Trevisani, Causin, Montecchio, Kastner, Heumann:

Analysing the relationship between ectomycorrhizal infection and forest decline using marginal models

Sonderforschungsbereich 386, Paper 143 (1999)

Analysing the relationship between ectomycorrhizal infection and forest decline using marginal models

M. Trevisani* R. Causin§ L. Montecchio§ C. Kastner **
C. Heumann**

March 10, 1999

Abstract

This statistical survey originates from the problem of discovering which relationship exists between root ectomycorrhizal infection and health status of forest plants. The sampling scheme takes observations from roots that come from sectors around the tree resulting in a hierarchical association structure of the observations. Marginal regression models are used to analyze the mean effect of the ectomycorrhizal state on a response variable proxy for the health degree of the plants.

Keywords: ectomycorrhizal infection, forest decline, generalized estimating equations, likelihood methods, marginal models

1 Introduction

This statistical survey originates from the problem of discovering which relationship exists between root ectomycorrhizal infection and health status of forest plants. The ectomycorrhiza is a symbiosis involving fine root apexes of plants and some species of fungi. It is characterized by an hyphal mantle outside the apex surface and an internal, extracellular net of hyphae, with morphological different features depending on the fungus and the plant species involved. This symbiosis gives to the plant an improvement of the up-take of water and nutrients and of resistance to biotic and abiotic stresses. Therefore the decline of the plant is expected when the proportion of the ectomycorrhized tips decreases (Causin, Montecchio and Mutto Accordi, 1996).

---

*Dept. of Statistical Sciences, Universita' di Padova, via San Francesco 33, 35121 Padova, Italy

§Institute of Patologia Vegetale, Universita' di Padova, stra. Romea 16, 35020 Legnaro (Padova), Italy

**Institute of Statistics, LMU München, Ludwigstr. 33, 80539 München, Germany

The study was conducted in a coastal, holm oak (*Quercus ilex* L.) wood in the Veneto region (Mesola forest, northeast Italy) where a large number of trees are declining in varying degrees. The survey was performed on coeval plants, growing in the same enviromental conditions. The trees were divided in three classes of decline identified by means of objective visual criteria such as defoliation degree, discolored leaves, epicormic twigs, dead twigs and branches, growing anomalies, not attributable to parasites, silvicultural damages or climatic events. In particular the ordered damage levels are a not-decline control class (class 1), an intermediate one (class 2) and, at last, a decline one (class 3).

The sample was collected as follows: for each class five trees were randomly selected, and for each plant the area below the crown was divided in six equal sectors. For each sector a cylindrical soil core (Ø 8 cm, height 10 cm) was collected beneath the little layer at a distance of approximately 1 meter from the collar. Fifteen roots were randomly drawn from each core and each root was cut after fifteen apexes were counted in subsequence on it, starting from the distal part. A scheme of the data structure is given in figure 1.

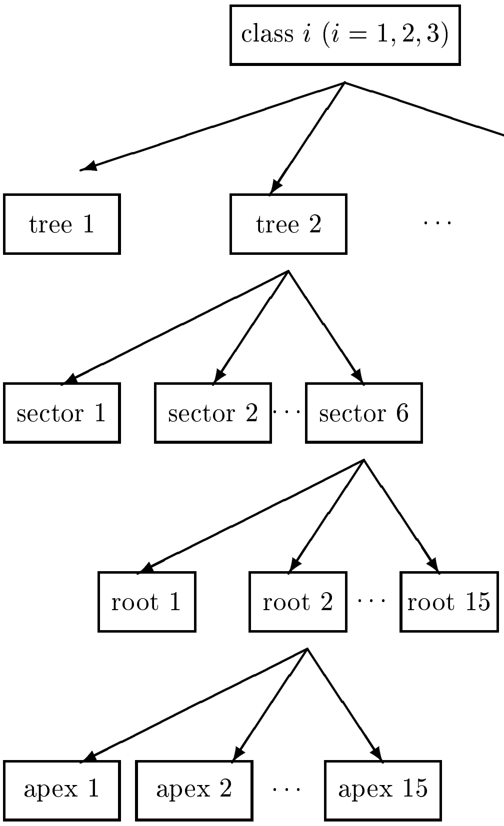

Figure 1: Data structure

The whole data set consists of 3 classes × 5 trees × 6 sectors × 15 roots × 15 apexes. That is 20 250 apexes altogether. Each of them was checked for the presence of ectomycorrhization, registering the vitality and the morphotype of the found ectomycorrhizae. The vitality of unmycorrhized apexes was also recorded.

The scope of this study is to analyze the relationship between the vitality of the apexes, presumable proxy for the plant health status, and the presence of ectomycorrhization. As there are further variables as length of the fragment and the number of its ramifications which should be also included, we can use a regression model to analyze the relationship.

Regression models—as the linear regression model for continuous response variables and the logistic regression model for binary response variables—assume that the observations are independent. As seen in figure 1 this assumption has to be violated in this case, because each 15 apexes cluster belongs to the same root, each 15 roots cluster is from the same sector and so on. Several methods for handling correlated data exist. For an overview see e. g. Fahrmeir and Tutz (1994).

In this paper we use marginal regression models to fit the data. Because a well stated hypothesis of scientific interest exists on the mean process, the influence of ectomycorrhizal state on the marginal probability of a response variable, proxy for the decline degree, is of prime interest. The correlation structure either of univariate or multivariate variables, considered up to now, can be a second purpose of analysis.

In section 2 we describe the data. Section 3 gives a short description of the three methods for estimating marginal regression models we used. In section 4 the results are presented. The results are discussed in section 5.

2 Data description

An exploratory analysis of the data was conducted to detect the possible difformity of ectomycorrhizal infection and its effect in the three classes of decline. First sample probabilities per apex were calculated (table 1). As for ectomycorrhizal phenomenon the values don't look as they were expected: the rate of infection is almost the same in the three classes. Instead the vitality of apexes varies from nearly 60% in the healthy class to 18% in the diseased one.

	Class	1	2	3
Mycorrhizae	Prob.	**0.285037**	**0.2441481**	**0.2598519**
	s.e.	0.005495063	0.005229081	0.005338294
Vitality	Prob.	**0.5874074**	**0.4265185**	**0.1863704**
	s.e.	0.005992537	0.006020172	0.004740041

Table 1: Sample probabilities per apex

Noticed ectomycorrhizal rate is not a discriminant feature but vitality rate is, the relationship between vitality and ectomycorrhizal spread among the apexes becomes the key point from now on. Contingency tables and conditional distributions (fig. 2) show a clear positive association between the two variables:

probability of dying is higher than that of surviving for a not-mycorrhizal apex, getting almost sure death in the diseased class, probability of living is nearly one for a mycorrhizal apex in the first two classes and is more than 0.5 in the third one. An immediate observation is that ectomycorrhizal portion of a radical system is usually hale and hearty (by a microscope observation ectomycorrhizal apexes are bright and coloured) and is the last surviving in dying trees.

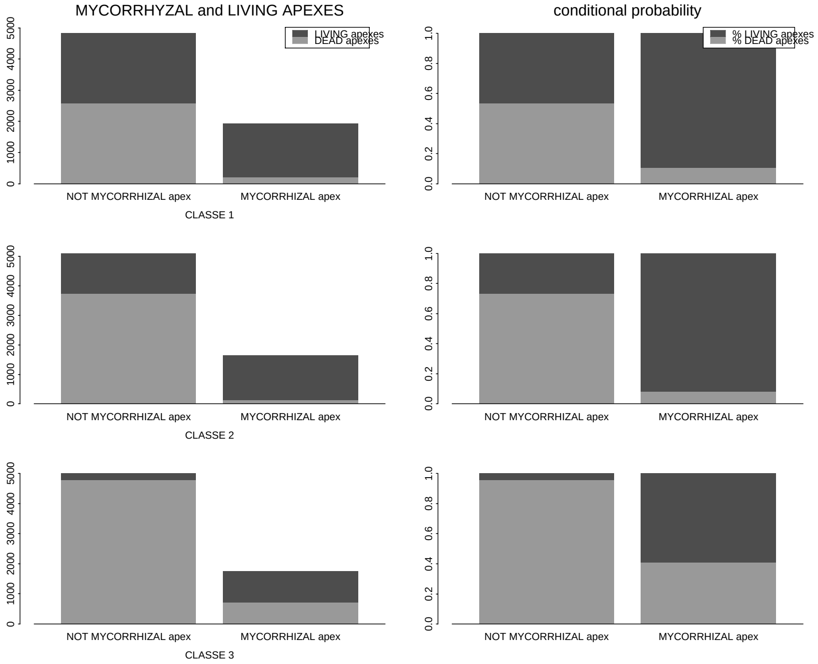

Figure 2: Conditional distribution of vitality given ectomycorrhizal status of apex

Mycorrhizae distribution is now examined. The frequency distribution of the number of ectomycorrhizae in every fifteen apexes-root has a shape considerably different from that of a standard binomial with the same parameters (fig. 3). Overdispersion in binary data can provide a general explanation of atypical distributions like this.

Sources of overdispersion can be heterogeneity of clusters—in this case trees, sectors as well as roots—and ignorance of dependence of units within clusters. With respect to the dependence of units, apexes infect each other within the small root and their dependence can be even seen as a longitudinal correlation. In fact ectomycorrhizal diffusion takes mainly place in two ways: during the growth of root and during its life. In the first case a temporal order in infection of subsequent apexes exists, in the last case a general spatial correlation is supposed to act. As regards the vitality distribution the same remark on overdispersion can be repeated yet noticing an emphasized dychotomous behaviour in the peaks of 0 and 15 apexes. For that summary of the two binomial variables into four

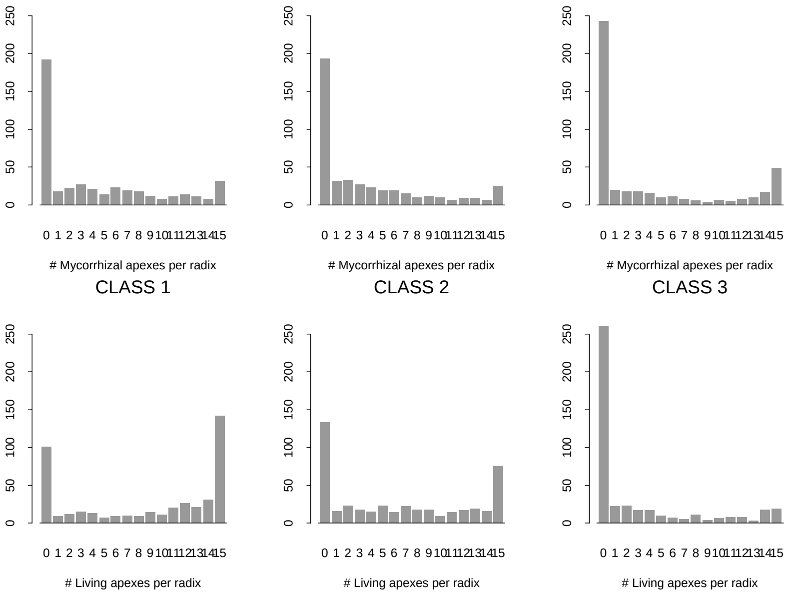

Figure 3: Frequency distributions of the number of ectomycorrhizal and living apexes per root

categories (figure 4) can help to put in focus the phenomenon.

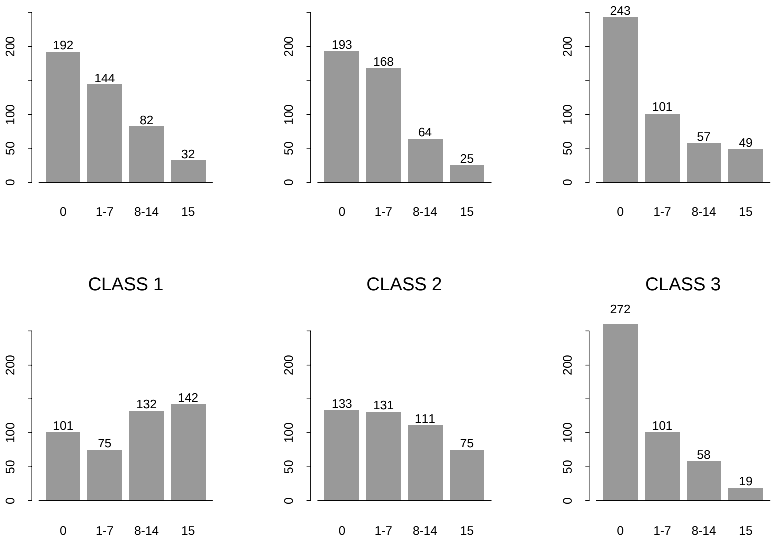

Figure 4: Four categories ectomycorrhizal and vitality distributions

The first two classes have a similar ectomycorrhizal distribution: infection is a rare event (pick of 0-roots) but when it succeds it spreads in the remaining three categories in a decreasing way rather uniformly—with a better ectomycorrhizal and vital state in the healthy class—. In the third class infection is still more rare but when it adheres it catches the whole root. Why these full infected roots die? A first interpretation of biological phenomenon can now be given. Left any attempt of explanation in terms of cause and effect, the kind of mutual association between the two variables has to be explored. It seems that in the damaged trees ectomychorrizal infection concentrates in few roots avoiding large portions of the root system: both a less infective ability of fungi involved and a loss of vitality of still surviving roots [1], in the unmycorrhized area, could be responsible. Moreover why in the third class some of the full ectomycorrhized roots die? An attempt of explanation is that in strongly declined plants the specific fungi in symbiosis with the dead observed roots were less capable to protect them from stresses in comparison with the fungi in symbiosis with living observed roots. But the difficulty to distinguish among morphotypes in dead ectomycorrhizae restrict the possibility to interpret the phenomenon more deeply.

The conditional relative frequencies of the four categories of root life given as many as categories of ectomycorrhizal state (fig. 5) can give a better insight of the sort of dependence between the two variables.

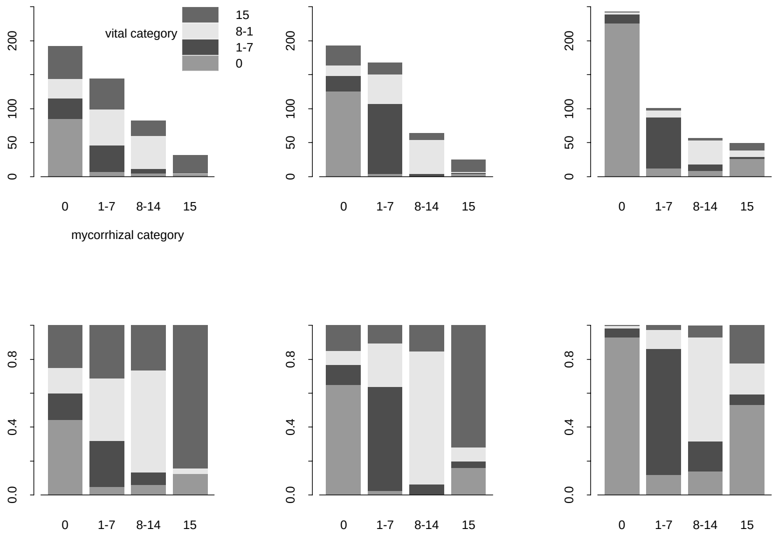

Figure 5: Absolute(above) and relative (beneath) frequencies of the four root life categories given the ectomycorrhizal ones with decline increase.

A positive association between pairs of the same category is evident in the second class and between pairs of the same and superior life category in the healthy class too. As foreseen, in the diseased class, the correlation is very strong for the first three categories while it is, in a certain sense, negative for the 15-mycorrhizae roots.

---

[1] Fungi produce ectomycorrhizae only with living apexes

Next figures describe distributions and possible influences of other variates in the phenomenon evolution. Figure 6 shows boxplots of root length given number of its ramifications, providing evidence that roots become more and more a tangle—shorter and with more ramifications—with decline increase.

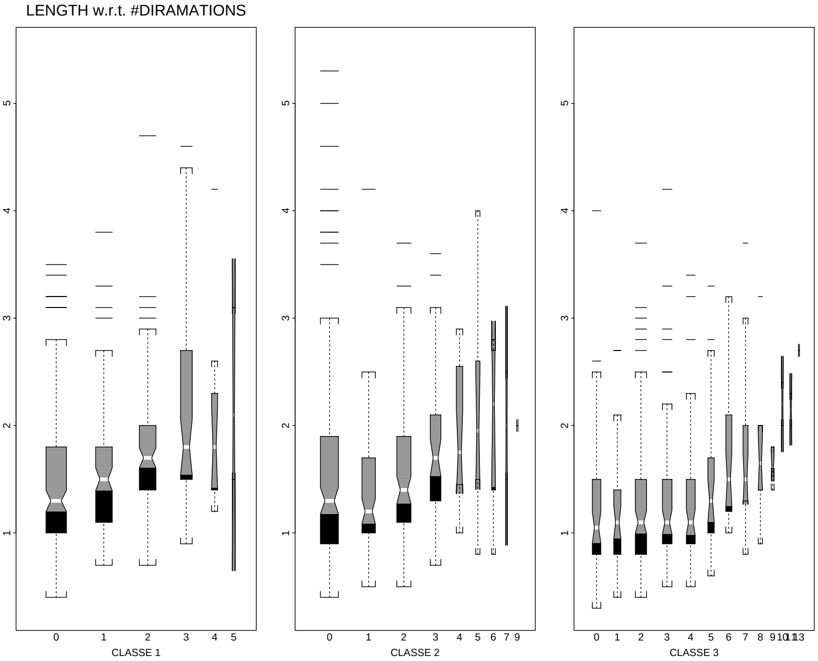

Figure 6: Length (in mm) w.r.t. number of ramifications of the root: boxplots with width proportional to the number of observations.

Figures 7 to 9 show conditional distributions of vitality and ectomycorrhizal four categories given three levels of length for each number of ramifications. The proportion of higher categories of vitality is usually larger with the increase of root length and the decreasing of ramification number. Mycorrhization generally shows analogous features. Figure 10 shows conditional distributions of vitality given ectomycorrhizal four categories for each class with respect to the level of a dummy factor variable, to be referred as sector dispersion factor. It is generated by classification of sectors in to three levels. In the first one ectomycorrhizal infection is distributed rather homogeneously in the four categories and the number of 0&0-roots (nor ectomycorrhizal neither alife) is low, as mean behaviour of class 1 is. In the last, on the contrary, ectomycorrhizal infection is sparse and concentrated (0-15 ectomycorrhizae) at the same time and there are many 0&0-roots, as in the third class. As expected in each class the typical sector category corresponds to the level of the class itself. It is rather evident that the typical bivariate distribution for each class is repeated in sectors classified into the three levels.

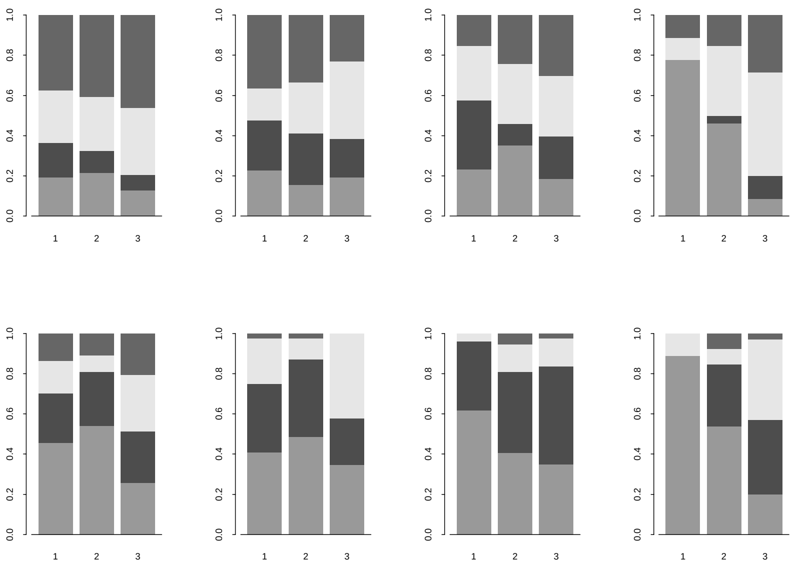

Figure 7: Vitality (above) and ectomycorrhizae (beneath), ramifications 0,1,2,3 (from left to right), length 0.4-1.3, 1.3 -1.8, 1.8 -4.7 (categories 1,2,3 in graphic)

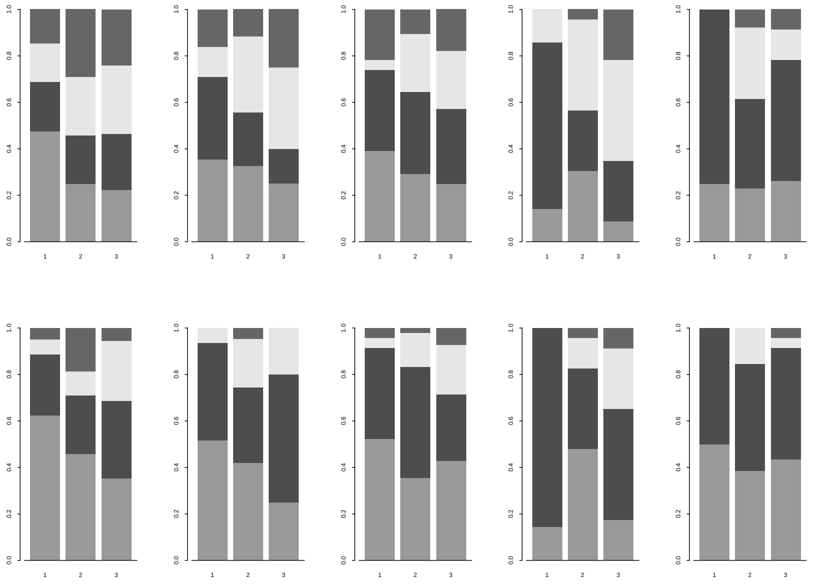

Figure 8: ramifications 0,1,2,3,4 (from left to right), length 0.4-1.1, 1.1-1.7, 1.7-5.3 (categories 1,2,3 in graphic)

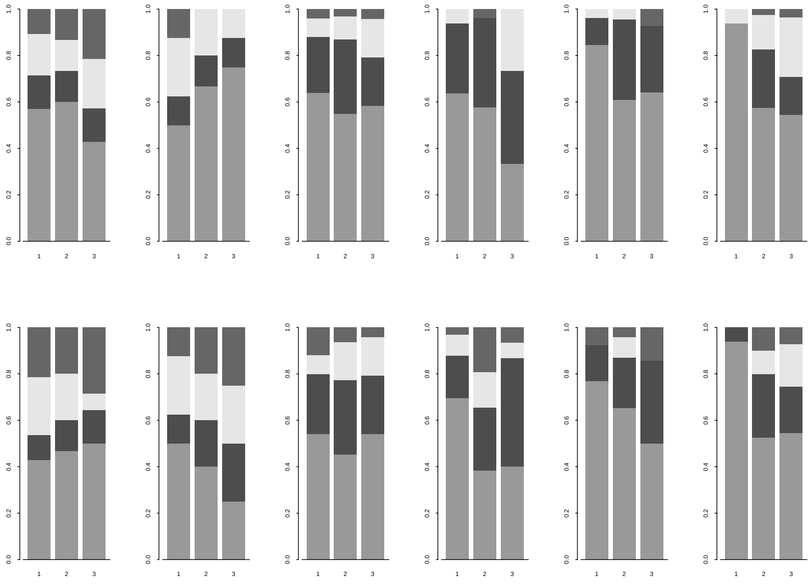

Figure 9: ramifications 0,1,2,3,4,5 (from left to right), length 0.3-1.0, 1.0-1.5, 1.5-7.5 (categories 1,2,3 in graphic)

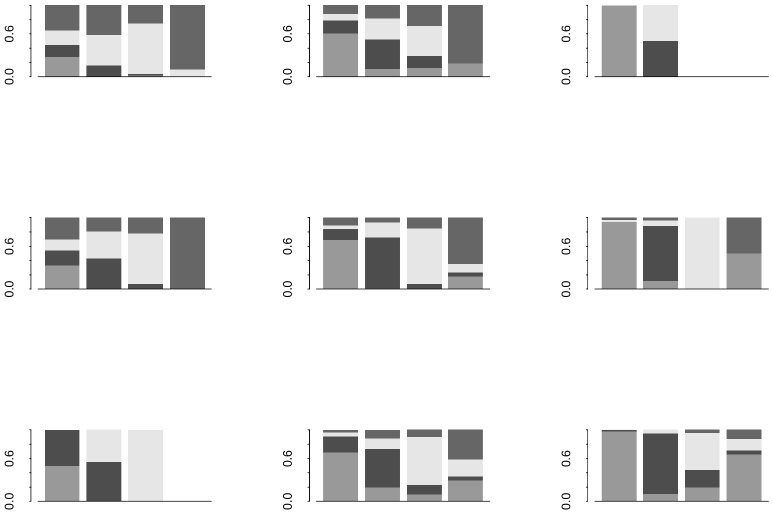

Figure 10: Class 1,2,3 (from top to bottom), sector disp. 1,2,3 (from left to right)

In the following we describe several marginal regression models for the binary response variable vitality. Clusters are defined by roots from which follows that the association between roots, between sectors and between trees is not taken into account.

3 Marginal regression models for correlated data

Let Y_{ij} be the vitality status of the jth apex within the ith root $j = 1, \ldots, T; i = 1, \ldots, n$. For each Y_{ij} several covariates are available, which can be summarized in the matrix $X_i = (x'_{i1}, \ldots, x'_{iT})'$. Marginal models specify the marginal expectation of the response given the covariates

$$\mathrm{E}(Y_{ij}|x_{ij}) = \mu_{ij} = g(x'_{ij}\beta)$$

where $g^{-1}(x'_{ij}\beta)$ is the so-called link function and β is the vector of parameters.

For binary data, the usual link function is the logit-link resulting in

$$\mathrm{E}(Y_{ij}|x_{ij}) = P(y_{ij} = 1|x_{ij}) = \frac{\exp(x'_{ij}\beta)}{1 + \exp(x'_{ij}\beta)}$$

3.1 Logistic marginal regression with markovian dependence

Within-root dependence among apexes is modeled like a Markov chain exploiting the one-way direction of the birth order and infection spread and of the nutrient flow from the older apexes to the younger ones[2]. Other possible sources of overdispersion have been neglected. The logistic marginal regression model proposed by Azzalini (1994) uses a markovian dependence structure. The association between two subsequent apexes is modeled in terms of an odds ratio of transition probabilities of a first order markovian chain[3]:

$$\psi = \frac{p_{11ij(j-1)}(1 - p_{00ij(j-1)})}{(1 - p_{11ij(j-1)})p_{10ij(j-1)}}$$

where $p_{klijj-1} = P(Y_{ij} = k|Y_{i(j-1)} = l)$. Solving

$$\begin{cases} \pi_{ij} = \pi_{i(j-1)}p_{11ij(j-1)} + (1 - \pi_{i(j-1)})p_{10ij(j-1)} \\ \psi = \frac{p_{11ij(j-1)}(1-p_{10ij(j-1)})}{(1-p_{11ij(j-1)})p_{10ij(j-1)}} \end{cases}$$

for $p_{11ij(j-1)}$ and $p_{10ij(j-1)}$ results in the transition probabilities

$$\begin{aligned} p_{10ij(j-1)} &= \frac{\delta_{ij(j-1)} - 1 + (\psi - 1)(\pi_{ij} - \pi_{ij-1})}{2(\psi - 1)(1 - \pi_{ij-1})} && (1) \\ p_{11ij(j-1)} &= p_{10ij(j-1)} \\ &\quad + \frac{1 - \delta_{ij(j-1)} + (\psi - 1)(\pi_{ij} + \pi_{i(j-1)} - 2\pi_{ij}\pi_{i(j-1)})}{2(\psi - 1)\pi_{i(j-1)}(1 - \pi_{i(j-1)})} && (2) \end{aligned}$$

with $\delta^2_{ij(j-1)} = 1 + (\psi - 1)\{(\pi_{ij} - \pi_{i(j-1)})^2\psi - (\pi_{ij} + \pi_{i(j-1)})^2 + 2(\pi_{ij} + \pi_{i(j-1)})\}$. Parameter estimates are obtained using standard maximum likelihood methods.

---

[2] The count order of the apexes has been inverted compared to the survey one. The first apex is the first to be born and to be liable to infection.

[3] The odds ratio is supposed constant across roots and apexes, but this condition can theoretically be relaxed.

Assuming independence among the roots besides a first order Markov chain structure within them, the log-likelihood is

$$l(\beta, \psi) = \sum_{i=1}^{n} l_i(\beta, \psi)$$

where

$$l_i(\beta, \psi) = \sum_{j=1}^{15} l_{ij}(\beta, \psi)$$

and

$$\begin{aligned} & l_i(\beta, \psi) && (3) \\ = \quad & y_{i1} \log \pi_{i1} + (1 - y_{i1}) \log(1 - \pi_{i1}) + \\ & + \sum_{j=2}^{15} y_{i(j-1)} y_{ij} \log p_{11ij(j-1)} + y_{i(j-1)} (1 - y_{ij}) \log(1 - p_{11ij(j-1)}) \times \\ & \times (1 - y_{i(j-1)}) y_{ij} \log p_{10ij(j-1)} (1 - y_{i(j-1)})(1 - y_{ij}) \log(1 - p_{10ij(j-1)}) \\ = \quad & y_{i1} \operatorname{logit} \pi_{i1} + \log(1 - \pi_{i1}) + \sum_{j=2}^{15} y_{ij} \operatorname{logit}(p_{1 y_{ij-1} jj-1}) + \\ & + \log(1 - p_{1 y_{ij-1} jj-1}) && (4) \end{aligned}$$

Inserting (1) and (2) in (3) the log-likelihood is parametrized in terms of the marginal means and the odds ratio; maximum likelihood estimates can then be obtained computing the first derivatives. In the context of repeated measures standard errors can be achieved by inversion of the sum of first derivatives squares[4], so avoiding the second derivatives computing. The hypothesis of orthogonality[5] of β and $\log \psi$ is an attractive feature of this model.

3.2 Mixed parameter model

The second approach is to use a 'mixed parameter model' for modeling marginal expectations and marginal pairwise associations. The underlying model for the joint distribution of the binary variates is a loglinear model and was e. g. introduced by Fitzmaurice and Laird (1993) for the context of marginal models. The joint density is given by

$$f(y; \Psi, \Omega) = P(Y_1 = y_1, \ldots, Y_T = y_T; \Psi, \Omega) = \exp\{y'\Psi + w'\Omega - A(\Psi, \Omega)\} \quad (5)$$

where $y = (y_1, \ldots, y_T)'$, $w = (y_1 y_2, y_1 y_3, \ldots, y_{T-1} y_T, \ldots, y_1 y_2 \cdots y_T)'$, $\Psi = (\Psi_1, \ldots, \Psi_T)'$ a parameter vector which can be interpreted as logits of conditional probabilities and $\Omega = (\omega_{12}, \omega_{13}, \ldots, \omega_{T-1T}, \ldots, \omega_{12\cdots T})'$ a vector representing conditional log odds ratios and contrasts of conditional log odds ratios.

---

[4] Quantity which approximates the Fisher information, at least for large n.

[5] It has been proved in some particular cases and it seems to hold in the model applications.

The expression

$$\exp\{A(\Psi,\Omega)\} = \sum_{y=(0,0,\dots,0)}^{y=(1,1,\dots,1)} \exp\{y'\Psi + w'\Omega\}$$

is the normalizing constant obtained by adding 2^T terms according to all the possible response profiles. The density (5) is a special case of the partial exponential families introduced by Zhao, Prentice and Self (1992). They have considered the following family

$$f(y;\Psi,\Omega) = \frac{1}{\Delta(\Psi,\Omega)} \exp\{y'\Psi + c(y,\Omega)\} \tag{6}$$

which includes the multinomial distribution and the multivariate normal distribution as particular cases. This presentation enables the modeling of mixed discrete and continuous response also. Choosing $\Delta(\Psi,\Omega) = \exp\{A(\Psi,\Omega)\}$ and $c(y,\Omega) = w'\Omega$ we get (5) as a special case of (6).

Zhao and Prentice (1990) have discussed the choice of a quadratic exponential family as the working likelihood for binary variables, i. e.,

$$f(y;\Psi,\Omega) = \frac{1}{\Delta(\Psi,\Omega)} \exp\{y'\Psi + w'\Omega\}\ ,$$

where w is found from the pairwise products $(y_1y_2),\dots,(y_{T-1}y_T)$ only, i. e., all interactions of order three and higher are neglected meaning thereby that they a priori are assumed to be zero. This parametrization was first proposed by Cox (1972). In fact we use such a quadratic model after a one-to-one transformation of the canonical parameters into marginal mean and pairwise marginal odds ratio parameters, see Heagerty and Zeger (1996), Glonek (1996) and Heumann (1998) for the details.

For the pairwise marginal log odds ratios Θ_i we used a model where the marginal log odds ratio α_1 between direct neighbor apexes (1 and 2, 2 and 3, and so on) is one parameter and the pairwise marginal log odds ratio α_2 between all other apexes is another:

$$\Theta_i(\text{apex } k,\ \text{apex } l) = \begin{cases} \alpha_1 & \text{if } |k-l| = 1, k \neq l \\ \alpha_2 & \text{if } |k-l| > 1, k \neq l \end{cases} \tag{7}$$

independent of the individual index i.

The maximum likelihood estimators $(\hat\beta,\hat\alpha)$ are then the solutions of the equations

$$s(\beta,\alpha) = \sum_{i=1}^{n} \begin{pmatrix} \frac{\partial\mu_i}{\partial\beta} & 0 \\ 0 & \frac{\partial\Theta_i}{\partial\alpha} \end{pmatrix}' \begin{pmatrix} I & 0 \\ \frac{\partial\nu_i}{\partial\mu_i} & \frac{\partial\nu_i}{\partial\Theta_i} \end{pmatrix}' \mathrm{Cov}(Y|W)_i^{-1} \begin{pmatrix} y_i - \mu_i \\ w_i - \nu_i \end{pmatrix} = 0\ ,$$

where $\frac{\partial\Theta_i}{\partial\alpha}$ is a design matrix independent of i resulting from the parametrisation in equation (7). ν_i is the vector of expectations of the pairwise products:

$\mathrm{E}(Y_1Y_2),\dots,\mathrm{E}(Y_{T-1}Y_T)$. $\frac{\partial\nu_i}{\partial\mu_i}$ and $\frac{\partial\nu_i}{\partial\Theta_i}$ result from the mentioned one-to-one transformation. $\mathrm{Cov}(Y|W)_i$ is the symmetric matrix

$$\begin{pmatrix} \mathrm{Var}(Y_i) & \mathrm{Cov}(Y_i, W_i) \\ \mathrm{Cov}(W_i, Y_i) & \mathrm{Var}(W_i) \end{pmatrix}$$

containing second moments in the upper left block, third moments in the upper right and lower left block and fourth moments on the lower right block.

An algorithm for estimating β and α using 'iterative proportional fitting within Fisher Scoring is described in Heagerty and Zeger (1996) and in more detail in Heumann (1998). Also a formula for estimating the variances of the estimates can be found there.

3.3 The generalised estimating equations

Contrary to the Azzalini approach and the mixed parameter model, the generalized estimating equations (GEE) are a semiparametric quasi-likelihood approach. It has first been proposed by Liang and Zeger (1986). Several extensions have been proposed, reviewed e.g. by Ziegler, Kastner and Blettner (1998).

The Generalized Linear Model—not taking into account that the observations within a cluster are correlated—will estimate the parameters using the Independence Estimating Equation (IEE)

$$\mathbf{s}(\beta) = \frac{1}{n}\sum_{i=1}^{n} D_i' V_i^{-1} \epsilon_i \tag{8}$$

where $D_i = \partial\mu_i/\partial\beta'$ is the diagonal matrix of first derivatives, V_i is the diagonal matrix of the variances $V_i = \mathrm{diag}(v_{it})$, where $v_{it} = \mathrm{Var}(Y_{it})$, and $\epsilon_i = y_i - \mu_i$. The solution of (8) is computed using the Fisher scoring algorithm. In (8), only the first two moments have to be specified. Therefore this is called a quasi-likelihood approach. Not taking into account the correlation between the observations, it is necessary to use the robust variance estimator proposed by White (1982):

$$\widehat{V(\hat\beta_{IEE})} = \left(\sum_{i=1}^{n} \hat D_i' \hat V_i^{-1} \hat D_i\right)^{-1} \left(\sum_{i=1}^{n} \hat D_i' \hat V_i^{-1} \hat\Omega_i \hat V_i^{-1} \hat D_i\right) \left(\sum_{i=1}^{n} \hat D_i' \hat V_i^{-1} \hat D_i\right)^{-1}. \tag{9}$$

$\sum_{i=1}^{n} \hat D_i' \hat V_i^{-1} \hat D_i$ is the estimated Fisher information matrix. The middle matrix on the right hand side of (9) consists of the estimated crossproducts of first order derivatives, also termed estimated outer product gradient (estimated OPG), with $\hat\Omega_i = (y_i - \hat\mu_i)(y_i - \hat\mu_i)'$.

The Independence Estimating Equations (8) can be inefficient, because V_i does not take the correlation between the observations into account. That is why Liang and Zeger (1986) modeled V_i as

$$V_i = A_i^{1/2} R_i A_i^{1/2}$$

where A_i is the matrix containing the variances on the diagonal, and R_i is a positive definite correlation matrix that should describe well the association structure. For estimating this matrix, Liang and Zeger (1986) used the method of moments. Prentice (1988) proposed a second estimating equation for the association which has teh same structure as (8). The correlation was taken as a measure for the association and the sample correlation is used as 'response' in the second estimation equation. This results in a second score equation

$$\mathbf{s}(\alpha) = \frac{1}{n}\sum_{i=1}^{n} E_i' U_i^{-1}(w_i - \rho_i) \tag{10}$$

where $E_i = \partial\rho_i/\partial\alpha'$ is the diagonal matrix of first derivatives, U_i is the matrix of the third and fourth moments, w_i is the vector of sample correlations and ρ is the correlation, which is modeled through the inverse of Fisher's z–transformation (Lipsitz, Laird and Harrington, 1991).

4 Results

We now apply the marginal regression models to the forest decline data. As there are 15 apexes per root, we have to assume independence between the roots within a sector for computational reasons. That is we have 1350 clusters of size 15. For the likelihood approach of the mixed parameter model we have to specify the joint distribution which consists of 2^{15} probabilities. This seems not to be computationally feasible. So for this approach we reduced the observations to the first ten apexes per cluster. The main goal of this analysis is to find out the relationship between mycorrhizal infection and vitality. Therefore modeling the mean structure is of primary interest.

The starting model only includes the class as covariate: it is dummy coded using class 1 as reference category. In model 2 ectomycorrhizal infection as binary covariate—using dummy coding—and the interaction between class and mycorrhizal infection are added. In model 3 the categorized ectomycorrhizal infection—which is cluster-constant—is included using dummy coding, too; also the interaction between class and the categorized ectomycorrhizal infection is modeled. Furthermore, factor—the sector dispersion factor, standing for the type of root distribution of ectomycorrhization and vitality within a sector—and the interaction factor×categorized ectomycorrhizal infection are added. Finally, in model 4 length and number of ramifications are considered. Number of ramifications is dummy coded, 0 and 1 are coded as 0, otherwise it is coded as 1.

The association is modeled using two parameters. The first one describes the association between subsequent apexes. The second one describes the pairwise association between each pair of apexes.

The results for models 1 to 4 are presented in table 2 to 5. For each estimation approach, parameters and standard deviations (in brackets) are given.

Significant results are highlighted. For the likelihood approaches the log likelihood is given, too. Note, that the results of the mixed parameter model are only based on 10 apexes per cluster (13 500 observations). Therefore, the likelihood values for the two different approaches within a model are not comparable.

The results of the generalized estimating equation approach of Prentice were only available for model 1. The algorithm did not converge for models 2 to 4. This could because of the restriction of the correlation which may be violated (see Prentice, 1988).

	IEE		GEE		Azzalini		Mixed param.	
Intercept	0.354	(0.080)***	0.355	(0.080)***	0.356	(0.055)***	0.328	(0.079)***
class2	-0.650	(0.109)***	-0.650	(0.109)***	-0.646	(0.078)***	-0.681	(0.112)***
class3	-1.828	(0.125)***	-1.830	(0.126)***	-1.835	(0.085)***	-1.817	(0.123)***
α_1			1.867	(0.045)***	3.932	(0.047)***	3.602	(0.075)***
α_2			1.404	(0.043)***			3.012	(0.081)***
Likelihood					-7221.402		-4639.210	

Table 2: parameter estimates, standard errors and significance (*: $p-$value $<$ 0.05, **: $p-$value < 0.01, ***: $p-$value < 0.001); α_1: association parameter for nearest neighbour, α_2: association parameter for all other associations; mixed parameter model is based on only 10 apexes per root

Model 1 estimates (table 2) point out a significant decrease in vitality from class1 to class 2 and much more strongly to class 3, a quite high correlation between either adiacent apexes or any two apexes in the same root, respectively around 0.7 and 0.6 in all estimating approaches.

	IEE		Azzalini		Mixed param.	
Intercept	-0.140	(0.091)	-0.072	(0.062)	-0.184	(0.088)*
class2	-0.868	(0.134)***	-0.922	(0.092)***	-0.862	(0.131)***
class3	-2.952	(0.212)***	-2.823	(0.138)***	-2.783	(0.193)***
mico	2.283	(0.248)***	2.384	(0.130)***	2.423	(0.180)***
C2mico	1.158	(0.382)***	1.432	(0.210)***	1.123	(0.281)***
C3mico	1.187	(0.345)***	1.390	(0.198)***	1.016	(0.270)***
α_1			5.216	(0.072)***	4.651	(0.101)***
α_2					3.892	(0.110)***
Likelihood			-4591.366		-2943.860	

Table 3: parameter estimates, standard errors and significance (*: $p-$value $<$ 0.05, **: $p-$value < 0.01, ***: $p-$value < 0.001); α_1: association parameter for nearest neighbour, α_2: association parameter for all other associations; mixed parameter model is based on only 10 apexes per root

In model 2 (table 3) both class factor and mycorrhiza indicator variable are significant. The class effect still acts but the presence of ectomycorrhiza involves very high vitality probability. With inclusion of ectomycorrhization model fit has been much improved and correlation has been estimated higher (producing an interesting articulate dependence between infection and vitality in near apexes).

	IEE		Azzalini		Mixed param.	
Intercept	0.149	(0.152)	0.180	(0.120)	0.072	(0.154)
class2	-0.323	(0.209)	-0.271	(0.162)*	-0.371	(0.214)
class3	-1.916	(0.413)***	-1.989	(0.266)***	-1.747	(0.358)***
myco	2.379	(0.226)***	2.523	(0.150)***	2.578	(0.212)***
cl2*myco	1.416	(0.381)***	1.545	(0.251)***	1.136	(0.333)***
cl3*myco	1.343	(0.373)***	1.435	(0.252)***	0.879	(0.340)**
myco(k1)	0.478	(0.218)*	0.365	(0.167)*	0.474	(0.235)*
myco(k2)	-0.093	(0.261)	-0.198	(0.197)	-0.240	(0.281)
myco(k3)	1.903	(0.865)*	1.055	(0.897)	1.045	(1.039)
cl2*myc1	-0.735	(0.286)**	-0.825	(0.211)***	-0.598	(0.298)*
cl3*myc1	-0.061	(0.510)	0.068	(0.332)	0.069	(0.465)
cl2*myc2	0.084	(0.396)	-0.220	(0.276)	0.013	(0.398)
cl3*myc2	-0.043	(0.607)	-0.240	(0.412)	-0.187	(0.564)
cl2*myc3	-1.506	(0.845)	-1.617	(0.561)**	-1.365	(0.753)
cl3*myc3	-1.353	(0.851)	-1.346	(0.562)**	-1.503	(0.763)*
factor2	-1.056	(0.205)***	-1.078	(0.161)***	-1.111	(0.210)***
factor3	-2.857	(0.484)***	-2.886	(0.273)***	-3.584	(0.406)***
fa2*myc1	0.327	(0.277)	0.473	(0.206)*	0.447	(0.288)
fa3*myc1	1.487	(0.573)**	1.741	(0.355)***	2.467	(0.523)***
fa2*myc2	0.252	(0.374)	0.568	(0.259)*	0.351	(0.363)
fa3*myc2	1.660	(0.667)*	2.109	(0.426)***	2.863	(0.613)***
fa2*myc3	-1.778	(0.830)*	-1.109	(0.901)	-0.631	(1.032)
fa3*myc3	-0.425	(1.065)	0.257	(0.993)	1.400	(1.182)
α_1			5.006	(0.072)***	4.431	(0.102)***
α_2					3.670	(0.110)***
Likelihood				-4416.508		-2840.33

Table 4: parameter estimates, standard errors and significance (*: $p-$value $<$ 0.05, **: $p-$value < 0.01, ***: $p-$value < 0.001); α_1: association parameter for nearest neighbour, α_2: association parameter for all other associations; mixed parameter model is based on only 10 apexes per root

Model 3 (table 4) adds factor and ,in minor way, the variable standing for the root infection size as sources of significant effect . Goodness of fit is a bit higher and correlation estimate a bit lower. Yet the known effects of class and myco apply, it is remarkable the decrease in vitality from factor1 to factor2 and factor3–expecially with myco(k0) and myco(k3)–, the stressed increase in myco(k0) and myco(k1) and the stressed decrease in myco(k3)–expecially in class2 and class3–. Model 4 (table 5) adds only the root length as source of significant effect. Analogous observations as those for model 3 apply with regard to model fit and correlation estimate.

	IEE		Azzalini		Mixed param.	
Intercept	0.038	(0.229)	0.116	(0.166)	-0.004	(0.224)
class2	-0.333	(0.211)	-0.286	(0.162)	-0.373	(0.214)
class3	-1.798	(0.422)***	-1.993	(0.273)***	-1.635	(0.363)***
myco	2.392	(0.225)***	2.549	(0.151)***	2.588	(0.213)***
cl2*myco	1.391	(0.379)***	1.575	(0.254)***	1.125	(0.333)***
cl3*myco	1.371	(0.375)***	1.369	(0.251)***	0.890	(0.341)**
myco(k1)	0.451	(0.221)*	0.333	(0.168)*	0.449	(0.236)
myco(k2)	-0.190	(0.267)	-0.310	(0.198)	-0.328	(0.283)
myco(k3)	1.805	(0.857)*	0.945	(0.895)	0.948	(1.041)
cl2*myc1	-0.683	(0.290)*	-0.775	(0.212)***	-0.558	(0.299)
cl3*myc1	-0.081	(0.509)	0.159	(0.336)	0.060	(0.466)
cl2*myc2	0.151	(0.400)	-0.160	(0.277)	0.078	(0.399)
cl3*myc2	-0.024	(0.607)	-0.076	(0.416)	-0.144	(0.566)
cl2*myc3	-1.441	(0.828)	-1.569	(0.569)**	-1.323	(0.754)
cl3*myc3	-1.355	(0.846)	-1.187	(0.567)*	-1.496	(0.766)
factor2	-1.065	(0.206)***	-1.148	(0.162)***	-1.129	(0.211)***
factor3	-2.783	(0.484)***	-2.908	(0.277)***	-3.528	(0.407)***
fa2*myc1	0.341	(0.278)	0.495	(0.207)**	0.476	(0.288)
fa3*myc1	1.461	(0.572)*	1.810	(0.357)***	2.466	(0.524)***
fa2*myc2	0.283	(0.374)	0.634	(0.260)**	0.382	(0.363)
fa3*myc2	1.630	(0.670)*	2.139	(0.428)***	2.821	(0.615)***
fa2*myc3	-1.740	(0.825)*	-1.002	(0.900)	-0.575	(1.035)
fa3*myc3	-0.445	(1.058)	0.349	(0.992)	1.389	(1.184)
length	0.141	(0.120)	0.268	(0.079)***	0.123	(0.112)
ram	-0.475	(0.276)	-0.134	(0.192)	-0.385	(0.278)
length*ram	0.125	(0.156)	-0.052	(0.108)	0.066	(0.156)
α_1			4.980	(0.072)***	4.423	(0.102)***
α_2					3.661	(0.110)***
Likelihood				-4405.720		-2836.420

Table 5: parameter estimates, standard errors and significance (*: $p-$value $<$ 0.05, **: $p-$value < 0.01, ***: $p-$value < 0.001); α_1: association parameter for nearest neighbour, α_2: association parameter for all other associations; mixed parameter model is based on only 10 apexes per root

5 Discussion

In this paper we used several estimating approaches for marginal models to analyze the relationship between ectomycorrhizal infection and vitality.

The first important goal was to demonstrate that an increase of decline of the epigean apparatus (classes from 1 to 3) corresponds to a decrease of root vitality. This was partially expected as it is well known that a deep and complex interrelation between epigean and hypogean apparatus exists. Hence it is realistic to presume that when damage symptoms are observed for a long time in one apparatus, sufference symptoms are detectable also in the other one.

The second goal was to confirm that the apex vitality is positively associated to the ectomycorrhizal degree. This also was expected and is probably due to the uptake (of nutrients and water) and to the (both biotic and abiotic) stress resistance induced by the ectomycorrhizal symbiosis.

Nevertheless the mutual influence between vitality and ectomycorrhization stands out at each level of data clustering, whether apex or root or sector. We attempt some possible interpretations. The presence of living ectomycorrhized apexes can protect the unmycorrhized neighbours from biological stresses. This fact derives from the capacity of ectomycorrhizal fungi to oppose the colonization of the closest rhizophere by some harmful soil microrganisms, both producing toxic compounds or antibiotics and modifying the root exudates. On the other hand this study demonstrates that vitality is a variable with a high correlation. In fact the death of proximal[6] apexes can cause nutrient flow interruptions, and consequently the death of the distal apexes. Furthermore the vitality degree of plants can influence the capacity of attracting symbionts or selecting the ectomycorrhizal fungi. In fact we can presume the existence of ectomycorrhizal fungi with more or less ability of infecting through the soil (e. g. by emanating hyphae or rhizomorphs), or more or less protective for plants.

At last the root length–generally higher in class1– turns out to be positively correlated with vitality. It can be interpretated as a root vegetative vigour index, higher in healthy plants and lower in the declining ones.

In our marginal models we neglected the association between roots within a sector, between sectors within a tree and between trees within a class. This might result in incorrect conclusions, if clusters are not independent. Possible solutions are to extend the likelihood approaches to the case of nested structures or to use different approaches like bayesian hierarchical models (Knorr-Held and Rasser, 1998).

For the mixed parameter model it was necessary—because of computational feasibility—to reduce the data to the first ten apexes per root. This subset is taken systematically (the first ten), so a comparison of the parameter estimates between estimation approaches should be done with care. An additional effect is that the parameter estimates are not as efficient as in the Azzalini approach.

The generalized estimating equation approach by Prentice did not converge when cluster-specific covariates are used in the mean model. This approach uses

---

[6] Proximal apexes consist of the older apexes along the root growth direction.

the correlation as measure of association, which is bounded. Another possibility is to use the odds ratio as measure of association (Fahrmeir and Pritscher, 1996).

Acknowledgements

This work was supported by the Deutsche Forschungsgemeinschaft.

References

Azzalini, A. (1994). Logistic regression for autocorrelated data with application to repeated measures, *Biometrika* **81**: 767–775.

Causin, R., Montecchio, L. and Mutto Accordi, S. (1996). Probability of ectomycorrhizal infection in a declining stand of common oak, *Ann. Sci. For.* **53**: 743–752.

Cox, D. R. (1972). The analysis of multivariate binary data, *Applied Statistics* **21**: 113–120.

Fahrmeir, L. and Pritscher, L. (1996). Regression analysis of forest damage by marginal models for correlated ordinal responses, *Journal of Enviromental and Ecological Statistics* **3**: 257–268.

Fahrmeir, L. and Tutz, G. (1994). *Multivariate Statistical Modelling Based on Generalized Linear Models*, Springer, New York.

Fitzmaurice, G. M. and Laird, N. M. (1993). A likelihood-based method for analysing longitudinal binary responses, *Biometrika* **80**: 141–151.

Glonek, G. V. F. (1996). A class of regression models for multivariate categorical responses, *Biometrika* **83**: 15–28.

Heagerty, P. J. and Zeger, S. L. (1996). Marginal regression models for clustered ordinal measurements, *Journal of the American Statistical Association* **91**: 1024–1036.

Heumann, C. (1998). *Likelihoodbasierte marginale Regressionsmodelle für korrelierte kategoriale Daten*, Peter Lang Europäischer Verlag der Wissenschaften, Frankfurt am Main.

Knorr-Held, L. and Rasser, G. (1998). Bayesian detection of clusters and discontinuities in disease maps, *SFB386 – Discussion paper*, Ludwig-Maximilians-Universität München.

Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models, *Biometrika* **73**: 13–22.

Lipsitz, S. R., Laird, N. M. and Harrington, D. P. (1991). Generalized estimating equations for correlated binary data: Using the odds ratio as a measure of association, *Biometrika* **78**: 153–160.

Prentice, R. L. (1988). Correlated binary regression with covariates specific to each binary observation, *Biometrics* **44**: 1033–1048.

White, H. (1982). Maximum likelihood estimation of misspecified models, *Econometrica* **50**: 1–25.

Zhao, L. P. and Prentice, R. L. (1990). Correlated binary regression using a generalized quadratic model, *Biometrika* **77**: 642–648.

Zhao, L. P., Prentice, R. L. and Self, S. G. (1992). Multivariate mean parameter estimation by using a partly exponential model, *Journal of the Royal Statistical Society, Series B* **54**: 805–811.

Ziegler, A., Kastner, C. and Blettner, M. (1998). The generalised estimating equations: An annotated bibliography, *Biometrical Journal* **40**: 115–139.