CS-BIGS 1(2): 183-202 http://www.bentley.edu/csbigs/vol1-2/bertrand.pdf

Using Factor Analyses to Explore Data Generated by the National Grapevine Wood Diseases Survey

Frédéric Bertrand
Laboratoire de Statistique, IRMA, ULP & CNRS (UMR 7501), France

Myriam Maumy
Laboratoire de Statistique, IRMA, ULP & CNRS (UMR 7501), France

Lionel Fussler
DRAF-SRPV Alsace, France

Nathalie Kobes
DRAF-SRPV Alsace, France

Serge Savary
UMR Santé Végétale, Centre INRA de Bordeaux, France

Jacques Grosman
DRAF-SRPV Rhône, France

The Grapevine Wood Diseases National Observatory yields a cohesive and large data set which may be analyzed with different approaches. In our study, we deal with complex data composed of quantitative and qualitative variables which evolve with time, since data for three successive years are available. The objective of the study was to produce the largest possible amount of information from this data set, in order to highlight main trends. To this aim, we used several data analysis techniques. Our study proceeds in three stages. First, relationships between the different variables are identified using bivariate measures of association and tests. Then factorial methods, namely multiple correspondence analysis and factor analysis of mixed data are used to look for multivariate dependencies between the variables of the dataset. Last, we use factor analysis of multi-tables, each table representing a year, in order to account for the successive years of data. The exposition is accessible to readers with an intermediate knowledge of statistics. A prior exposure to multiple correspondence analysis is quite useful for reading the article.

Keywords: *Fisher-Freeman-Halton exact test, Multiple correspondence analysis, Factor analysis of mixed data, Factor analysis of multi-tables.*

Foreword

It is easy to perform the following analyses with any statistical software packages that provides MCA techniques and especially with **R** (R Development Core Team, 2007) and the FactoMineR package (Husson, Lê and Mazet, 2007). Indeed this **R** package is fully integrated in the **R** Commander interface (Fox *et al.*, 2007) and therefore all the **R** code is generated automatically using menus and popup windows. Some of the procedures of the FactoMineR package were slightly enhanced to produce the graphical outputs that were included throughout this paper. All the **R** code lines written by the authors are available upon simple request to the corresponding author.

1. Background

Since November 8th 2001 and the banning of the use of sodium arsenite in every sector of agriculture, in France then in Europe, there has no longer been any authorized way of treating grapevine trunk diseases, namely eutypa dieback (ED), esca decline (ES) and black dead arm (BDA); see Figure 1 for photos of afflicted grapevine. Sodium arsenite is a chemical made of arsenic and its use was forbidden for the sake of wine growers, since there was hard evidence that this chemical is carcinogenic. Wine growers have been deeply concerned by this drastic measure since they fear that it may result in a steady rise of grapevine trunk diseases rates in French vineyards. Unfortunately, almost every time, the final stage of any of the grapevine trunk diseases is the death of the grapevine plant. Moreover, we should bear in mind that the scientific community lacks accurate studies on the epidemiology of these diseases; no other cure could be quickly found and its use recommended.

In order to collect all the data and organize the activities of all the research teams working on grapevine trunk diseases a national technical group was founded in 2001. This national technical group decided to carry out the National Grapevine Wood Diseases Survey for several years. We emphasize the fact that no other survey in the world was ever carried out with the same scale and with as many teams involved. Indeed, up to now, such surveys have only been focused on a small area and on some varieties of vine, whereas the data that were collected by the National Grapevine Wood Diseases Survey deal with all the main French vine varieties and with all French vine-growing regions. The duration of the survey was already extended from three years to six years and is to be extended to a total duration of nine years in order to provide a reliable set of data for longitudinal data analysis. The primary objective[1] of the survey is to collect enough data to decide whether the banning of the use of sodium arsenite will result in a steady increase in grapevine trunk disease rates, in order to be able to evaluate the economic scope of such an interdiction. The secondary objectives are to identify some of the factors that explain the variability of grapevine trunk disease rates and eventually design several experiments in order to validate hypotheses that were highlighted by the survey's results.

2. Features of the survey

The memorandum DGAL/SDQPV/N2003-8085 published on May 19th 2003[2] specifies the features of the survey. We now recall its main features:

- Every French vine-growing region suffers from at least one of the grapevine trunk diseases and hereby was to be included in the survey. However, in fact, only 12 areas defined as vine-growing regions or administrative divisions were sampled: 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. These areas are the widest vine-growing regions: the area covered by the grapevines in these regions accounts for 95,5 % of the total area covered by grapevines in France, see Table 1 and Figure 2. Corsica was not included in the survey since, as an island, it has its own specificity.
- At least 25 parcels per vine-growing region and per vine variety were to be surveyed, in order to suitably depict the repartitions, the frequencies and the intensities of the diseases according to the vine-growing region and per vine variety.
- For any of the parcels that were randomly chosen among the observation networks 300 grapevines were marked and spotted. These 300 grapevines were divided in 10 locations, randomly chosen in the parcel, of 30 grapevines. The size of the parcels was not taken into account for the selection of a parcel since it has no interest from an epidemiological point of view.

[1] The goals of the study are stated in the memorandum DGAL/SDQPV/N2004-8126 which is available at the following address: http://www.agriculture.gouv.fr/spip/IMG/pdf/dgaln20048126z.pdf.
[2] This memorandum is available at the following address: http://www.agriculture.gouv.fr/spip/IMG/pdf/dgaln20038085.pdf.

- The observations were carried out on the same locations, hence on the same grapevines, during the three years of the survey.

- Grapevine trunk diseases were searched for at two particular stages of vine growth: blossoming for the eutypa dieback and veraison (change of color of the grape berries) for the esca decline and the black dead arm. At this very moment, dead or non producing grapevines were also counted.

- The only preset criterion for the choice of a parcel was the vine variety.

The features of the survey were specified by the Department of agriculture and the national technical group, to whom the ITV[3] and the INRA[4] belong, formally agreed to them. From a statistical point of view, the sample sizes chosen for this survey are of the same magnitude as those that are usually used in that kind of studies.

As for the criterion for the choice of a parcel, the vine variety is a simple way of identifying the parcels and is therefore well suited for such a large scale study since more than 40 different groups are involved in the data collection. From a biological point of view, a wide range of vulnerabilities was observed between the vine varieties even among the same vine-growing region. As a result, the vine variety was guessed to be one of the most interesting factors and the survey was designed to make it possible to have that hypothesis thoroughly investigated.

For any of the parcels that were randomly chosen, additional information about the characteristics and the cropping habits were collected using questionnaires, see Figure 3.

3. Preliminary statistical study–Bivariate tests

The esca decline (ES) shows all the same symptoms as black dead arm (BDA), see Figure 1 for a visual proof of that statement. As a consequence, these two diseases are always confounded in a same disease that we will call esca/BDA. We have three years of data, 2003, 2004 and 2005, at our disposal for these two diseases. In order to account for the variation of the intensities of the diseases among the population of cropped parcels as well as for the mortality rate of the grapevines the following factors and variables were chosen: the vine-growing region –factor–, the vine variety –factor–, the age of the parcel[5] –variable–, the density of planting –variable–, the rootstock that was used –factor–, the management of pruning residues –factor–, what happened to the dead wood –factor–, pre-pruning –factor–, the pruning method –factor–, the date of beginning and end of the pruning of the grapevine –two variables– and the number of times sodium arsenite was used from 1999 to 2001, just before the use of that cure was banned –factor–. Other data, such as the vigor of the grapevine –factor–, the type of soil –factor–, the area of the parcel –variable–, were also collected. However, these two factors and this variable were not included in the study because of a low level of response or a low reliability of the answers.

Table 1. French administrative regions and the area covered by grapevines. Regions sampled for the survey are in bold.

Wine-producing region	Area covered by grapevines (*ha*)	Percentage of total area covered by grapevines
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-Ardennes[6]	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[7]	41,000.29	4.7
Pays de la Loire	37,877.60	4.3
Picardie	2,810.02	0.3
Poitou-Charentes	80,794.61	9.2
Provence-Alpes-Côte d'Azur	99,850.91	11.4
Rhône-Alpes	57,371.59	6.6
Total	873,702.20	100.0
Total surveyed	834,245.90	95.5

We recall that the incidence rate of a disease in a parcel is the percentage of grapevine trunks that suffer from the disease in that parcel and that the mortality rate is estimated by summing, in the parcel, the number of dead or missing grapevines and the number of young

[3] Institut Technique de la Vigne et du Vin (Technical Institute of Grapevine and Wine).

[4] Institut National de la Recherche Agronomique (French National Institute for Agricultural Research).

[5] The age of the parcel is equal to the age of most of the grapevines that belongs to that parcel.

[6] The Champagne-Ardennes vine-growing region set up its own local grapevine wood disease survey.

[7] The Midi-Pyrénées vine-growing region was to be sampled in 2003 and 2004 yet it in fact began to send back results in 2005 and therefore could not be included in the analysis for the 2003 to 2005 period.

Figure 1. Grapevines suffering from:
Top left: Eutypa dieback
Top right: Black dead arm
Bottom right: Esca

Photos: Jacques Grosman

"complants" (recently replaced grapevines still not producing grapes), and dividing by the total number of grapevines, about 300, whose state was observed in that parcel.

The dataset is made of 701 parcels surveyed in 12 vine-growing regions. Twenty-six different vine varieties[8] belong to the dataset. These data were collected following a detailed protocol during the three years of the study. We removed 4 vine varieties, the Alphonse Lavallée, the Italia, the Lival and the Mourvèdre from the dataset since they accounted only for a total of 1.28 % of the total sample size. We also removed from the database any of the parcels for which there was an outlier or a missing value.

In order to investigate our primary and secondary objectives, we used several statistical methods. We started with a three-step preliminary data analysis.

First, we checked the dependencies between the variables in our dataset; this analysis showed that the influence of the age of the parcel was strong on the incidence and mortality rates, especially on the incidence rate of the esca/BDA. This led us to identifying mainly four different stages (see Table 2) of relationship between the age of the parcel and the three variables that we wanted to explain. These relationships frankly differed from a stage to another with even sign reversals in the coefficients of correlation between the age of the parcel and the incidence rates of the two diseases and the mortality rate. The details of this work are available upon request from the corresponding author[9].

Then, we carried out many non parametric Kruskal-Wallis tests (Siegel and Castellan, 1988). The multiple comparisons performed after these bivariate tests between a factor and a variable enabled us, for instance, to design classes for different sensitivities to the two diseases among

[8] The Alphonse Lavallée, the Pinot Auxerrois, the Cabernet Franc, the Cabernet Sauvignon, the Carignan, the Chardonnay, the Chenin, the Cinsault, the Gamay, the Gewurztraminer, the Grenache, the Italia, the Lival, the Melon, the Merlot, the Mourvèdre, the Muscat de Hambourg, the Muscat à Petits Grains, the Pinot Noir, the Poulsard, the Riesling, the Sauvignon, the Savagnin, the Syrah, the Trousseau and the Ugni Blanc.

[9] These analyses are detailed in the report by Lionel Fussler.

the different vine varieties. These results confirmed what was known up to now about the relative weakness of some of the vine varieties to the two diseases.

As a consequence of not only the Kruskal-Wallis tests but also the many bivariate tests such as independence χ^2 tests (Agresti, 1990), and the Fisher-Freeman-Halton

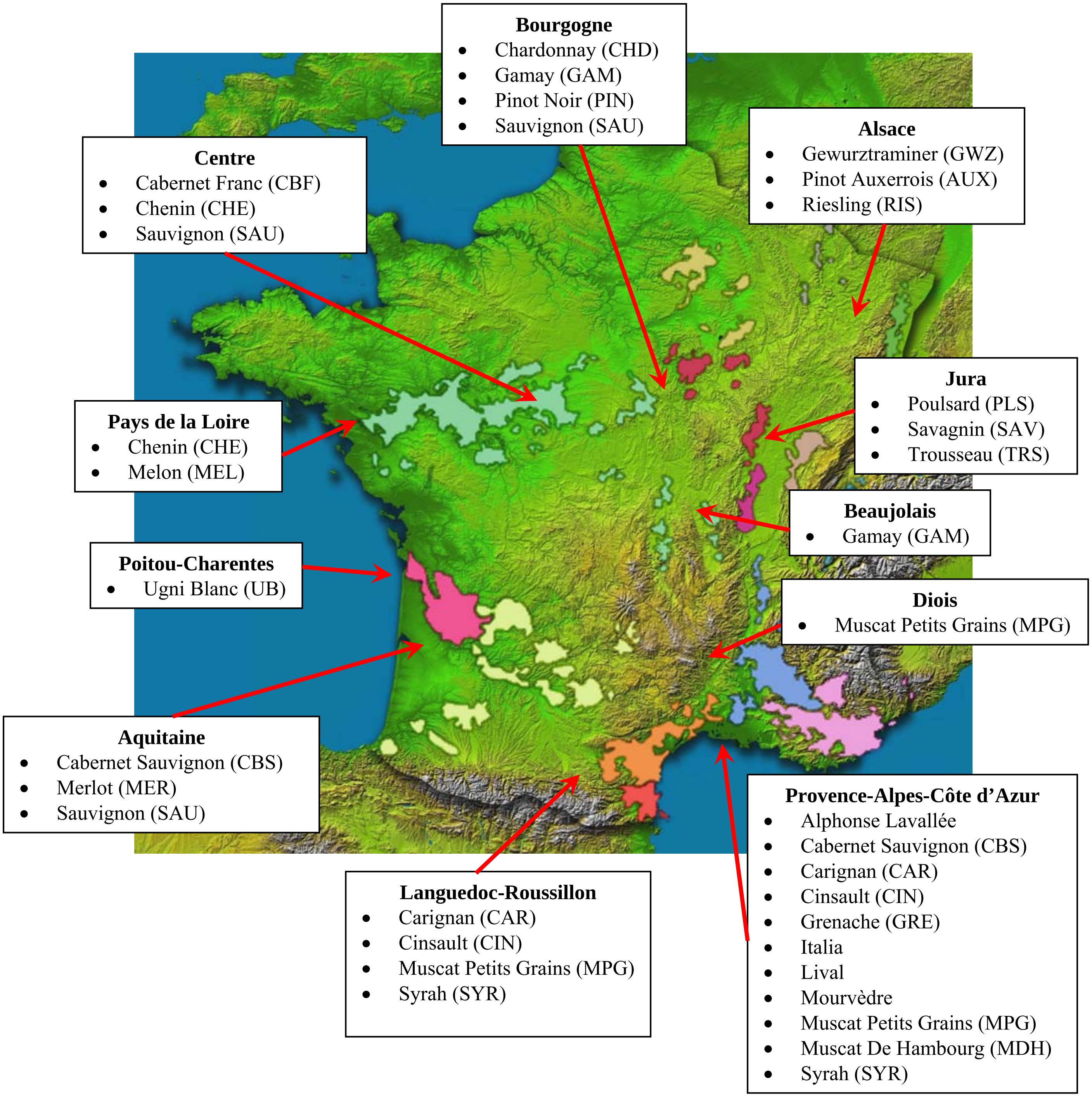


Figure 2. French vine-growing regions and vine varieties produced there. The tags that were created to identify the vine varieties that were part of the analyses are put within brackets. Source: Background image from NASA's Earth Observatory : The Topography of France[10]

[10] This image is available at the following address: http://earthobservatory.nasa.gov/Newsroom/NewImages/images.php?img_id=15360.

Vineyard of:		**Organism(s):**	
VINE GROWER			
Name First name:		**Company name:**	
Address:		**Phone number:**	
		Fax number:	
		Email:	
PARCEL			
Reference:			
Location:		**Vine variety (clone):**	
Name:		**Rootstock (clone):**	
		Year of planting:	
Pruning method:		**Area:**	
		Density of planting:	
Type of soil:		**Vigour:**	high, medium or low
WINTER TREATMENTS			
Sodium arsenite in 1999:	yes or no	**Other:**	
Sodium arsenite in 2000:	yes or no	**Other:**	
Sodium arsenite in 2001:	yes or no	**Other:**	
DISEASE PREVENTION MADE IN 2004			
Date of pruning:	accuracy of at least a two weeks time period	**Pre-pruning:**	yes or no
Management of pruning residues:	crushed or removed	**Cleaning dead wood:**	removed or not
Other grapevine trunk diseases prevention: “égourmandage”, “perçage des souches”, “traitement des plaies de taille”, “environnement de la parcelle”, …			
When did disease prevention start?			write down the year
OTHER REMARKS			

Figure 3. English translation of the questionnaire sent to any of the vine growers that were surveyed.

.

exact tests (Freeman and Halton, 1951) that we conducted[11], we were able to reduce the number of variables from 15 to 8 which were to be part of the statistical analysis and only selected those which were linked to the incidence rate of any of the two diseases or to the mortality rate.

These 8 variables are the following:

- vine-growing region,
- vine variety,
- age of the parcel,
- rootstock,
- management of pruning residues,
- pre-pruning,
- pruning method,
- number of times sodium arsenite was used between in 1999 and 2001.

The main objective of this case study is to derive the main trends of evolution of the grapevine trunk diseases from an epidemiological point of view.

4. Multiple Correspondence Analysis (MCA)

The dataset is made up of three tables, one per year. In this section, we consider the mean table of the three tables of data. As a consequence, the values of the incidences of the ED, the ES/BDA and the rate of mortality used in this part of the study, are the means of the values taken over the three years of the survey.

For a tutorial on basic and advanced multiple correspondence analysis (MCA), we point out the books of Greenacre (1984, 1993), Bry (1995, 1996) and of Greenacre and Blasius (2006).

[11] These analyses are detailed in the report by Lionel Fussler.

Application to the dataset generated by the National Grapevine Wood Diseases Survey

The reason why an MCA was chosen first is that this method allows the simultaneous handling of the two types of attributes –quantitative and qualitative through the recoding of the quantitative variables– and does not imply any a-priori assumptions on the variables, such a linear relationship between them or homoscedasticity.

The MCA was applied in two steps. At first we conducted a multiple correspondence analysis with only the following active variables:

- the incidence rate of eutypa dieback,
- the incidence rate of esca/bda,
- the mortality rate of the grapevines,
- the age of the parcel.

The main purpose of the MCA was to spotlight the dependencies between the grapevine trunk diseases and the mortality rate of the grapevines. As a consequence of the results of the bivariate tests, we had to include the age of the parcel to account for some of the variation of the mortality rate of the grapevines and of the incidence rates of the grapevine trunk diseases. The reasons why the vine variety (22 levels) and the vine-growing regions (11 levels) were not selected as active variables for this first analysis are mainly due to both their high number of levels and their unbalanced factor levels.

We then studied the influence of the remaining variables as supplementary qualitative variables: these variables were represented on the very display we built during the first step. Yet for the sake of readability, we plotted each of the supplementary qualitative variables on a different graph:

- vine-growing region,
- vine variety,
- number of times sodium arsenite was used between 1999 and 2001,
- rootstock,
- pre-pruning,
- management of pruning residues,
- pruning method.

In order to perfom these analyses and use an MCA, it was mandatory to derive factor levels by segmenting the range of continuous variables such as the incidence rates of the two diseases and the mortality rate.

These factor levels were designed not only to comply with the results of the bivariate tests but also to set up balanced factor levels for each variable in order to avoid any shrinkage of the axes by factor levels whose sizes are too low.

We now explain, following the recommendation by Savary et *al.* (1995), how to encode the quantitative variables into classes, *i.e.* define quantitative boundaries of classes and encode the values of the quantitative variables according to these boundaries. Performing this encoding process with care will allow the investigator to:

- define the boundaries so that they represent the (maximum possible) error made in the measurement of each variable (variables with low accuracy should be represented by a few, broad classes, while variables with high accuracy should be represented by a larger number of classes),
- link the definition of classes with key-values, thresholds, or any information that might be available beforehand.

The process of converting quantitative data into coded data is flexible, different options being available depending on the variable at hand and no statistical restriction made.

In our data set, the three rates and the age of the parcel were four continuous variables. The rates were categorized into three classes: low, medium and high not only on epidemiological grounds but also according to the dependencies between these three variables and the age of the parcel that were spotted during the preliminary analyses. So for every rate, three grades from 0 to 2 were considered (Table 2). The range of the age of the parcel was segmented into four intervals following the four stages identified in Section 3. Table 2 sums up the designed factor levels for the continuous variables.

Further analysis of the resulting coded data by means of contingency tables depends on how well populated the classes are, and therefore on the number of classes relative to the size of the sample. Therefore we checked the consistency of the former dependencies spotted when the coded variables were continuous and the new dependencies we now spotted using χ^2 tests and Fisher-Freeman-Halton exact tests between the coded variables and the factors of our dataset.

It is a common recommendation to maintain principal axes in such as way that their cumulated percentage of inertia is above 80 %. Yet it can be very difficult, when dealing with MCA, to understand to what phenomena higher order principal axes are related. Table 3 indicates that if we keep only the four first principal axes, we still retain 60 % of the total inertia which is rather good. Table 4 gives the values of the coordinates, contributions and squared cosine of all the levels of the active variables for the first four principal axes of the MCA. These axes are then displayed on Figure 4 (first and second principal axes), and on Figure 5 (third and fourth principal axes).

Table 2. Variables and their levels

Variable	*Symbol*	*Factor levels*	*Definition of factor levels*	*Unit*
Diseases and mortality				
Eutypa Dieback	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	%
Mortality	Mort	Mort0; Mort1; Mort2	Mort0: 0 ≤ Mort < 3; Mort1: 3 ≤ Mort < 10; Mort2: 10 ≤ Mort	%
Active variable				
Age of the parcel	Age	Age0; Age1; Age2; Age3	Age0: 0 ≤ Age < 15; Age1: 15 ≤ Age < 25; Age2: 25 ≤ Age < 40; Age3: 40 ≤ Age	years
Supplementary variables				
Vine-growing region	-	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: Provence-Alpes-Côte d'Azur; PCH: Poitou-Charentes; PDL: Pays de la Loire	none
Vine variety	-	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	none
Number of uses of sodium arsenite	Ars	Ars0; Ars1; Ars2; Ars3	Ars0: 0 time; Ars1: 1 time; Ars2: 2 times; Ars3: 3 times	none
Rootstock	-	101-14; 161-49; 3309C; 41B; R110; SO4	Name of the rootstock used	none
Pre-pruning	PreP	PreP0; PreP1	PreP0: no pre-pruning was done; PreP1: grapevine was pre-pruned	none
Pruning residues	-	Crushed; Burned; Removed	What happened to the pruning residues	none
Pruning method	-	Royat cordon; Gobelet; Guyot	Name of the method used to prune the grapevine	none

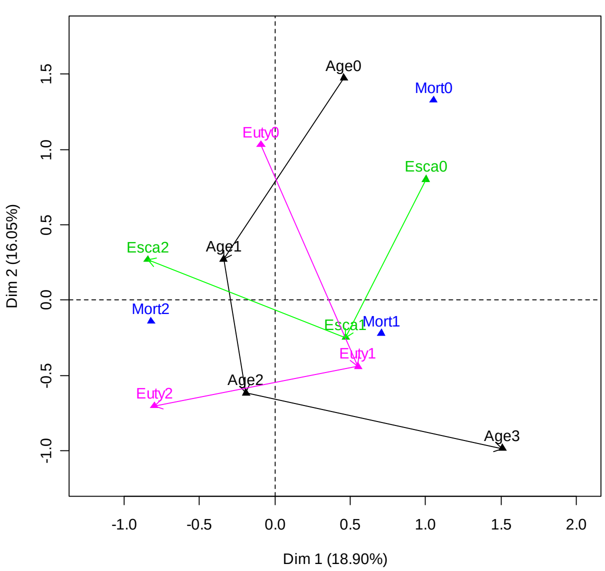


Figure 4. *Display of the first and second principal axes of the MCA.*

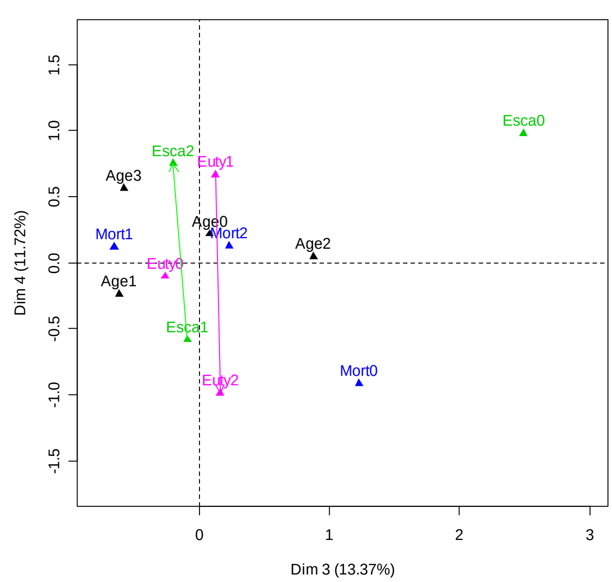


Figure 5. *Display of the third and fourth principal axes of the MCA.*

Table 3. Eigenvalues and percentages of inertia with respect to the principal axes of the MCA.

Axis	*Singular eigenvalue*	*Percentage of inertia Individual*	*Cumulated*	*Axis*	*Singular eigenvalue*	*Percentage of inertia Individual*	*Cumulated*
1	0.652	18.904	18.904	6	0.454	9.164	79.745
2	0.601	16.049	34.953	7	0.435	8.411	88.156
3	0.549	13.372	48.326	8	0.415	7.661	95.818
4	0.514	11.720	60.046	9	0.307	4.182	100.000
5	0.487	10.536	70.582				

Table 4. Coordinates, contributions and squared cosines of the levels of the active variables for the first four principal axes of the MCA.

	Mass	*Quality*	*Relative Inertia*	*Coord. Dim.1*	*Inertia Dim.1*	*Cosine² Dim.1*	*Coord. Dim.2*	*Inertia Dim.2*	*Cosine² 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
Age0	0.033	0.866	0.096	0.458	0.016	0.032	1.472	0.201	0.335
Age1	0.105	0.579	0.064	-0.340	0.029	0.084	0.270	0.021	0.053
Age2	0.087	0.654	0.073	-0.192	0.007	0.020	-0.617	0.091	0.202
Age3	0.025	0.902	0.100	1.508	0.132	0.248	-0.986	0.066	0.106
				Coord. Dim.3	*Inertia Dim.3*	*Cosine² Dim.3*	*Coord. Dim.4*	*Inertia Dim.4*	*Cosine² Dim.4*
Euty0				-0.263	0.020	0.036	-0.103	0.003	0.006
Euty1				0.124	0.005	0.011	0.667	0.174	0.314
Euty2				0.159	0.005	0.008	-0.985	0.225	0.313
Esca0				2.489	0.263	0.334	0.981	0.047	0.052
Esca1				-0.091	0.004	0.011	-0.584	0.185	0.454
Esca2				-0.200	0.013	0.024	0.749	0.201	0.341
Mort0				1.226	0.147	0.201	-0.913	0.093	0.112
Mort1				-0.653	0.140	0.277	0.119	0.005	0.009
Mort2				0.230	0.021	0.051	0.125	0.007	0.015
Age0				0.080	0.001	0.001	0.217	0.006	0.007
Age1				-0.613	0.131	0.273	-0.238	0.023	0.041
Age2				0.879	0.222	0.409	0.045	0.001	0.001
Age3				-0.579	0.027	0.037	0.566	0.030	0.035

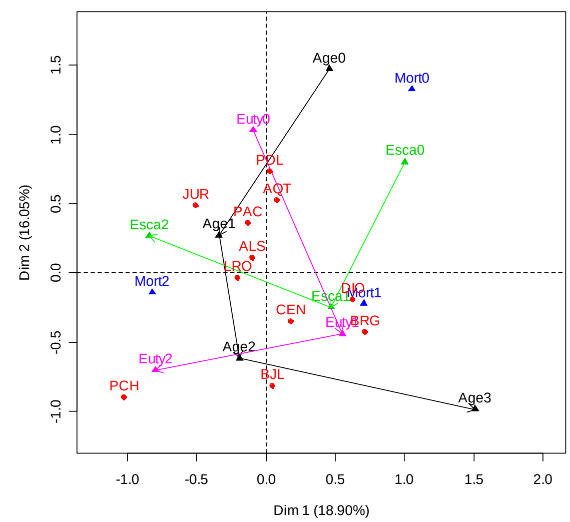
The analyses depicted here are the basis for the display of the supplementary qualitative variables. Since the supplementary variables do not contribute to the construction of the axes of the MCA, we will only provide their coordinates and squared cosine in Table 5. The levels of some of the supplementary variables are displayed on the first and second principal axes of the MCA, see Figure 6, 7 and 8.

The main results are the following:

- high mortality rates –Mort2– are associated with high incidence levels of the two trunk diseases –Esca2 and Euty2–,
- high incidence levels of eutypa dieback are associated with old grapevines –from 25 to 40 years old– and high incidence levels of esca decline/black dead arm

Table 5. Coordinates and Squared Cosines of the Levels of the Supplementary Variables

Variables and their levels	*1st principal axis*		*2nd principal axis*	
Vine-growing region	*Coordinates*	*Cosine²*	*Coordinates*	*Cosine²*
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
Sodium arsenite				
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
Rootstock				
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
Vine variety				
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
Pre-pruning				
PreP0	-0.097	0.010	0.020	0.000
PreP1	0.121	0.015	-0.025	0.001
Pruning residues				
Crushed	-0.156	0.026	0.060	0.004
Burned	0.586	0.211	-0.287	0.050
Removed	0.207	0.014	0.141	0.007
Pruning method				
Royat cordon	-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



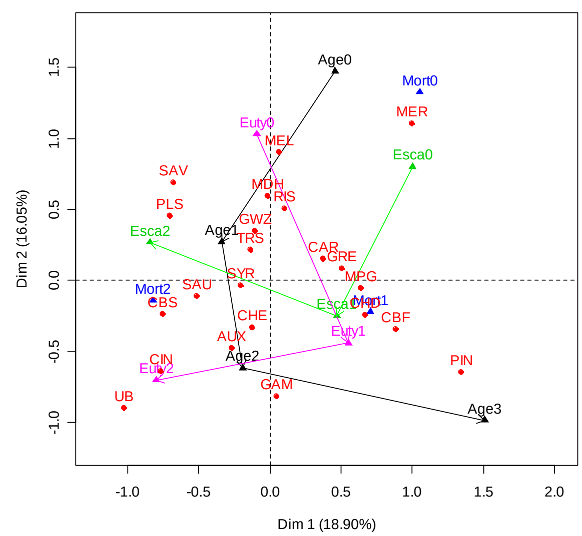


Figure 6. Display on the first and second principal axes of the MCA of the two supplementary variables, the vine-growing region (on top) and the vine variety (on bottom).

are associated with young grapevines –from 15 to 25 years old–,

- incidences seem to evolve in opposite ways: if the level of incidence of one of the two diseases is high then the level of incidence of the other one will be low,
- some vine-growing regions or vine varieties are linked to high level of vulnerabilities to both diseases, e.g. Poitou-Charentes and Euty2 or Poulsard and Esca2,

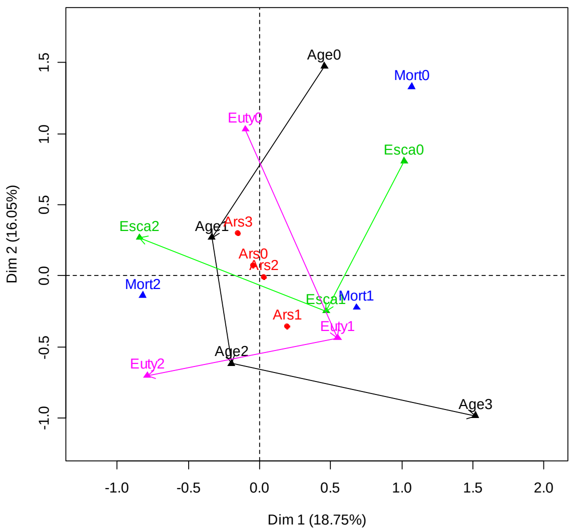


Figure 7. *Display on the first and second principal axes of the MCA of the number of uses of sodium arsenite.*

- the use of sodium arsenite does not seem to account for significant changes in the evolution of the diseases,
- the association between burning pruning residues and medium levels of incidence of the two diseases –Euty1 and Esca1–, is most probably due to the link between vine-growing regions and vine varieties and the fact that the burning technique to dispose of pruning residues is mainly used in Bourgogne.

5. Multiple Factor Analysis for Mixed Data

We just conducted an MCA analysis on the data and therefore we had to turn the quantitative variables –the age of the parcel, the rates of incidence and the rate of mortality- into qualitative ones by cutting the range of values into three classes or four classes. Since this step was handled with care this should have enabled us to spot even non linear associations between the variables we analyzed. Yet it is also possible to perform a multiple factor analysis for mixed data (FAMD) (Escofier, 1979, Pagès 2002, Pagès 2004), i.e. for a dataset with both quantitative and qualitative active variables.

There are two advantages and one possible drawback in applying this method. The first advantage is to be able to maintain an equal importance for any of the 4 quantitative variables or the qualitative variable which participate in the analysis as active variables. Indeed during the process of construction of the directions of maximal inertia the influence of all variables, quantitative as well as qualitative, will be balanced (p.97 Pagès 2004), and therefore will not depend on the number of the levels

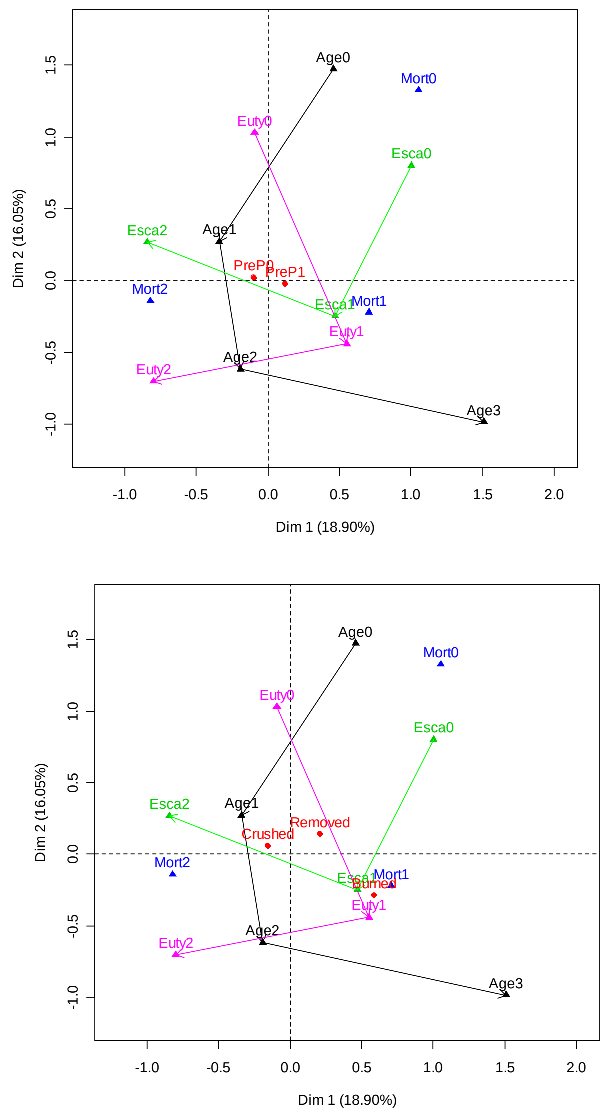


Figure 8. *Display on the first and second principal axes of the MCA of the two supplementary variables pre-pruning (top) and management of pruning residue (bottom).*

of the qualitative variable. The second advantage, following the recommendations of the article written by Pagès (2004, p. 94)), is to not have to encode continuous variables (here, there are four) as factors. The drawback is that we will not be able to spot any non linear association between the quantitative variables.

The reader may not be familiar with such techniques, so we briefly describe them following the article written by Pagès (2004).

5.1 Describing the method

5.1.1 Data and notation

Let us consider I units. Every unit i has a weight w_i such that $\sum_{i=1}^{I} w_i = 1$. These units are described by:

- Q quantitative variables; these variables are always assumed to have zero mean and are scaled to unit variance. The use of standardized variables is mandatory since we use both quantitative and qualitative active variables.
- J qualitative variables; the j th variable has K_j levels, so we have a total of $K = \sum_{j=1}^{J} K_j$ levels for all qualitative variables; we denote by w_{k_j} the proportion of the units that have level k_j.

The total number T of quantitative variables and of indicator functions for each qualitative level is equal to $Q+K$. Let x_{iq} denote the value of unit i for the q th variable, $q = 1 \ldots Q$, x_{ij} denote the level of unit i for the j th variable and let x_{ik_j} equal 1 if i has level k_j for variable j, else let x_{ik_j} equal zero.

5.1.2 Displaying the variables in R^I

Let us consider R^I the space of functions defined on I. The matrix D defined by the weights of the units provides this space with a diagonal metric. More precisely we have D equal to:

$$D(i,j) = \begin{cases} 0 & \text{if} \quad i \neq j \\ w_i & \text{if} \quad i = j \end{cases}.$$

As in a normalized principal component analysis (PCA), the variables will be represented by vectors whose lengths are equal to 1.

As in a MCA, the j th variable is represented by the set N_j of K_j vectors representing the K_j centered indicator functions of the levels of the j th variable. This set of vectors spans a vector space E_j of dimension $K_j - 1$ of functions that are centered and constant on every set of the partition defined by the levels of the j th variable. In order for N_j to inherit the same properties regarding inertia as in a MCA, we have to set the weight of the k_j th indicator function to $1/w_{k_j}$. This weight should be set to $1/\left(Q \cdot w_{k_j}\right)$ to have exactly the same inertia than in a MCA. Indeed in a MCA the inertias are averaged by the number of variables. Yet using these weights would lead to an inadequate property – averaging– since for a factorial analysis of mixed data qualitative variables are compared to quantitative ones whose inertias are not averaged. Some PCA software tools are not designed to assign weights to columns. As a consequence we prefer to divide the values of the k_j th indicator functions by $\sqrt{w_{k_j}}$.

5.1.3 Displaying the units in R^T

The dimension of the vector space R^T is equal to the sum of the number of quantitative variables, Q, and of the number of indicator functions, K. We provide this vector space with the usual Euclidean metric. The distance between two units i and j is then given by the following formula:

$$d^2(i,l) = \sum_{q \in Q} (x_{iq} - x_{lq})^2 + \sum_{j \in J} \sum_{k \in K_j} \frac{1}{w_{k_j}} (x_{ik_j} - x_{lk_j})^2,$$

with as a sub case the distance between a unit and the center of gravity of the all the units. This center of gravity is located at O, the origin of the vector space R^T, as soon as the variables are centered. This is an assumption we made on the quantitative variables at the very beginning of this section. As to the indicator functions that were derived from the qualitative variables and divided by $\sqrt{w_{k_j}}$ in order to assign adequate weights to the columns, the mean of the k_j th column is equal to $\sqrt{w_{k_q}}$. These renormalizations lead to the following formulas:

$$d^2(i,O) = \sum_{q \in Q} x_{iq}{}^2 + \sum_{j \in J} \sum_{k \in K_j} \left(\frac{x_{ik_j}}{\sqrt{w_{k_j}}} - \sqrt{w_{k_j}} \right)^2$$
$$= \sum_{q \in Q} x_{iq}^2 + \sum_{j \in J} \frac{1 - w_{j(i)}}{w_{j(i)}},$$

with $j(i)$ the level that the j th variable takes for the i th unit and $w_{j(i)}$ the corresponding proportion.

5.1.4 Graphical outputs

As in any factor analysis we display:

- The units by their projection on the inertia axes. We denote by $F_s(i)$ the projection of the i^{th} unit on the axis F_s of order s.
- The quantitative variables using their correlation coefficients with the axes F_s.
- The levels of the j^{th} qualitative variable using the centers of gravity of the corresponding units. We denote by $F_s(k_j)$ the projection on the axis F_s of order s of the center of gravity of the units featuring the k^{th} level of the j^{th} qualitative variable.

5.2 Applying FAMD to the data generated by the National Grapevine Wood Diseases Survey

In order to be able to put side by side the results we are about to obtain with those we obtained from the MCA, we will proceed in a somewhat similar way to that of the MCA. We first build the factorial axes using the four main variables. The first four active variables will remain the same and are quantitative: the incidence of eutypa dieback, the incidence of esca/BDA, the mortality of grapevine trunks and the age of the parcel. The fifth active variable is a qualitative variable chosen among all those we decided to include in the analysis and that were, up to now, used as supplementary variables: vine-growing region, vine variety, number of uses of sodium arsenite, rootstock, pre-pruning, pruning residues, pruning method.

5.2.1 *Vine-growing regions and vine varieties*

Since vine-growing regions and vine varieties are two strongly dependent variables we will analyze them together, see Figure 9, top and bottom panels.

A first point is that the grapevine mortality is close to being proportional to the sum of the vectors associated with the two diseases. A second point is that the diseases are opposed to each other, which agrees with the results we got with the MCA: the two diseases are likely to not completely –or not at all– evolve in the same way. The age of the parcel is still strongly linked with the second factor axis, as was found using the MCA.
We spot again, as with the MCA, an association between esca/BDA and the vine-growing region Jura or between Poitou-Charentes and eutypa. As to the associations between the vine varieties and the diseases, the Poulsard, the Savagnin and the Trousseau are close to the esca/BDA, as was found with the MCA.

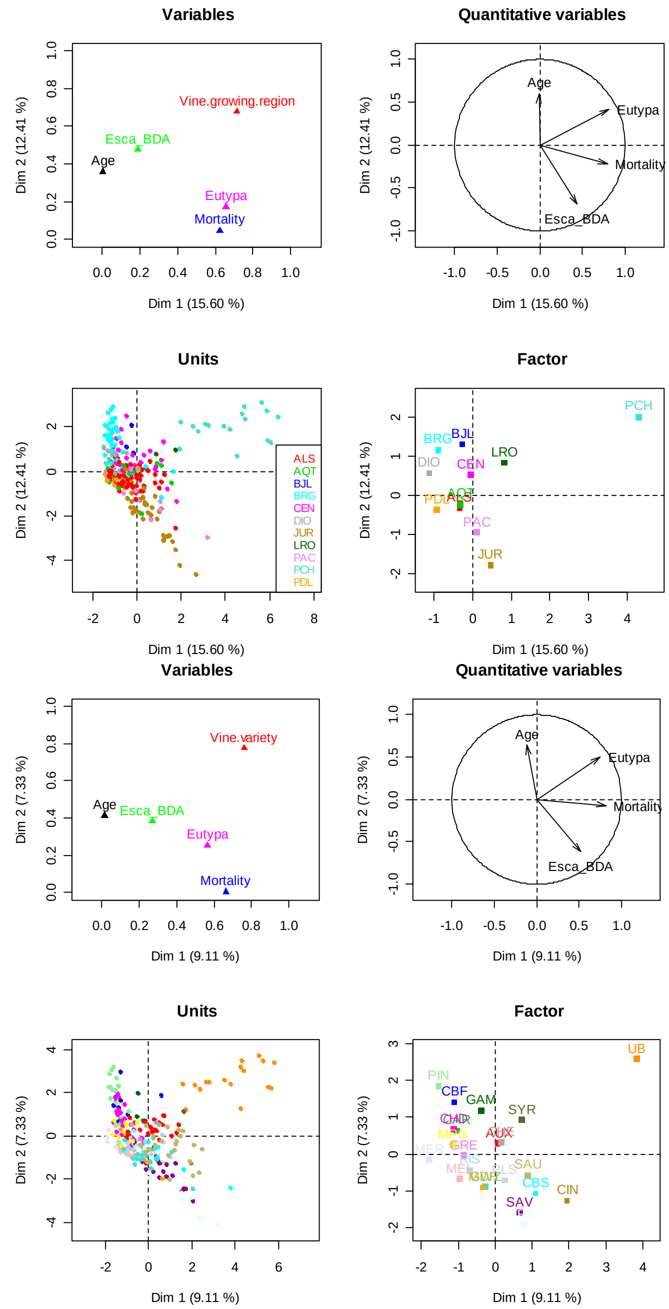


Figure 9. Display on the first and second principal axes of the FAMD with the two qualitative variables vine-growing region (top) and the vine variety (bottom).

5.2.2 *Sodium arsenite*

Figure 10 displays our results related to the number of uses of sodium arsenite.

The association between diseases and mortality is still strong as well as the location of the variable age of the parcel along the second factor axis. There is an opposition between the rise in mortality of the grapevines and the non-use or twice use of sodium arsenite as well as very close locations of high incidence rates of eutypa

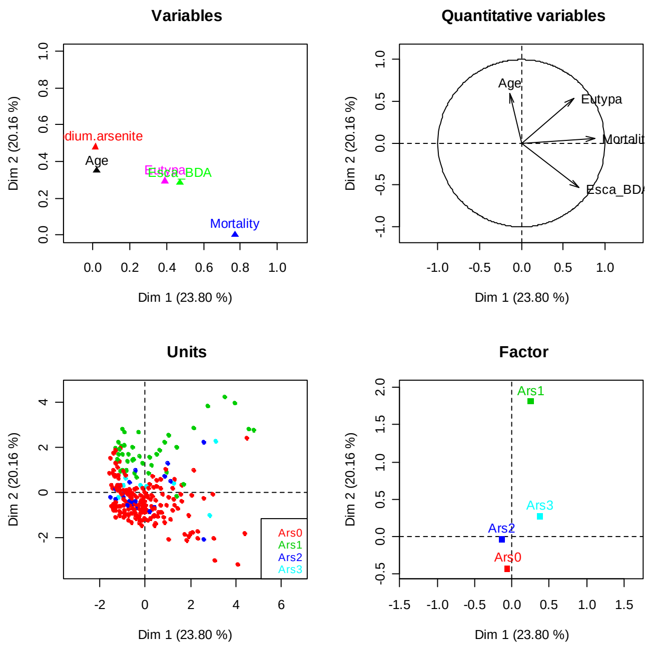


Figure 10. Display on the first and second principal axes of the FAMD with the qualitative variable given by the number of uses of sodium arsenite.

dieback and the three times repeated use of sodium arsenite. However only a few parcels were treated three times during the period from 1999 to 2001 and there is no evidence that the use of sodium arsenite accounts for a decrease in the incidence rate of eutypa dieback. This result is to be confirmed by other experiments. Another way to understand this relationship is that parcels with high incidence rates of eutypa dieback were treated every year by vine growers.

5.2.3 Pre-pruning and management of pruning residues

The pre-pruning and the management of pruning residues, two disease prevention policies, seemed to have no effect on the incidence rates of the two diseases and on the mortality rate of the grapevines when analyzed with the MCA. Is it still the case with the FAMD; see Figure 11.

First of all, we notice the same association pattern between the incidence of eutypa dieback, the incidence of esca/BDA, the mortality of the grapevines and the age of the parcel as those we spotted with the MCA: the higher the incidence of the eutypa dieback and of the esca/BDA, the higher the mortality rate.

We emphasize the fact that although there seems to be an opposition between the mortality of the grapevines or the esca/BDA and the pre-pruning of the wine ("PreP1"),

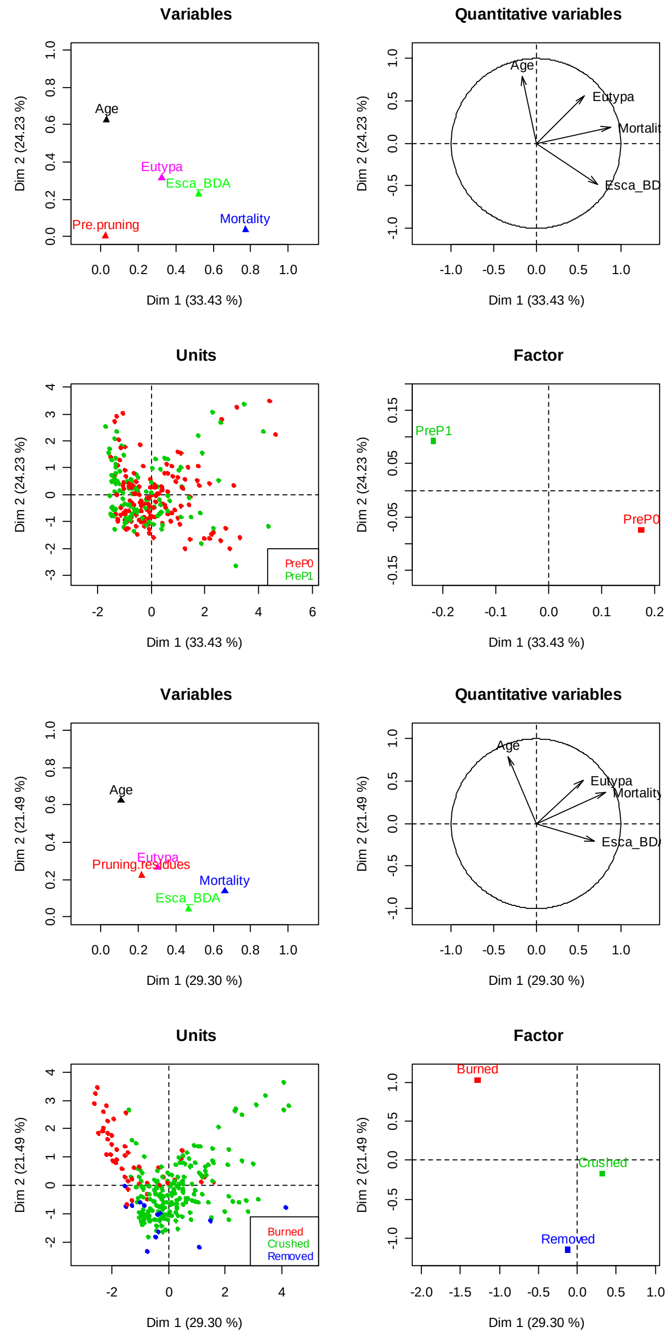


Figure 11. Display on the first and second principal axes of the FAMD with the qualitative variables pre-pruning (top panel) and management of pruning residues (bottom panel).

which was not spotted by the MCA, this result is not backed up by the plot of the association of the the PreP factor with the two first factor axes since these accounts for almost none of the variability of PreP factor.

Burning the pruning residues is opposed to the incidence of the diseases and to the mortality rate. It is more difficult to derive conclusions about the removal of the pruning residues since the number of parcels where it was used is very small (7 out a sample of 191 parcels).

6 Multi-table correspondence analysis

Multi-table correspondence analysis (CA) –also known as *K*–tables correspondence analysis– (Bry 1996, Escofier and Pagès 1998, Cazes 2004) will be used in order to study the yearly evolution of the incidences rates of the two diseases and of the mortality of the grapevines. Up to now, our analyses only focused on the mean values of these three variables over the three years: 2003, 2004 and 2005. As a consequence the time factor was not taken into account. The multi-tables correspondence analysis will enable us to check whether or not one of the years has a high impact on the average association patterns we found before. This is a matter of high concern since the summer of 2003 was very hot in France.

6.1 Description of the method

This statistical tool was applied to the data tables corresponding to the three years 2003, 2004 and 2005 for the same units and variables. We now describe more precisely the way we used multi-tables correspondence analysis. We begin with the analysis which does not take the age of the parcel into account:

- First we bind together in the same table the data for the incidence rates of the two diseases and of the mortality rate for the three years 2003, 2004 and 2005. We get a table with 321 rows –one for each of the 3*107 units– and three –EUTY_Tot, ESCA_Tot and MORT_Tot– columns. These three new variables will be the active ones. We then compute the Burt table B_{Tot} for this table which is the sum over the three years of the yearly Burt tables 2003, 2004 and 2005, B_{2003}, B_{2004} et B_{2005}.
- Each yearly table is made up with 107 rows and 3 columns –EUTY_0X, ESCA_0X and MORT_0X, where X is equal to 3 for the year 2003, 4 for the year 2004 and 5 for the year 2005.
- We lay on the right of the table B_{Tot} the three yearly Burt tables B_{2003}, B_{2004} and B_{2005} of the incidence of diseases and the mortality rate, each one with 13 rows and 13 columns. The columns of the three Burt tables B_{2003}, B_{2004} and B_{2005} with be used as supplementary variables in the upcoming correspondence analysis.
- We lay on the bottom of the table B_{Tot} the three yearly Burt tables B_{2003}, B_{2004} and B_{2005} of the incidence of diseases and the mortality rate, each one with 13 rows and 13 columns. The rows of the three Burt tables B_{2003}, B_{2004} and B_{2005} will be used as supplementary units in the upcoming correspondence analysis.
- Then we complete the empty cells of this global table B with zeros.
- Finally we performe the correspondence analysis of the table B as described in Table 6 : the 9 variables EUTY_Tot.0, EUTY_Tot.1, EUTY_Tot.2, ESCA_Tot.0 ESCA_Tot.1, ESCA_Tot.2, MORT_Tot.0, MORT_Tot.1 and MORT_Tot.2 were active variables and the 27 variables EUTY_03.0, EUTY_03.1, EUTY_03.2, EUTY_04.0, EUTY_04.2, EUTY_04.3, EUTY_05.0, EUTY_05.1, EUTY_05.2, ESCA_03.0, ESCA_03.1, ESCA_03.2, ESCA_04.0, ESCA_04.1, ESCA_04.2, ESCA_05.0, ESCA_05.1, ESCA_05.2, MORT_03.0, MORT_03.1, MORT_03.2, MORT_04.0, MORT_04.1, MORT_04.2, MORT_05.0, MORT_05.1 and MORT_05.2 were supplementary ones.

Table 6. How to perform a correspondence analysis of the table B.

		Actives Variables	Supplementary Variables		
Active Units		B_{Tot}	B_{2003}	B_{2004}	B_{2005}
Supplementary Units		B_{2003}	0	0	0
		B_{2004}	0	0	0
		B_{2005}	0	0	0

We then added the age of the parcel as an active variable and followed the same scheme of analysis.

6.2 Applying multi-table correspondence analysis to the data generated by the National Grapevine Wood Diseases Survey

We carried out two analyses: the first one does not take into account the age of the parcel and the second one does.

6.2.1 Ignoring the age of the parcel

Tables 7 and 8 display the numerical results of the CA. It is legitimate to restrict our analyses to the four principal axes since these axes account for 88 % of the total inertia.

The number of points on Figure 12 is so high that we decide to split it into several sub-diagrams for the sake of readability, see Figures 13 and 14.

For the esca/BDA the levels of the three years are close which means a stable pattern for the incidence rates through the three years of the survey.

For eutypa dieback and mortality, the two years 2004 and 2005 are close and (for all levels of the variable Euty and Mort) whereas the year 2003 stands apart. This seems to highlight a stable expression of the symptoms and of the mortality in 2004 and 2005 after a variation between 2003 and 2004. The mean incidence for eutypa-dieback was: 3.5% in 2003, 3.5% in 2004, 3.6% in 2005. As a consequence the multi-tables correspondence analysis spotted differences in the incidence patterns of eutypa dieback whilst the mean values of incidences where close to be equal. Yet the age of the parcel was not included in this analysis. This may have resulted in misleading results. We will check this matter in the following subsection.

Table 7. Eigenvalues and percentages of inertia with respect to the principal axes of the K–tables CA.

Axis	*Singular Eigenvalue*	*Percentage of inertia*	
		Individual	*Cumulated*
1	0.495	33.326	33.326
2	0.420	23.944	57.269
3	0.347	16.387	73.657
4	0.323	14.178	87.835
5	0.250	8.498	96.333
6	0.164	3.667	100.000

Table 8. Coordinates, contributions and square cosine of the levels of the active variables for the first four principal axes of the K–tables CA without taking the age of the parcel into account.

	Coord. Dim.1	*Coord. Dim.2*	*Coord. Dim.3*	*Coord. Dim.4*	*Inertia Dim.1*	*Cosine² Dim.1*	*Inertia Dim.2*	*Cosine² Dim.2*	*Inertia Dim.3*	*Cosine² Dim.3*	*Inertia Dim.4*	*Cosine² Dim.4*
EUTY_Tot.0	0.118	0.324	-0.210	-0.393	0.010	0.043	0.103	0.322	0.063	0.135	0.255	0.473
EUTY_Tot.1	0.180	-0.402	-0.318	0.763	0.012	0.035	0.083	0.177	0.076	0.111	0.503	0.637
EUTY_Tot.2	-0.519	-0.277	0.919	-0.017	0.077	0.204	0.031	0.058	0.494	0.641	0.000	0.000
ESCA_Tot.0	0.794	0.562	0.153	0.398	0.176	0.441	0.123	0.221	0.013	0.016	0.104	0.111
ESCA_Tot.1	0.113	-0.475	0.153	-0.183	0.009	0.036	0.214	0.634	0.032	0.066	0.054	0.095
ESCA_Tot.2	-0.751	0.419	-0.369	0.035	0.224	0.591	0.097	0.185	0.110	0.143	0.001	0.001
MORT_Tot.0	0.932	0.240	0.197	-0.088	0.316	0.786	0.029	0.052	0.029	0.035	0.007	0.007
MORT_Tot.1	-0.064	-0.619	-0.400	-0.167	0.002	0.006	0.234	0.524	0.143	0.219	0.029	0.038
MORT_Tot.2	-0.561	0.334	0.188	0.190	0.174	0.558	0.086	0.198	0.040	0.063	0.047	0.064
EUTY_03.0	0.510	0.254	-0.076	-0.389		0.175		0.043		0.004		0.102
EUTY_03.1	0.235	-0.545	-0.322	0.690		0.028		0.154		0.054		0.246
EUTY_03.2	-0.253	-0.293	0.967	-0.005		0.028		0.038		0.414		0.000
ESCA_03.0	0.749	0.621	0.187	0.310		0.236		0.162		0.015		0.040
ESCA_03.1	0.198	-0.551	0.008	-0.124		0.028		0.220		0.000		0.011
ESCA_03.2	-0.615	0.104	-0.490	0.354		0.194		0.005		0.123		0.064
MORT_03.0	0.938	0.256	0.208	-0.056		0.425		0.032		0.021		0.002
MORT_03.1	0.051	-0.750	-0.385	-0.070		0.001		0.314		0.083		0.003
MORT_03.2	-0.252	0.249	0.303	0.386		0.041		0.040		0.060		0.097
	Coord. Dim.1	*Coord. Dim.2*	*Coord. Dim.3*	*Coord. Dim.4*	*Inertia Dim.1*	*Cosine² Dim.1*	*Inertia Dim.2*	*Cosine² Dim.2*	*Inertia Dim.3*	*Cosine² Dim.3*	*Inertia Dim.4*	*Cosine² Dim.4*
EUTY_04.0	-0.029	0.307	-0.280	-0.391		0.001		0.073		0.060		0.118
EUTY_04.1	0.101	-0.251	-0.324	0.855		0.005		0.033		0.056		0.388
EUTY_04.2	-0.590	-0.207	0.896	0.017		0.152		0.019		0.350		0.000
ESCA_04.0	0.809	0.585	0.175	0.381		0.271		0.142		0.013		0.060
ESCA_04.1	0.030	-0.456	0.122	-0.322		0.001		0.155		0.011		0.077
ESCA_04.2	-0.780	0.477	-0.356	-0.028		0.308		0.115		0.064		0.000
MORT_04.0	1.011	0.386	0.218	-0.078		0.456		0.066		0.021		0.003
MORT_04.1	-0.091	-0.517	-0.431	-0.352		0.005		0.157		0.109		0.073
MORT_04.2	-0.616	0.418	0.120	0.128		0.238		0.110		0.009		0.010
EUTY_05.0	-0.091	0.418	-0.254	-0.400		0.006		0.125		0.046		0.114
EUTY_05.1	0.162	-0.318	-0.308	0.794		0.014		0.054		0.051		0.339
EUTY_05.2	-0.633	-0.321	0.905	-0.050		0.170		0.044		0.347		0.001
ESCA_05.0	0.895	0.360	0.020	0.670		0.309		0.050		0.000		0.173
ESCA_05.1	0.083	-0.401	0.356	-0.132		0.005		0.118		0.093		0.013
ESCA_05.2	-0.758	0.446	-0.350	0.011		0.298		0.103		0.064		0.000
MORT_05.0	0.860	0.104	0.166	-0.139		0.367		0.005		0.014		0.010
MORT_05.1	-0.195	-0.561	-0.381	-0.070		0.022		0.182		0.084		0.003
MORT_05.2	-0.679	0.299	0.190	0.141		0.284		0.055		0.022		0.012

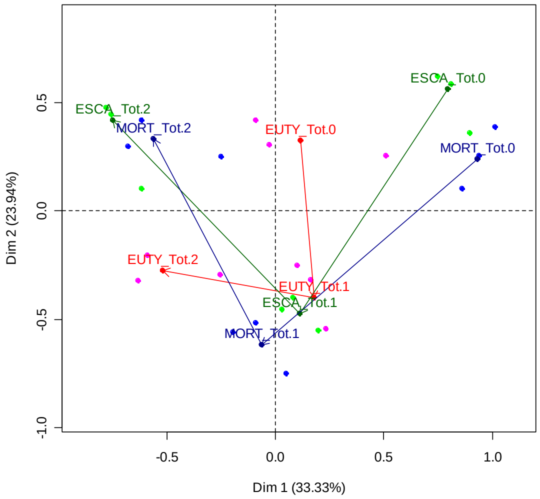


Figure 12. Display on the first and second principal axes of the K–tables CA of the means per year without taking into account the age of the parcel.

Figure 12 displays the graphical output of the CA. The analysis of the second factor plane highlights a stable opposition through time for the increase of the two diseases.

6.2.2 *Taking the age of the parcel into account*

We added the age of the parcel as an active variable, built the global Burt table B^{age} using the three yearly ones B^{age}_{2003}, B^{age}_{2004} and B^{age}_{2005} and performed a CA of the global Burt table B^{age} using the 13 variables AGE.0, AGE.1, AGE.2, AGE. 3, EUTY_Tot.0, EUTY_Tot.1, EUTY_Tot.2, ESCA_Tot.0 ESCA_Tot.1, ESCA_Tot.2, MORT_Tot.0, MORT_Tot.1 and MORT_Tot.2 as active variables and the 27 variables EUTY_03.0, EUTY_03.1, EUTY_03.2, EUTY_04.0, EUTY_04.1, EUTY_04.2, EUTY_05.0, EUTY_05.1, EUTY_05.2, ESCA_03.0, ESCA_03.1, ESCA_03.2, ESCA_04.0, ESCA_04.1, ESCA_04.2, ESCA_05.0, ESCA_05.1, ESCA_05.2, MORT_03.0, MORT_03.1, MORT_03.2, MORT_04.0, MORT_04.1, MORT_04.2, MORT_05.0, MORT_05.1 and MORT_05.2 as supplementary variables. The high number of points on Figure 15 leads us to split it into several sub-diagrams (Figures 16 and 17).

The values of the inertia led us again to consider only the first four axes, accounting for about 72 % of the total inertia of the whole dataset. However, we emphasize that the fifth axis could be interesting to study as well.

For all the factors and for any given level of these factors, the annual points of these levels lie close and the pattern they draw remains somewhat similar for the three years of the study 2003, 2004 and 2005. We note (see Figures 16 and 17) that the path of the mortality rate stays stable during the three years while the paths of the two incidence rates change slightly during this period. These variations could not be spotted by only looking at the incidence rates which were almost constant for the three

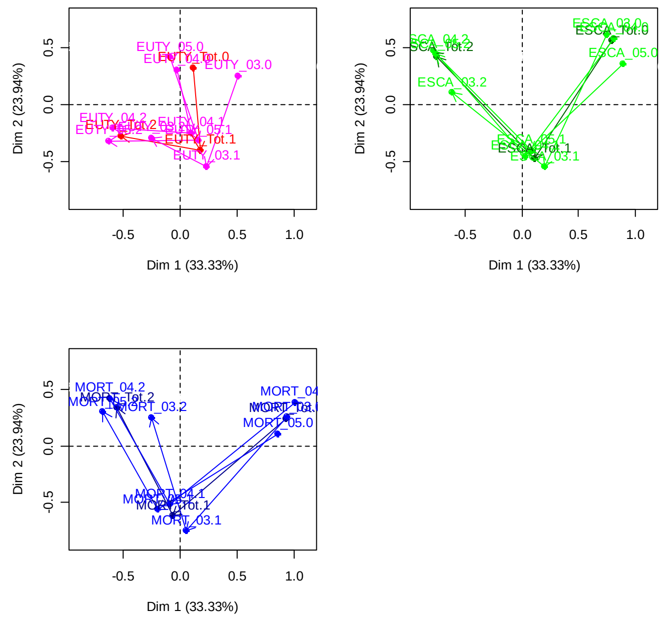


Figure 13. Display on the first and second principal axes of the K–tables CA of the means per year without taking into account the age of the parcel for the eutypa dieback disease only.

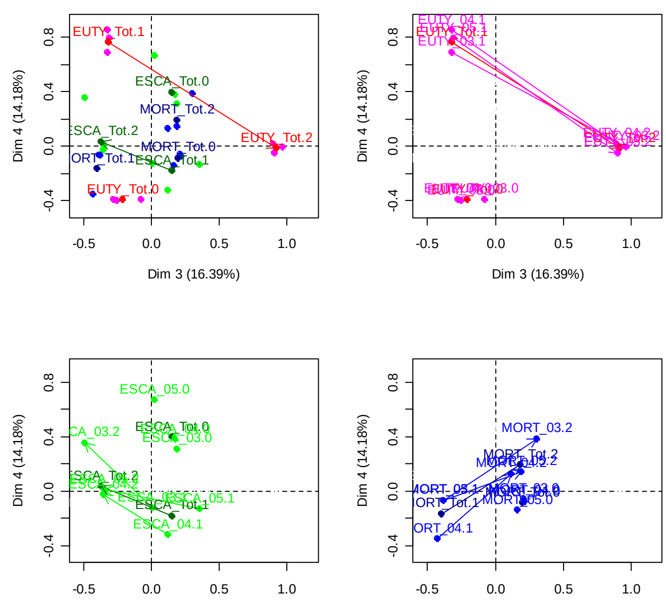


Figure 14. Display on the third and fourth principal axes of the K–tables CA of the means per year without taking into account the age of the parcel for the eutypa dieback disease only.

Table 9. Eigenvalues and percentages of inertia with respect to the principal axes of the K–tables CA.

Axis	*Singular Eigenvalue*	*Percentage of inertia Individual*	*Cumulated*	*Axis*	*Singular Eigenvalue*	*Percentage of inertia Individual*	*Individual*
1	0.439	30.190	30.190	6	0.217	7.335	88.629
2	0.332	17.265	47.456	7	0.183	5.225	93.854
3	0.289	13.043	60.498	8	0.166	4.301	98.155
4	0.269	11.322	71.821	9	0.109	1.845	100.000
5	0.246	9.473	81.294				

Table 10. Coordinates, contributions and square cosine of the levels of the active variables for the first four principal axes of the K–tables CA taking the age of the parcel into account.

	Coord. Dim.1	*Coord. Dim.2*	*Coord. Dim.3*	*Coord. Dim.4*	*Inertia Dim.1*	*Cosine² Dim.1*	*Inertia Dim.2*	*Cosine² Dim.2*	*Inertia Dim.3*	*Cosine² Dim.3*	*Inertia Dim.4*	*Cosine² Dim.4*
AGE.0	1.236	-0.348	-0.011	-0.012	0.222	0.633	0.031	0.050	0.000	0.000	0.000	0.000
AGE.1	-0.067	-0.156	-0.364	-0.161	0.002	0.013	0.023	0.069	0.374	0.374	0.038	0.073
AGE.2	-0.328	0.130	0.213	0.292	0.053	0.245	0.015	0.038	0.103	0.103	0.113	0.194
AGE.3	0.180	0.656	0.866	-0.509	0.004	0.012	0.082	0.152	0.266	0.266	0.075	0.092
EUTY_Tot.0	0.253	-0.290	-0.046	-0.219	0.043	0.243	0.099	0.321	0.008	0.008	0.086	0.183
EUTY_Tot.1	-0.039	0.540	0.355	-0.050	0.001	0.002	0.179	0.412	0.179	0.179	0.002	0.004
EUTY_Tot.2	-0.567	0.018	-0.343	0.599	0.088	0.311	0.000	0.000	0.114	0.114	0.262	0.347
ESCA_Tot.0	0.714	-0.059	0.460	0.383	0.135	0.453	0.002	0.003	0.188	0.188	0.104	0.131
ESCA_Tot.1	-0.022	0.339	-0.315	0.022	0.000	0.002	0.131	0.428	0.368	0.368	0.001	0.002
ESCA_Tot.2	-0.464	-0.540	0.216	-0.306	0.081	0.296	0.193	0.402	0.064	0.064	0.095	0.129
MORT_Tot.0	0.861	-0.017	-0.160	0.224	0.257	0.802	0.000	0.000	0.028	0.028	0.046	0.054
MORT_Tot.1	-0.155	0.439	-0.132	-0.365	0.010	0.043	0.141	0.349	0.032	0.032	0.149	0.242
MORT_Tot.2	-0.442	-0.338	0.210	0.143	0.103	0.440	0.105	0.256	0.099	0.099	0.029	0.046
EUTY_03.0	1.417	-0.296	0.076	0.166		0.538		0.024		0.002		0.007
EUTY_03.1	0.110	0.025	-0.383	-0.145		0.009		0.000		0.105		0.015
EUTY_03.2	-0.158	0.389	0.151	0.363		0.016		0.099		0.015		0.087
ESCA_03.0	0.303	0.725	0.881	-0.491		0.024		0.135		0.199		0.062
ESCA_03.1	0.518	-0.126	-0.115	-0.084		0.189		0.011		0.009		0.005
ESCA_03.2	-0.023	0.639	0.253	-0.080		0.000		0.234		0.037		0.004
MORT_03.0	-0.408	0.120	-0.330	0.736		0.081		0.007		0.053		0.263
MORT_03.1	0.699	-0.155	0.402	0.408		0.232		0.011		0.077		0.079
MORT_03.2	0.005	0.436	-0.285	-0.049		0.000		0.145		0.062		0.002
	Coord. Dim.1	*Coord. Dim.2*	*Coord. Dim.3*	*Coord. Dim.4*	*Inertia Dim.1*	*Cosine² Dim.1*	*Inertia Dim.2*	*Cosine² Dim.2*	*Inertia Dim.3*	*Cosine² Dim.3*	*Inertia Dim.4*	*Cosine² Dim.4*
EUTY_04.0	-0.471	-0.212	0.226	-0.403		0.123		0.025		0.028		0.090
EUTY_04.1	0.822	0.009	-0.114	0.245		0.359		0.000		0.007		0.032
EUTY_04.2	-0.141	0.586	-0.150	-0.301		0.012		0.210		0.014		0.055
ESCA_04.0	-0.255	-0.128	0.281	0.298		0.045		0.011		0.055		0.062
ESCA_04.1	1.105	-0.353	-0.004	-0.074		0.376		0.038		0.000		0.002
ESCA_04.2	-0.128	-0.296	-0.337	-0.204		0.012		0.062		0.081		0.030
MORT_04.0	-0.377	-0.001	0.235	0.200		0.097		0.000		0.038		0.027
MORT_04.1	0.188	0.590	0.888	-0.473		0.009		0.090		0.204		0.058
MORT_04.2	0.118	-0.305	-0.012	-0.275		0.011		0.074		0.000		0.060
EUTY_05.0	-0.048	0.439	0.513	-0.008		0.001		0.111		0.152		0.000
EUTY_05.1	-0.590	-0.057	-0.322	0.544		0.170		0.002		0.051		0.145
EUTY_05.2	0.718	-0.051	0.466	0.349		0.240		0.001		0.101		0.057
ESCA_05.0	-0.054	0.287	-0.324	-0.025		0.002		0.064		0.082		0.000
ESCA_05.1	-0.471	-0.610	0.189	-0.307		0.125		0.210		0.020		0.053
ESCA_05.2	0.978	-0.112	-0.117	0.260		0.453		0.006		0.006		0.032
MORT_05.0	-0.127	0.322	-0.183	-0.441		0.010		0.066		0.021		0.124
MORT_05.1	-0.445	-0.430	0.229	0.086		0.133		0.124		0.035		0.005
MORT_05.2	1.187	-0.395	-0.104	-0.128		0.406		0.045		0.003		0.005

years –2003, 2004 and 2005–; for instance the mean incidences, for the eutypa were 3.5 % in 2003, 3.5 % in 2004 and 3.6 % in 2005.

The analysis of the second factor plane highlights a steady opposition through time for the increase of the two diseases.

Other procedures of data analysis were used to study this dataset, namely Multiple Factor Analysis (MFA) (Escofier and Pagès 1998) and the ACT–STATIS method (Lavit, Escoufier, Sabatier and Traissac 1994), using the FactoMineR (Husson, Lê and Mazet 2007) and the ade4 (Chessel, Dufour and Thioulouse 2004) packages for R (Team 2007). The results we obatined this way were similar to those of the *K*–tables CA.

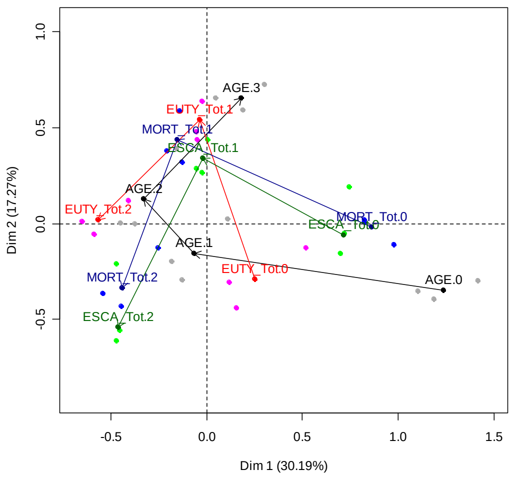


Figure 15. Display on the first and second principal axes of the K–tables CA of the means per year taking into account the age of the parcel.

7 Conclusion

This case study highlights the interest in the joint use of several statistical procedures. Its main objective is to provide an example of the use of adequate tools to analyze quantitative, qualitative and dynamic data. Moreover from a biological point of view, the associations between the variables of our survey were hardly studied before. We aimed to deeply explore the dataset by emphasizing the analysis of the consequences of the banning of sodium arsenite on the incidence rates of the diseases and the mortality rate as well as on their evolutions.

The multiple correspondence analysis allowed us to depict the associations between the main variables. The factorial analysis of mixed data confirmed these results and enabled us to spot other associations between the variables. The multi-table correspondence analysis took time into account and therefore provided us an analysis of the evolution of the two diseases and of the mortality rate.

One of the main results of these analyses is that the disease rates depend highly on the vine variety as well as on the vine-growing region. We also shed light on the fact that eutypa dieback is mainly linked with the age of the grapevine whereas higher esca/BDA

Other statistical methods were used to investigate this dataset from a modeling point of view, which will be the core of a upcoming article. For instance, both binary and ordinal logistic regressions models were fitted and the results they highlighted do not depart from those we got with the factor analyses we conducted throughout this analysis.

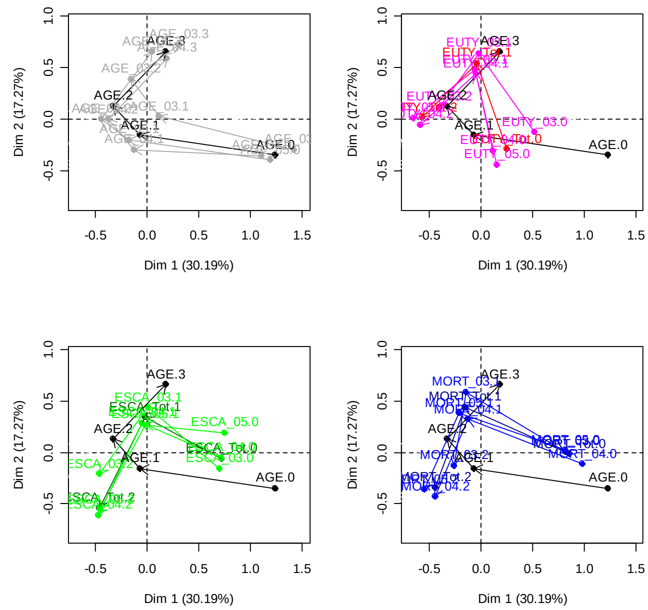


Figure 16. Display on the first and second principal axes of the K–tables CA of the means per year taking into account the age of the parcel.

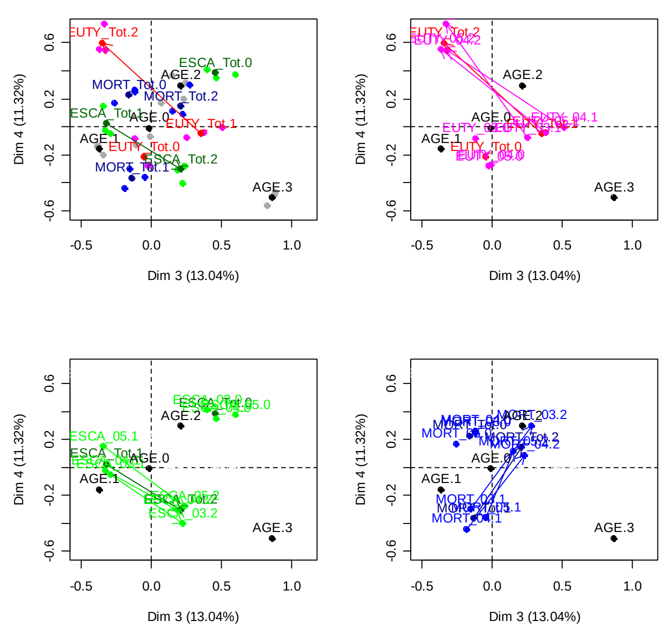


Figure 17. Display on the third and fourth principal axes of the K–tables CA of the means per year taking into account the age of the parcel.

Acknowledgements

The authors are grateful to Editor-in-Chief Dominique Haughton and two referees for their helpful and valuable comments.

Use of the dataset

Permission is granted to use the dataset generated by the National Grapevine Wood Diseases Survey to Academic purposes only. Any scientific research works or results based on the use of the dataset generated by the National Grapevine Wood Diseases Survey must be authorized by the French Agriculture Office; please write to Jacques.grosman@agriculture.gouv.fr.

Corespondence: fbertran@math.u-strasbg.fr

REFERENCES

Afresti, A. (1990), *Categorical data analysis*, Wiley, New York.

Bry, X. (1995), *Analyses factorielles simples*, Economica, Paris.

Bry, X. (1996), *Analyses factorielles multiples*, Economica, Paris.

Cazes, P. (2004), Quelques méthodes d'analyse factorielle d'une série de tableaux de données, *Revue MODULAD*, Paris.

Chessel, D., A.-B Dufour and J. Thioulouse. (2004), The ade4 package –I– One-table methods, ***R** News*, 4, 5–10.

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. and J. Pagès. (1998), *Analyses factorielles simples et multiples*, 3rd edition, Dunod, Paris.

Fox, J. and with contributions from M. Ash, T. Boye, S. Calza, A. Chang, P. Grosjean, R. Heiberger, G.J. Kerns, R. Lancelot, M. Lesnoff, S. Messad, M. Maechler, D. Putler, M. Ristic and P. Wolf. (2007), *Rcmdr:* **R** *Commander*, http://www.r-project.org, http://socserv.socsci.mcmaster.ca/jfox/Misc/Rcmdr/.

Freeman, G.H. and J.H. Halton. (1951), Note on an exact treatment of contingency, goodness of fit and other problems of significance, *Biometrika*, 38, 141–149.

Greenacre, M. J. (1984), *Theory and Applications of Correspondence Analysis*, Academic Press, New-York.

Greenacre, M. J. (1993), *Correspondence Analysis in Practice*, 1st edition, Academic Press, New-York.

Greenacre, M. J. and J. Blasius. (2006), *Multiple Correspondence Analysis and related methods*, Chapman-Hall, Boca-Raton.

Husson, F., S. Lê and J. Mazet. (2007), *FactoMineR: Factor Analysis and Data Mining with* **R**, http://factominer.free.fr, http://www.agrocampus-rennes.fr/math/.

Lavit, Ch., Y. Escoufier, R. Sabatier and P. Traissac. (1994), The ACT (STATIS Method), *Computational Statistics and Data Analysis*, 18, 97–119.

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.

Savary, S., L. V. Madden, J. C. Zadoks and H. W. Klein-Gebbinck. (1995), Use of Categorical Information and Correspondence Analysis in Plant Disease Epidemiology, *Advances in Botanical Research* incorporating *Advances in Plant Pathology*, **21,** 213–240.

Siegel, S. and N. J. Castellan. (1988), *Nonparametric Statistics for the Behavioural Sciences*, 2nd edition, McGraw-Hill Inc, New York.

R Development Core Team (2007), **R**: *A Language and Environment for Statistical Computing*, **R** Foundation for Statistical Computing, Vienna, http://www.R-project.org.