous nous sommes posés la question quel rôle 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'année 2003 a été globalement moins bien renseignée. Enfin 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 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.

4.1 Présentation

Cette méthode a été effectuée sur les données des trois années 2003, 2004 et 2005 pour les mêmes individus et les mêmes variables, suivant le protocole suivant :

- Tout d'abord, nous avons regroupé dans un tableau l'ensemble des observations des incidences des deux maladies, du taux de mortalité et de l'âge de la parcelle sur les trois années obtenant donc un tableau de 321 lignes, 107 observations par année, et quatre colonnes (Euty, Esca, Mort, Âge) afin d'obtenir un tableau de Burt, B_{Tot} , des sommes des effectifs.
- Ensuite, nous avons juxtaposé à droite du tableau B_{Tot} les trois tableaux de Burt annuels B_{2003} , B_{2004} et B_{2005} des observations des incidences des deux maladies, du taux de mortalité et de l', chacun avec 13 lignes et 13 colonnes. Leurs colonnes seront mises en supplémentaires lors de l'utilisation de l'AFC.
- Nous avons également juxtaposé en dessous du tableau B_{Tot} les trois tableaux de Burt annuels B_{2003} , B_{2004} et B_{2005} . Leurs lignes seront mises en supplémentaires lors de l'utilisation de l'AFC.
- Puis, nous avons complété les cases vides de ce grand tableau B par des zéros.
- Enfin, nous avons réalisé l'AFC de B comme décrit au Tableau 11.

	Colonnes Actives	Colonnes supplémentaires		
Individus Actifs	B_{Tot}	B_{2003}	B_{2004}	B_{2005}
Individus supplémentaires	B_{2003}	0	0	0
	B_{2004}	0	0	0
	B_{2005}	0	0	0

TABLEAU 11
Stratégie d'analyse par AFC du tableau B

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

Au regard des inerties reportées dans le Tableau 12, nous retenons uniquement les quatre premiers axes, qui regroupent environ 72% de l'information. 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.

Axe	Valeur propre singulière	Pourcentage de l'inertie	
		Individuel	Cumulé
1	0,439	30,190	30,190
2	0,332	17,265	47,456
3	0,289	13,043	60,498
4	0,269	11,322	71,821
5	0,246	9,473	81,294
6	0,217	7,335	88,629
7	0,183	5,225	93,854
8	0,166	4,301	98,155
9	0,109	1,845	100,000

TABLEAU 12

Valeurs propres et inerties associées aux axes factoriels de l'ACM multi-tableau

	Coord.	Inertie	Cosinus ²	Coord.	Inertie	Cosinus ²
	<i>Dim.1</i>	<i>Dim.1</i>	<i>Dim.1</i>	<i>Dim.2</i>	<i>Dim.2</i>	<i>Dim.2</i>
Âge0	1,236	0,222	0,633	-0,348	0,031	0,050
Âge1	-0,067	0,002	0,013	-0,156	0,023	0,069
Âge2	-0,328	0,053	0,245	0,130	0,015	0,038
Âge3	0,180	0,004	0,012	0,656	0,082	0,152
Euty0	0,253	0,043	0,243	-0,290	0,099	0,321
Euty1	-0,039	0,001	0,002	0,540	0,179	0,412
Euty2	-0,567	0,088	0,311	0,018	0,000	0,000
Esca0	0,714	0,135	0,453	-0,059	0,002	0,003
Esca1	-0,022	0,000	0,002	0,339	0,131	0,428
Esca2	-0,464	0,081	0,296	-0,540	0,193	0,402
Mort0	0,861	0,257	0,802	-0,017	0,000	0,000
Mort1	-0,155	0,010	0,043	0,439	0,141	0,349
Mort2	-0,442	0,103	0,440	-0,338	0,105	0,256

	Coord.	Inertie	Cosinus ²	Coord.	Inertie	Cosinus ²
	<i>Dim.3</i>	<i>Dim.3</i>	<i>Dim.3</i>	<i>Dim.4</i>	<i>Dim.4</i>	<i>Dim.4</i>
Âge0	-0,011	0,000	0,000	-0,012	0,000	0,000
Âge1	-0,364	0,167	0,374	-0,161	0,038	0,073
Âge2	0,213	0,052	0,103	0,292	0,113	0,194
Âge3	0,866	0,189	0,266	-0,509	0,075	0,092
Euty0	-0,046	0,003	0,008	-0,219	0,086	0,183
Euty1	0,355	0,103	0,179	-0,050	0,002	0,004
Euty2	-0,343	0,075	0,114	0,599	0,262	0,347
Esca0	0,460	0,130	0,188	0,383	0,104	0,131
Esca1	-0,315	0,149	0,368	0,022	0,001	0,002
Esca2	0,216	0,041	0,064	-0,306	0,095	0,129
Mort0	-0,160	0,020	0,028	0,224	0,046	0,054
Mort1	-0,132	0,017	0,032	-0,365	0,149	0,242
Mort2	0,210	0,054	0,099	0,143	0,029	0,046

	Coord.	Cosinus²	Coord.	Cosinus²	Coord.	Cosinus²	Coord.	Cosinus²
	<i>Dim.1</i>	<i>Dim.1</i>	<i>Dim.2</i>	<i>Dim.2</i>	<i>Dim.3</i>	<i>Dim.3</i>	<i>Dim.4</i>	<i>Dim.4</i>
Euty03_0	1,417	0,538	-0,296	0,024	0,076	0,002	0,166	0,007
Euty03_1	0,110	0,009	0,025	0,000	-0,383	0,105	-0,145	0,015
Euty03_2	-0,158	0,016	0,389	0,099	0,151	0,015	0,363	0,087
Esca03_0	0,303	0,024	0,725	0,135	0,881	0,199	-0,491	0,062
Esca03_1	0,518	0,189	-0,126	0,011	-0,115	0,009	-0,084	0,005
Esca03_2	-0,023	0,000	0,639	0,234	0,253	0,037	-0,080	0,004
Mort03_0	-0,408	0,081	0,120	0,007	-0,330	0,053	0,736	0,263
Mort03_1	0,699	0,232	-0,155	0,011	0,402	0,077	0,408	0,079
Mort03_2	0,005	0,000	0,436	0,145	-0,285	0,062	-0,049	0,002
Euty04_0	-0,471	0,123	-0,212	0,025	0,226	0,028	-0,403	0,090
Euty04_1	0,822	0,359	0,009	0,000	-0,114	0,007	0,245	0,032
Euty04_2	-0,141	0,012	0,586	0,210	-0,150	0,014	-0,301	0,055
Esca04_0	-0,255	0,045	-0,128	0,011	0,281	0,055	0,298	0,062
Esca04_1	1,105	0,376	-0,353	0,038	-0,004	0,000	-0,074	0,002
Esca04_2	-0,128	0,012	-0,296	0,062	-0,337	0,081	-0,204	0,030
Mort04_0	-0,377	0,097	-0,001	0,000	0,235	0,038	0,200	0,027
Mort04_1	0,188	0,009	0,590	0,090	0,888	0,204	-0,473	0,058
Mort04_2	0,118	0,011	-0,305	0,074	-0,012	0,000	-0,275	0,060
Euty05_0	-0,048	0,001	0,439	0,111	0,513	0,152	-0,008	0,000
Euty05_1	-0,590	0,170	-0,057	0,002	-0,322	0,051	0,544	0,145
Euty05_2	0,718	0,240	-0,051	0,001	0,466	0,101	0,349	0,057
Esca05_0	-0,054	0,002	0,287	0,064	-0,324	0,082	-0,025	0,000
Esca05_1	-0,471	0,125	-0,610	0,210	0,189	0,020	-0,307	0,053
Esca05_2	0,978	0,453	-0,112	0,006	-0,117	0,006	0,260	0,032
Mort05_0	-0,127	0,010	0,322	0,066	-0,183	0,021	-0,441	0,124
Mort05_1	-0,445	0,133	-0,430	0,124	0,229	0,035	0,086	0,005
Mort05_2	1,187	0,406	-0,395	0,045	-0,104	0,003	-0,128	0,005

TABLEAU 13

Coordonnées, inerties et cosinus carrés associés aux modalités des différentes variables pour les quatre premiers axes factoriels de l'ACM multi-tableau

Le nombre de points à interpréter sur la Figure 23 est important. C'est pourquoi nous avons opté pour une étude fragmentée de « sous graphes », qui constituent la Figure 24. Chacun des graphes de la Figure 24 est en effet associé à l'une des quatre variables « Âge », « Euty », « Esca » et « Mort ».

Les trajectoires et les points étiquetés de la Figure 23 sont le résultat de l'AFC du tableau de Burt B_{Tot} tandis que les points sans étiquette correspondent aux projections des colonnes des tableaux supplémentaires B_{2003} , B_{2004} et B_{2005} et donc aux valeurs annuelles.

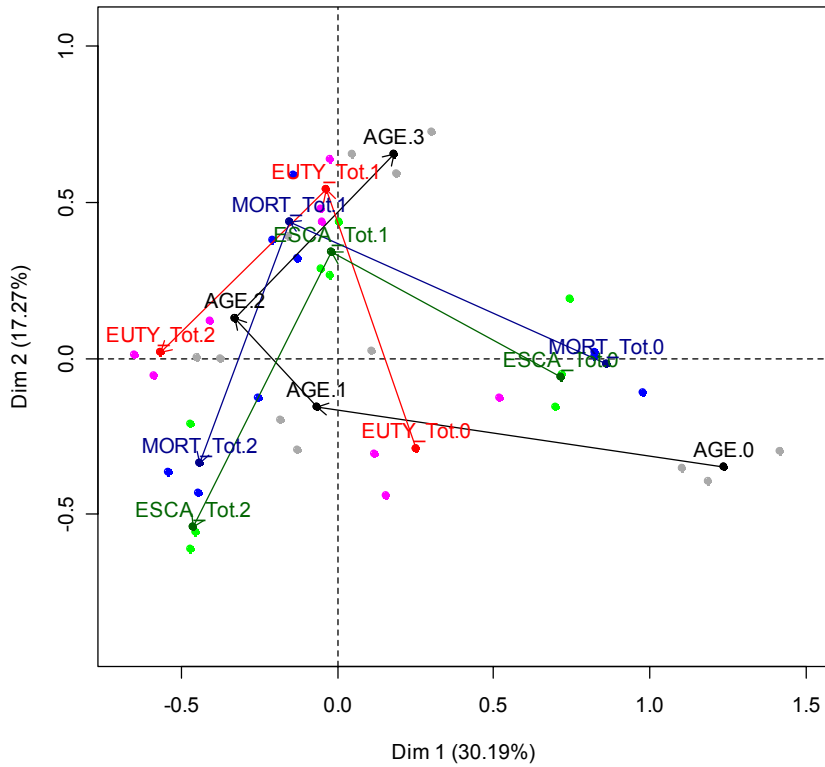


FIGURE 23

Premier plan factoriel présentant les modalités annuelles et moyennes de toutes les variables

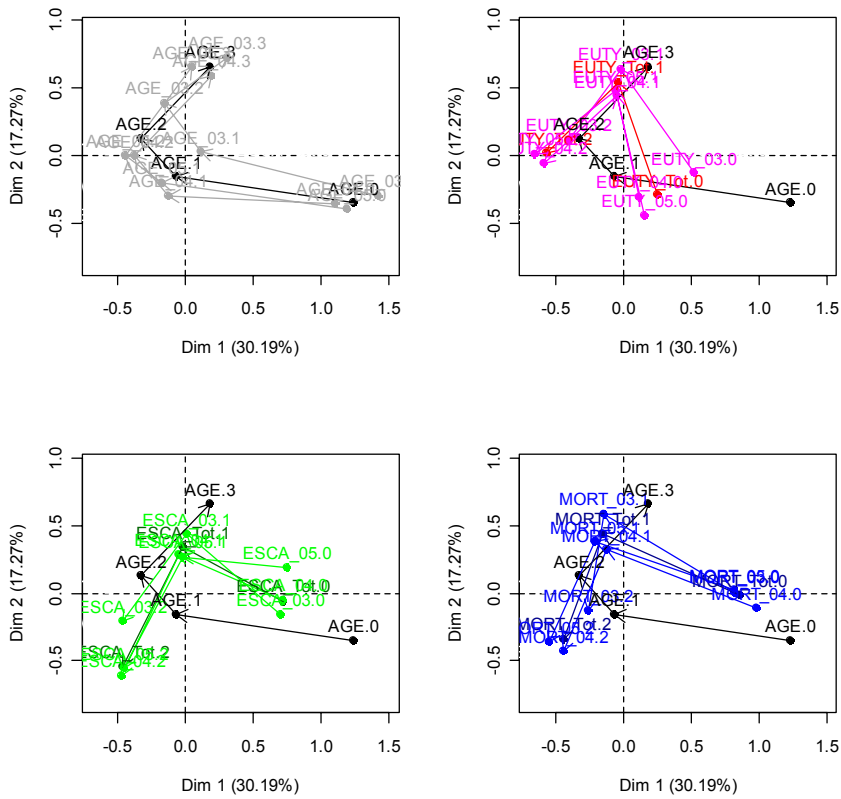


FIGURE 24

Premier plan factoriel présentant les trajectoires annuelles par variable

La Figure 25 reproduit les analyses des Figures 23 et 24 mais pour le deuxième plan factoriel.

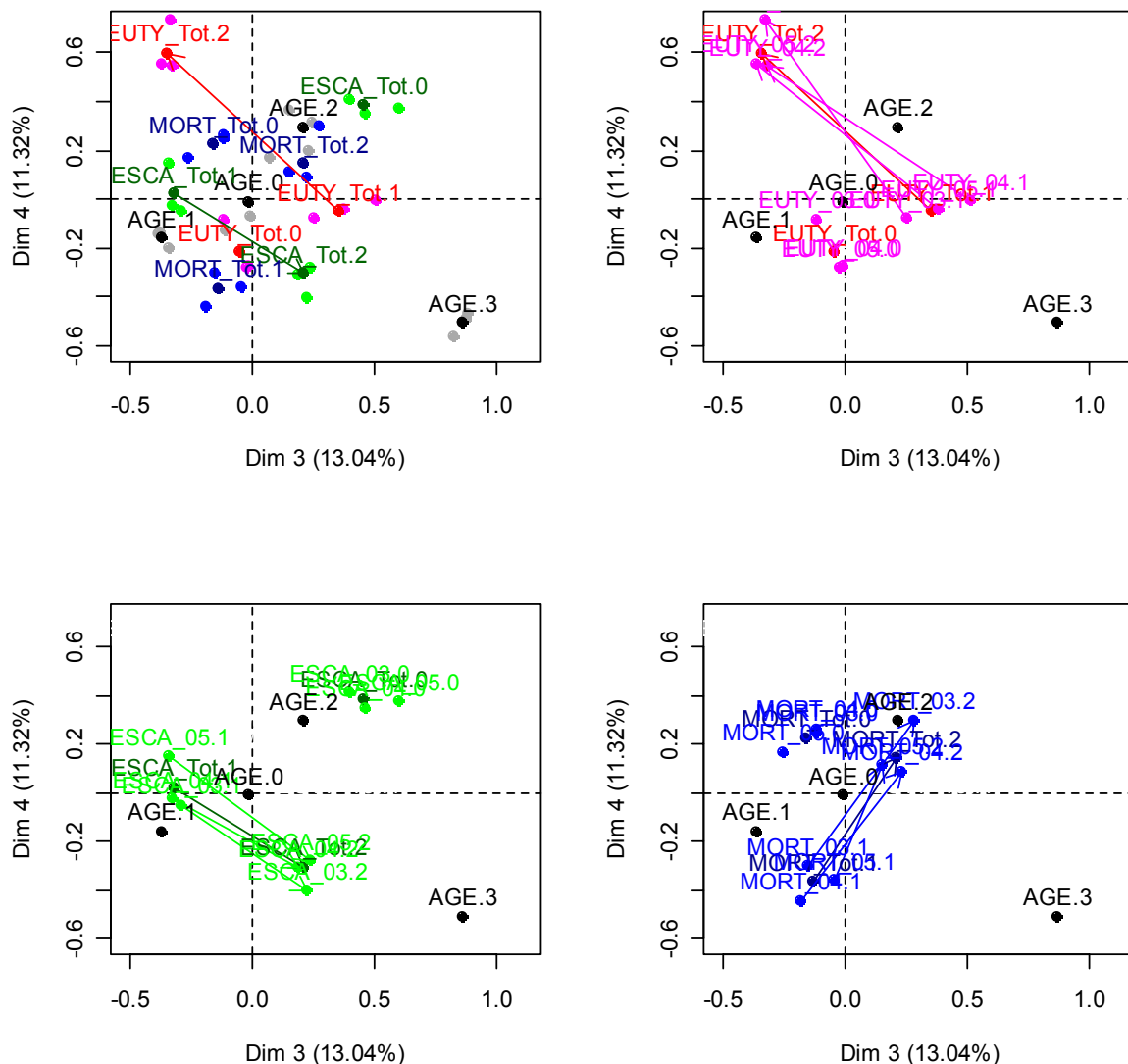


FIGURE 25

Deuxième plan factoriel présentant les modalités annuelles et moyennes et l'étude de la stabilité des trajectoires annuelles des variables « Euty », « Esca » et « Mort »

De manière générale, nous retrouvons les résultats déjà constatés lors des ACM et des ACP à savoir que le premier axe factoriel est représentatif de l'intensification des maladies et de la mortalité. Le deuxième axe est lui représentatif de l'apparition des maladies et de la mortalité ainsi que de l'âge de la parcelle mis en parallèle. L'ensemble des années sont proches et ce pour l'ensemble des modalités. 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.

Les troisième et quatrième axes eux semblent toujours montrer l'opposition entre les évolutions des deux maladies, les fortes incidences d'eutypiose ne coïncidant pas avec de fortes incidences d'esca/BDA par exemple. Cette opposition entre les phases d'intensification

pour les deux maladies, en violet et en vert, est stable pour les trois 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 multiples multi-tableaux.

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 des analyses des correspondances multiples. En effet seules des liaisons de type linéaires seront pleinement mises en évidence par les analyses en composantes principales. L'analyse factorielle de données mixtes, contrairement à l'analyse des 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 le 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.

6 Annexes

6.1 Logiciels et programmes

Tous les calculs réalisés dans cet article ont été menés à bien avec le logiciel **R** (Team 2007) sauf les Figures 13, 14, 15, 16 et 17 qui ont été produites avec le logiciel Minitab (Minitab Inc 2000). 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 Terminologie des maladies du bois de la vigne

Nous donnons ici quelques définitions, tirées des deux ouvrages GALET (1988) et GALET P. (1991), de termes du vocabulaire viticole ou de l'épidémiologie végétale.

Cep de vigne : Le cep de vigne est synonyme de pied de vigne.

Cépages : Les cépages sont des variétés de l'espèce *Vitis Vinifera*, la vigne européenne. Il existe plus de 6000 cépages à travers le monde, dont les raisins diffèrent par :

- leur goût : certains sont plus ou moins acides, plus ou moins sucrés, et développent diverses saveurs;
- leur couleur : la pellicule peut être blanche ou colorée, de même pour la pulpe;
- leur grosseur : en général, les baies de raisins de table sont plus grosses et plus charnues que les baies de raisins de cuve, destinées à l'élaboration du vin.

Parmi les vignes en production en France, il y a une cinquantaine de cépages principaux. Les cépages les plus plantés sont :

- en rouges : Carignan (CAR), Grenache (GRE), Merlot (MER), Cabernet Sauvignon (CBS), Syrah (SYR), Gamay (GAM), Cabernet Franc (CBF), Pinot Noir (PIN);
- en blancs : Ugni Blanc (UB), Chardonnay (CHD), Sauvignon (SAU), Sémillon, Melon (MEL), Chenin (CHE), Colombard, Riesling (RIS).

Clone : Un clone est une descendance végétative d'une variété conforme à une souche de vigne choisie pour l'identité, les caractères phénotypiques et l'état sanitaire de la plante-mère.

Floraison : La floraison est une phase végétative de la vigne au cours de laquelle se produit la fécondation des grappes, ce qui va permettre la formation des baies. La floraison se déroule généralement mi-juin et dure de 10 à 15 jours environ.

Greffage : Le greffage est une méthode qui consiste à fixer sur un porte-greffe résistant au phylloxéra un greffon d'origine locale.

Greffés-soudés : Les Greffés-soudés sont des fractions de sarments ou rameaux herbacés de vigne assemblées entre elles par greffage, dont la partie souterraine est racinée, comprenant une variété greffon (à fruits) et une variété porte-greffe.

Greffon : Un greffon est un fragment de végétal portant un ou plusieurs bourgeons, qui seront soudés, par la greffe, au porte-greffe. Le greffon donnera la partie aérienne de l'ensemble et appartient à une variété sélectionnée pour la qualité de ses fruits.

Invasion phylloxérique : Le phylloxera est un insecte s'attaquant aux racines de vigne européenne. Il est d'origine américaine et a été introduit en 1864 dans le Gard avec des plants de vigne américaine. Identifié par Emile Planchon en 1868, il a progressivement envahi les différents vignobles de France puis d'Europe. Le greffage sur des plants américains résistants a permis de venir à bout de ce fléau.

ITV : L'Institut Technique de la Vigne et du Vin est un organisme technique professionnel de recherche et d'expérimentation sur la vigne et le vin. Voir le site d'ITV France, www.itvfrance.com.

Phylloxera : Le Phylloxera Vastatrix est un puceron parasite qui s'attaque aux racines de la vigne. Cet insecte provoque la mort des ceps par sa piqûre. Par extension, nom de la maladie de la vigne causée par ce puceron. Le Phylloxera fut introduit accidentellement en France avec des plants américains vers 1865 et détruisit plus de la moitié du vignoble. Le vignoble français fut reconstitué par des greffes sur plants américains résistant aux parasites dits porte-greffes.

Plant de vigne : Il faut distinguer d'une part les racinés composés d'une seule variété et d'autre part les greffés-soudés qui comportent une variété greffon et, une variété porte-greffe réunies par une soudure.

Porte-greffe : Le porte-greffe est la partie inférieure de la vigne (d'origine américaine) sur laquelle est fixée le greffon. Voir phylloxera.

Sarment : Le sarment est un rameau de vigne de l'année qui fait l'objet d'une taille en hiver.

Souche de vigne : La souche de vigne est synonyme de pied de vigne ou encore de cep de vigne.

SRPV : Le Service Régional de la Protection des Végétaux, service de la Direction Régionale de l'Agriculture et des Forêts, assure un rôle d'appui technique, d'expertise et de contrôle en matière de protection phytosanitaire. Le SRPV élabore et diffuse des Avertissements Agricoles, qui constituent la base d'une agriculture raisonnée et durable. Ce sont des bulletins techniques hebdomadaires qui ont pour objectifs :

- de prévenir les producteurs sur l'évolution parasitaire et les différentes stratégies de lutte dans le respect des Bonnes Pratiques Agricoles,
- de diffuser les informations réglementaires concernant l'évolution et les consignes d'utilisation des produits phytopharmaceutiques,
- de fournir au public les messages spécifiques orientés sur la protection de l'applicateur et de l'environnement (bonnes pratiques phytosanitaires).

Taille de la vigne : La taille hivernale et les travaux en vert constituent le cœur du métier de vigneron. Ces travaux permettent de limiter la quantité récoltée, de provoquer une meilleure maturité du raisin et d'augmenter la qualité du vin. La taille Cordon de Royat est une taille courte sur charpente longue : le cordon. La taille Guyot est une taille longue sur charpente courte qui est surtout pratiquée dans les secteurs gélifs. La taille en gobelet, dont la pratique a été répandue par les Romains, est une taille courte généralement à bras très court ce qui favorise le développement des maladies.

Véraison : La véraison est le moment où le grain, qui était primitivement vert et dur, devient, par perte de chlorophylle, translucide chez les cépages blancs ou se colore en rouge chez les cépages colorés en rouge ou noir.

6.3 Photographies des symptômes foliaires des maladies



FIGURE A

Plants de vignes malades de :

- *l'eutypiose (en haut à gauche),*
- *le black dead arm (en haut à droite),*
- *l'esca (en bas à droite)*

Photographies : Jacques Grosman



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

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 ¹⁷	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

¹⁷ La région Champagne-Ardenne dispose de son propre observatoire des maladies du bois de la vigne.

Rhône-Alpes	57 371,59	6,6
Total	873 702,20	100,0
Total enquête	834 245,90	95,5

TABLEAU A
Tableau des surfaces viticoles par régions administratives françaises

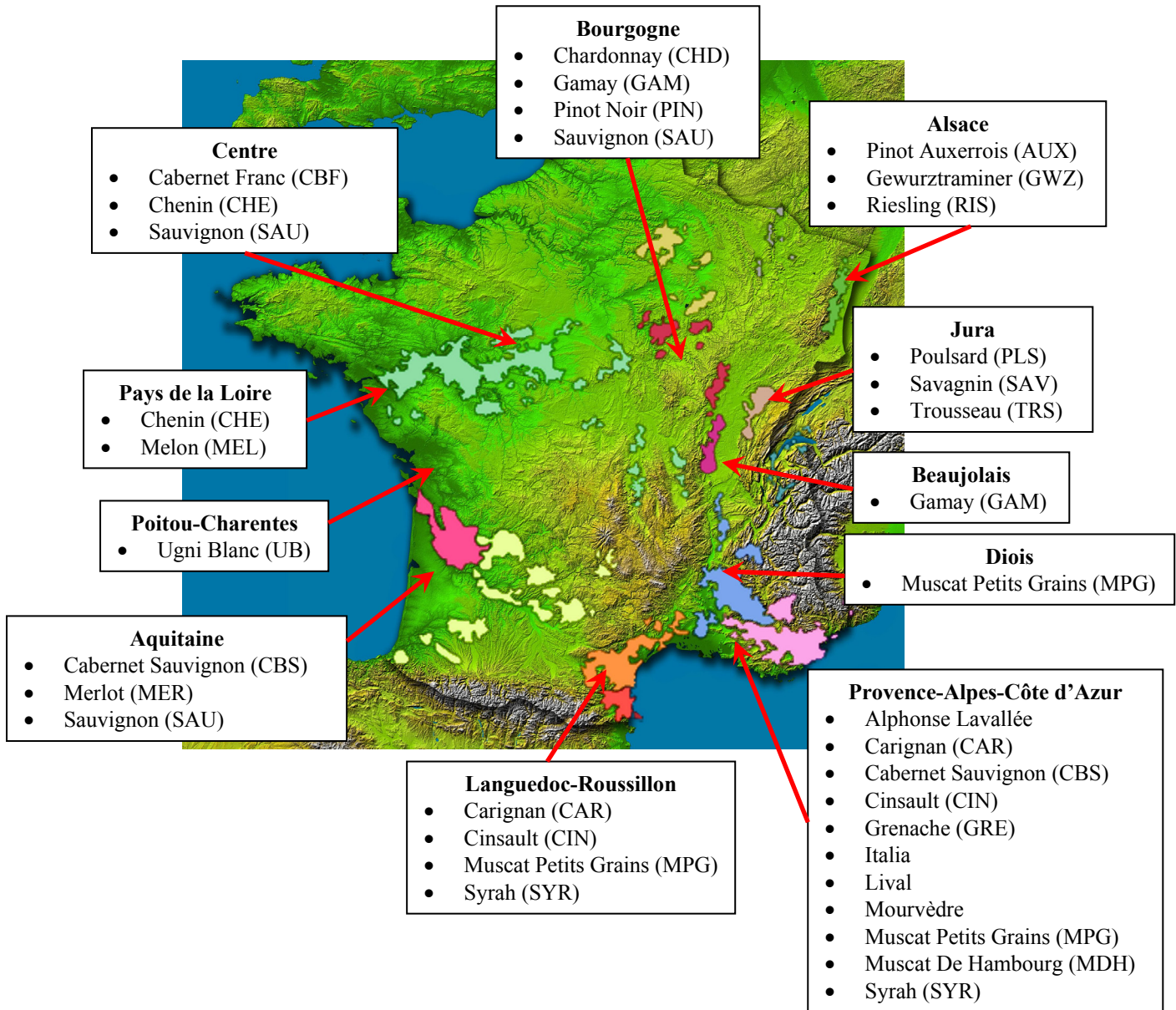


FIGURE B
Principales régions viticoles françaises et cépages majoritairement cultivés dans ces régions.
Les cépages présents dans le jeu de données analysé sont accompagnés du code qui a été introduit pour les identifier au Tableau B

6.5 Questionnaire d'enquête

Fiche de parcelle pour l'Observatoire National des Maladies du Bois de la Vigne		
Vignoble de :		Organisme(s) :
EXPLOITANT		
Nom Prénom :		Raison sociale :
Adresse :		Tél. :
		Fax :
		Mail :
PARCELLE		
Référence :		Cépage (clone) :
Commune :		Porte-greffe (clone) :
Lieu dit :		Année de plantation :
		Surface :
Type de taille :		Densité de plantation :
		Vigueur : forte, moyenne ou faible
Type de sol :		
TRAITEMENTS D'HIVER		
Arsénite en 1999 :	oui ou non	Autre :
Arsénite en 2000 :	oui ou non	Autre :
Arsénite en 2001 :	oui ou non	Autre :
PROPHYLAXIE RÉALISÉE EN 2004		
Date de taille :	indiquer au moins la quinzaine	Prétaillage : oui ou non
Sarments :	broyage ou enlèvement	Nettoyage bois mort : enlèvement ou non
Pratiques contre les maladies du bois :	égourmandage, perçage des souches, traitement des plaies de taille, environnement de la parcelle, ...	
Date de début de la mise en œuvre de mesures prophylactiques		indiquer l'année
AUTRES OBSERVATIONS		

FIGURE C

Questionnaire distribué à l'ensemble des collaborateurs de l'enquête

6.6 Codage des variables

Variable	Symbole	Catégories	Définition des catégories	Unité
<i>Maladies et mortalité</i>				
Eutypiose	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 Petits Grains ; PIN : Pinot Noir ; PLS : Poulsard ; 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

TABLEAU B
Tableau descriptif des variables et de leurs modalités

Références

- AGRESTI A. (1990), *Categorical data analysis*. Wiley, New York.
- BRY X. (1995), *Analyses factorielles simples*. Economica, Paris.
- BRY X. (1996), *Analyses factorielles multiples*. Economica, Paris.
- 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*, Paris.
- 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.
- 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.
- 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.
- GALET P. (1988), *Précis de viticulture*, 5^{ème} édition, Imprimerie Dehan, Montpellier.
- GALET P. (1991), *Précis de pathologie viticole*, Imprimerie Dehan, Montpellier.
- GOODMAN L.A. et KRUSKAL W.H. (1954), Measures of Association for Cross
- 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.
- Classifications. *Journal of the American Statistical Association*, 49, 732–764.
- LAVIT Ch., ESCOUFIER Y., SABATIER R. et TRAISSAC P. (1994), The ACT (STATIS Method). *Computational Statistics and Data Analysis*, 18, 97–119.
- MINITAB Inc. (2000), *Minitab 13.31*. <http://www.minitab.com>.
- PAGÈS J. (2002), Analyse factorielle multiple appliquée aux variables qualitatives et aux données mixtes. *Rev. Statistique Appliquée*, L (4), 5–37.
- PAGÈS J. (2004), Analyse factorielle de données mixtes. *Rev. Statistique Appliquée* LII (4), 93–111.
- SIEGEL S. et CASTELLAN N.J. (1988), *Nonparametric Statistics for the Behavioural Sciences*. 2^{ème} édition, McGraw-Hill Inc, New York.
- 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>.