



# **A statistical model to detect asymptomatic infectious individuals with an application in the *Phytophthora alni*-induced alder decline**

Chabi Fabrice Elegbede, J-Claude Pierrat, Jaime Aguayo-Silva, Claude Husson, Fabien Halkett, Benoit Marçais

## **► To cite this version:**

Chabi Fabrice Elegbede, J-Claude Pierrat, Jaime Aguayo-Silva, Claude Husson, Fabien Halkett, et al.. A statistical model to detect asymptomatic infectious individuals with an application in the *Phytophthora alni*-induced alder decline. *Phytopathology*, 2010, 100 (11), pp.1262-1269. 10.1094/PHYTO-05-10-0140 . hal-01195036

**HAL Id: hal-01195036**

**<https://hal.science/hal-01195036>**

Submitted on 18 Apr 2019

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

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

# A Statistical Model to Detect Asymptomatic Infectious Individuals with an Application in the *Phytophthora alni*-Induced Alder Decline

Chabi Fabrice Elegbede, Jean-Claude Pierrat, Jaime Aguayo, Claude Husson, Fabien Halkett, and Benoît Marçais

First, third, fourth, fifth, and sixth authors: INRA, UMR1136 Interaction Arbres/Microorganismes, F-54280 Champenoux, France and Nancy-Université, UMR1136 Interaction Arbres/Microorganismes, F-54000 Nancy, France; and first and second authors: INRA, Agro-ParisTech, UMR1092 LERFoB, F-54042 Nancy, France.

Accepted for publication 4 June 2010.

## ABSTRACT

Elegbede, C. F., Pierrat, J. C., Aguayo, J., Husson, C., Halkett, F., and Marçais, B. 2010. A statistical model to detect asymptomatic infectious individuals with an application in the *Phytophthora alni*-induced alder decline. *Phytopathology* 100:1262-1269.

In some diseases—in particular, tree root infection—stages of infection and inoculum production level and timing are not readily observable because of uncertainty or time lags in symptom appearance. Here, we pose a criterion, based on relative hazard of disease symptoms, to discriminate between healthy and asymptomatic infected individuals. We design a statistical procedure to estimate the criterion for a 6-year survey of alder decline along a northeastern French river. Individual tree symptom hazard was modeled with Cox's regression model, taking estimation

of local infection pressure as a risk factor. From an inoculum production experiment, we thereafter assessed the inoculum production level of target trees, including symptomatic and asymptomatic trees ranked according to their symptoms hazard. Using receiver operating characteristic methods, we first evaluated the criterion performance and determined the discrimination threshold to sort out asymptomatic individuals into healthy and infected. Then, we highlighted the fact that the infected asymptomatic trees were among the major inoculum producers whereas severely declining and dead trees were found to be poor inoculum sources.

*Additional keywords:* spatial point pattern analysis, survival analysis.

In epidemiology, disease dynamics is often modeled using compartmental models (14,22) which divide the hosts into different categories, taking the stages of infection into account. Defining the categories is not always straightforward because some stages of disease are not easily observed for some host–pathogen interactions. Difficulties arise particularly when infected hosts are in an incubation period or when they do not show obvious symptoms of the disease even after infection. Determining the importance of the host as an inoculum source (hereafter referred to as “host infectious level”), detecting hosts in the incubation period, and describing host behavior during disease development are some of the main challenges in plant epidemiology (11). The problem is especially acute in the case of root diseases of trees such as root rot or *Phytophthora* disease for which symptoms are not readily observed and often take time to impact tree health in visible ways (1,2,10,27).

In recent years, a new lethal disease has caused a considerable decline of alder along European rivers. The pathogenic agent, *Phytophthora alni*, was detected for the first time in the United Kingdom in 1993 (3,4,13). In ensuing years, the pathogen has been detected throughout Europe and, in particular, in France (38). Surveys conducted in France revealed that 16% of alder trees present along the watercourse in the Rhin-Meuse Basin were diseased (39). The disease is worrying for river managers because alder plays an important role in riparian ecology. As a deep-rooted species, it contributes to maintenance of the soil structure of river banks and reduces the effects of erosion. Alder root systems under

water shelter fish from predation and strong river currents (5). Given the importance of ecosystem services provided by alder trees, epidemiological study appeared essential to monitor and save this component of the riparian system but actual knowledge of alder decline is impeded by the poor identification of infectious individuals.

Whereas disease symptoms (37), the characteristics of the responsible pathogen and its detection (4,20), and risk factors of epidemics (23,39) have been documented, the evolution of a tree during the course of the disease and, particularly, when trees start to produce inoculum (i.e. become infectious) is still poorly known. Two external signs of the disease exist: (i) crown decline (i.e., sparse foliage due to abnormally small, yellow leaves and defoliation) and (ii) cankers on the trunk (i.e., dark necrosis, with sometimes tarry exudates, due to bark-killing at the base of the stem). Neither symptom is strictly linked with the presence of the disease; infected trees may show both, either, or neither of the symptoms. In surveys, approximately half of the declining trees were found to have no canker while many cankered trees did not decline (39). Cankers appear quickly, usually a few months after infection (31), and often result from direct infection of the trunk base during flooding through infection of lenticels by flagellated zoospores. In contrast, crown decline is believed to appear more slowly, expressing only past infection, probably because it results from progressive root system destruction. Although the lesions observed on the collar and base of the stem may girdle and kill the tree, in particular seedlings, crown decline was considered to be induced mainly by root destruction. Root colonization occurs through infection of fine roots by the zoospores. Infected fine roots are a good inoculum source under conditions such as flooding or wet soil, whereas the capacity of cankers to become inoculum sources appears limited (24,37). Thus, only trees with root infection should be considered to be infectious. However, many of these infectious trees may express no visible symptoms

Corresponding author: C. F. Elegbede; E-mail address: chabi.elegbede@gmail.com

\*The e-Xtra logo stands for “electronic extra” and indicates that Figure 1 appears in color online.

doi:10.1094/PHYTO-05-10-0140

© 2010 The American Phytopathological Society

(neither crown decline nor canker). Identifying these asymptomatic infectious trees is critical to predict disease evolution. Assuming that asymptomatic trees with a high risk of future crown decline are likely candidates because they should present root infection, we analyzed the risk of crown decline.

Survival analysis methods (18) are often used in epidemiological studies to analyze the length of time until the occurrence of an event; in general, infection or mortality (29). Studies have focused on predicting the hazard (instantaneous rate) of the event, comparing the survival distributions of monitored individuals in order to determine the most likely to be affected by the pathogens, and determining the impact of potential risk factors. Cox's regression model (6) is the most popular survival regression model used (18). To develop that model, variables which act as risk factors for individual under scrutiny are needed. Unfortunately, no tree characteristics that might act as a risk factor could be identified for infection by *P. alni* (39). Therefore, we decided to take into account the tree local environment impact. Such an approach in epidemiology, plant ecology, and forest sciences is often done using point pattern analysis within spatial statistics tools (8,12,36). Here, we used the number of infected neighbors and distance from these to assess local probability of infection with a non-parametric kernel estimation method (12,41).

Determining a tree infectious status from a decline likelihood requires the definition of a decision threshold that enables discrimination between asymptomatic healthy and infected trees. Receiver operating characteristic (ROC) analysis, a statistical means for evaluating the performance of a diagnostic test (17), is well adapted to this type of problems. This approach has been used in plant disease management to develop decision-making or diagnosis tools when the infection process is difficult to measure directly (19,26,43). The use of an ROC analysis implied a previous description of infection status of trees using a reference diagnostic test, namely a gold standard test (17).

The objective of the present study is to elaborate and validate a statistical procedure to discriminate between infectious and healthy (i.e., noninoculum producer) alder trees. This work is divided into three parts. First, we conducted a statistical analysis of the pluri-annual survey of a large alder population to assess the individual tree hazard of crown decline. Second, from inoculum production experiments we determined the infectious status of symptomatic and asymptomatic trees sorted out according to the risk of crown decline. Third, we tested the accuracy of our statistical procedure to discriminate asymptomatic infectious trees.

## MATERIALS AND METHODS

**Observation site.** In 2002, a survey was carried out on alder in northeastern France along the Sarre River, a tributary of the Moselle. The surveyed area encompassed 4.2 km along the river and extended 20 m inland from the river bank. Different sampling densities were used. In areas with low tree density, all trees were sampled whereas, in areas with high tree density, sampling intensity of 15 to 57% was used. Annually, in September and October, health of the sampled trees was assessed, all alder trees reaching a height of 1.30 m were recruited, and dead trees were removed from the survey. In total, 2,850 living alder trees were sampled over a period of 6 years, 2002 to 2007.

For each sampled alder, information was collected on its year of recruitment to the sample, its coordinates (obtained by GPS or computed from positions of previously mapped trees), its distance from the river, its number of stems (for coppiced trees), and its health status. Tree health was defined by two scales: absence or presence of a canker on the trunk, noted by 0 or 1, respectively; and status of the crown, rated on a 1-to-4 scale, where 1 = healthy foliage (leaves with normal size and color and dense crown); 2 = defoliation of 10 to 50%, with small and yellowish leaves; 3 = defoliation of >50%, with small yellowish leaves and dead

branches; and 4 = dead tree (no leaves, bark dead at the trunk level). The diameter of trees at a height of 1.30 m from ground level was measured at 2-year intervals in 2002 to 2006. The basal area of the trees (i.e., cross-sectional area) was computed from diameter measurements.

**Model to assess plant risk to crown decline.** *Structure of the model.* Individual tree risk of crown decline was assessed through survival analysis (i.e., analysis of the time delay until appearance of crown decline symptoms = first rating of 2 or 3 for crown status independently of canker occurrence). Thus, if  $T$  is the random variable describing time before appearance of crown decline, the survival function  $S(t)$  to crown decline is defined at  $t$  by:

$$S(t) = \Pr[T > t] = 1 - F(t) \quad (1)$$

where  $F$  is a cumulative distribution function of  $T$ . The function  $S(t)$  captures the probability that a tree will not have experienced decline yet at a specified time  $t$  (18). In order to assess  $S(t)$ , the hazard function  $h(t)$  (18), which is the instantaneous risk of crown decline at time  $t$  provided that the trees remained healthy until that time, is defined and linked to  $S(t)$  by:

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{\Pr[t \leq T < t + \Delta t | T \geq t]}{\Delta t}$$

and (2)

$$h(t) = -\frac{\dot{S}(t)}{S(t)} = -\frac{d}{dt} \ln[S(t)]$$

where  $\dot{S}(t)$  is the derivative of  $S(t)$  with respect to  $t$ . Therefore,  $S(t) = \exp(-\int_0^t h(u) du)$ . In this study, the hazard function was estimated using the Cox's regression model (6,18), which assumes that the hazard function is the product of two functions, one depending only on time,  $h_0(t)$ , and one depending only on the risk factors,  $\exp(\beta'Z)$ . This model is defined by:

$$h(t) = h(t, Z) = h_0(t) \exp(\beta'Z) \quad (3)$$

with  $Z' = [Z_1, \dots, Z_p]$ , a  $p$  vector of covariates (risk factors);  $\beta' = [\beta_1, \dots, \beta_p]$ , a vector of  $p$  parameters ( $\exp(\beta'Z)$  characterizing how the hazard function changes according to subject covariates); and  $h_0(t)$ , the baseline hazard function (the hazard function when  $Z' = [0, \dots, 0]$ ), which characterizes how the hazard function changes according to the survival time.

The coefficients of the model were estimated by the methods of partial likelihood (6,18), and only right censoring were taken into account in this study. Censoring occurred essentially when crown decline could not be observed; for instance, when trees disappear from the study prior to crown decline (e.g., because they are felled by fishermen). The entry date in the study was taken as the year of recruitment and date of decline as the first year of crown symptom observation (i.e., first rating of 2 or 3 for crown status independently of canker occurrence). The use of partial likelihood is based on the assumption that there were no tied times (i.e., declined trees were supposed not to have the same crown decline time). Unfortunately, because trees were surveyed once a year, tied survival times were observed and taken into account in the partial likelihood by the Efron approximation (18,30).

The Cox's model is semiparametric because the baseline hazard  $h_0(t)$  is estimated by a nonparametric method. Also, the model is said to be a proportional-hazard model because, for two observations  $i$  and  $j$  that differ in their covariates  $Z_i$  and  $Z_j$ , the relative hazard:

$$\frac{h(t, Z_i)}{h(t, Z_j)} = \frac{h_0(t) \exp(\beta'Z_i)}{h_0(t) \exp(\beta'Z_j)} = \frac{\exp(\beta'Z_i)}{\exp(\beta'Z_j)}$$

is independent of time  $t$ . This proportional hazards assumption was checked using tests based on the scaled Schoenfeld (16,33) statistic. Precisely, a  $\chi^2$  test is computed for each covariate, along with a global test for the model in whole. Subsequently, relative hazard of crown decline (RHCD) was used as a measure of risk of crown decline and was estimated for trees with healthy crowns relative to the average subject tree stem from the data set.

The statistical analysis and parameter estimation was realized using the package Survival of the R Software (R Foundation for Statistical Computing, Vienna). Several models were tested to identify pertinent covariates to assess crown decline hazard (using equation 3). The Bayesian information criterion (BIC) (34) was used for model selection.

*Tree characteristics as input explicative variables.* Except for tree basal area  $g$  (used as a surrogate for the size of tree root system available for infection), several tree characteristics were tested. Distance  $d$  to river bank was considered because water is the main vector of the infectious propagules and trees along the riversides are probably more subjected to infection. Presence or absence of canker on the trunk, indicating probable infection, was considered through a dummy variable  $p$ . Finally,  $cop$ , a dummy variable, was considered for type of alder tree (either coppiced or not).

*Local infection pressure as input explicative variable.* Availability of a local source of inoculum impacts root destruction and probability of crown decline and, thus, needs to be taken into account when estimating RHCD. Unfortunately, no tree characteristics that might act as risk factors for infection by *P. alni* could be identified. To overcome this problem, we used a surrogate variable, based on local dynamics of new canker occurrence. Indeed, because canker is a quick response to infection, it was assumed to be a better measure of current infection dynamics than crown decline status. Thus, local annual probability of canker, denoted  $c_t(x)$  (i.e., the probability that a tree, located at  $x$  and without cankers in year  $t - 1$ , got cankered in year  $t$ ) was estimated for each tree.

To estimate the local risk of canker, we used methods based on spatial point pattern analysis (12). In such methods, the spatial point distribution of a set of location describing an event (for instance, cankered tree) may be characterized by its intensity, which is the mean number of events per unit of area. In a two-dimensional context, this intensity is the density of sampled trees per unit of surface, and its estimation is similar to a bivariate probability density estimation (12). Thus, using the point pattern layer, generated by the bidimensional georeferencing of trees over the survey area, the spatial point distribution of trees was statistically evaluated to obtain a smooth estimation of the local risk of canker applying the bidimensional kernel method (12,41). The kernel intensity estimation is an interpolating and smoothing technique for generalizing event location to an entire area. It consists of placing a moving tree-dimensional function (the kernel) of a given radius (also called bandwidth or window) that visits each point in turn on a smooth grid superimposed over the study area. The kernel function weights events within its sphere of influence according to their distance from the point at which the intensity is being estimated. For each point of the smooth grid, the sum of event individual kernels is calculated. Then, a smoothed surface of the target event spatial distribution is generated. In our study, this technique was applied to compute the ratio of the intensity estimation of new cankered trees in year  $t$  to the intensity estimation of noncanker trees in year  $t - 1$  (population at risk the previous year). This ratio provides spatialized estimates of the probability of new canker over the entire study area. We used the quartic kernel and a fixed bandwidth of 20 m. This estimation was done using the package Splan of the R software (R Foundation for Statistical Computing). More details using mathematical formula are given in the Appendix.

Local probabilities of canker were handled as time-independent covariates in the model and each year period was used as a separate observation. The predictive effect of  $c_t(x)$ ,  $c_{t-1}(x)$ , and  $c_{t-2}(x)$  (local probabilities of cankers in year  $t$ ,  $t - 1$ , and  $t - 2$ , respectively, for a located tree in  $x$ ) until 3 years before the crown decline was checked.

In addition, this estimation was used to check whether the river geometry affects tree infection. Indeed, estimated local probability of cankers for straight parts and curved parts of the river separating outside and inside of curves were compared.

**Inoculum quantification experiment for model validation.** *P. alni* soil inoculum was quantified at the base of trees, using a baiting method (15,42), to assess the relationship between the hazard of crown decline and tree infection status. Also, this enabled us to assess inoculum production at base of trees and to identify which type of tree contributes the most to inoculum production.

*Sampling of trees to collect soil for P. alni detection.* Trees without neighbors in a 1-m radius were selected, taking into account their size, crown status, and, for asymptomatic trees, RHCD in the following fall (relative to the average subject within the data set, computed from the previously developed model). Four groups of trees were investigated: T1, trees with healthy crowns, no canker, and values of predicted RHCD of 0.1–2; T2, trees with crown rated as 2 (i.e., moderate decline, with or without canker); T3, trees with crown rated as 3 (i.e., severe decline, with or without canker); and T4, dead trees. In all, 10 to 25 trees were selected for each group, with an effort made to include both seedlings (considered as trees with their basal area <20 cm<sup>2</sup>) and large trees in each group. This was repeated in two consecutive years on different trees. Soil at the base of selected trees was sampled at 3 and 7 May 2007 (92 trees) and at 23 and 25 June and 7 July 2008 (89 trees). Between two and four soil samples were taken per tree, depending on the trunk diameter, at 10 to 50 cm away from stem base. Litter was brushed aside and  $\approx$ 500 ml of soil was taken with a soil borer at 5 to 20 cm of depth. All the samples from the base of a tree were pooled in a plastic bag, brought back to the laboratory, and kept at 4°C until processed.

*Baiting method for soil inoculum quantification.* To quantify inoculum in a sample, 200 ml of soil was placed in a plastic container and 500 ml of distilled water was added to it. Three well-developed, similar-sized leaves of rhododendron cv. Cunningham's White were placed in each container floating on the surface of the water as baits to quantify the amount of zoospores produced. The containers were incubated at a controlled temperature of 19 to 20°C in the dark for 3 to 4 days, until necrosis appeared on the leaves. The leaves were then removed from the containers, dried, and kept at 4°C for 24 h in order to facilitate counts of necrotic spotting. Pathogen-induced necrosis appears as  $\approx$ 0.5-mm-diameter dark spots on the abaxial leaf surface. The number of necrotic spots induced by *Phytophthora* spp. per leaf was recorded, after which  $\leq 10$  of the necrotic spots per soil sample were cut with a sterilized scalpel and placed in 96-wells plates for DNA extraction to determine the proportion of necrosis that were induced by *P. alni*. Negative and positive controls (i.e., piece of rhododendron leaf incubated with either sterilized soil or soil artificially contaminated with *P. alni*) were included in each plate. DNA of leaf tissues was extracted using a plant DNA extraction kit (DNeasy96 plant kit; Qiagen, Courtaboeuf, France) following the manufacturer's instructions. Polymerase chain reaction (PCR) tests with the species-specific primer pairs PA-F/R were performed (21). PCR products were separated by gel electrophoresis and visualized under UV light after ethidium bromide staining in order to determine the proportion of *Phytophthora* lesions that were induced by *P. alni* for each soil sample. The total number of necrotic spots induced by *P. alni* was then estimated to define the infectious level of trees at the base of which each soil was sampled (i.e., amount of inoculum produced by those trees).

**Assessment of the RHCD accuracy and a decision threshold for tree discrimination.** We assumed that experiments for *P. alni* detection at the tree base were accurate enough to be used as a gold standard test for an ROC analysis. Trees were classified in two subgroups according to *P. alni* presence (group of cases  $I+$ ) or absence (group of controls  $I-$ ) in the soil baiting test, and the accuracy of a diagnostic test based on RHCD ( $T+$  when  $\text{RHCD} > \text{RHCD}_0$  and  $T-$  otherwise,  $\text{RHCD}_0$  being an RHCD threshold to determine) was assessed. The accuracy of such diagnostic tests with binary outcomes is assessed in terms of the probability  $\text{Pr}(T+|I+)$  (namely, sensitivity, which also referred to the true positive proportion) that the test correctly classifies an infected tree as positive ( $T+$ ), and the probability  $\text{Pr}(T-|I-)$  (namely, the specificity, which also referred to the true negative proportion) that the test correctly classifies an uninfected tree as negative ( $T-$ ). Thus, to assess classification accuracy of the diagnostic test based on RHCD and choose an optimal decision threshold, the relationship between sensitivity and specificity was analyzed graphically. Precisely, we first plotted the ROC curve (sensitivity as function of  $[1 - \text{specificity}]$  at all possible decision thresholds), which enabled the assessment of the test without specifying a particular threshold value. Area under this curve (AUC) provided an overall measure of accuracy for the test. Briefly, AUC can range from 0.5 (noninformative or random discrimination strategy;

corresponds to diagonal  $y = x$ ) to 1 (perfect discrimination or accuracy). The larger the area, the better the test is (see 17 for more details about the meaning of the AUC). In a second step, because both sensitivity and specificity are of importance in our study, we chose a decision threshold which corresponds to a compromise between them. This was done using curve of sensitivity and specificity as function of the decision thresholds.

Finally, the probability that *P. alni* was detected in the soil at the base of sampled trees was analyzed using a logistic regression model to describe how tree characteristics (e.g., crown status, as defined above, and tree size) might impact *P. alni* detection in soil. A logistic regression model generally is used to calculate the probability of a given binary outcome (here, detection or non-detection of *P. alni*) as a function of a set of explanatory variables. For further details regarding logistic models, it is suggested to consult a reference work on generalized linear models (28).

## RESULTS

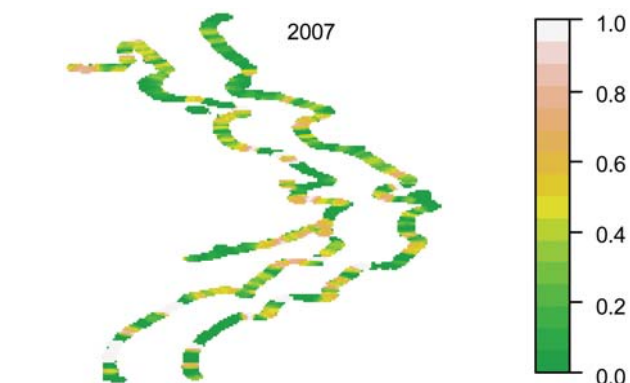
**Local risk of canker.** Local annual probability of cankers (i.e., the probability that the trees without cankers in the previous year got cankered in the current year) was estimated for 2003 onward. Estimations for 2007 are shown in Figure 1. Comparison for straight parts and curved parts of the river separating outside and inside of curves showed that the local risk of cankers did not depend on the river geometry in 2007 ( $F = 0.07$ ,  $df = 2$ ,  $P = 0.8$ ). The same results were obtained the other years (not shown). Pearson's correlation between probabilities of canker for two different years (Fig. 2) was computed to check the degree to which these values were related. In general, there were small positive correlations (Pearson  $r$  of 0.1 to 0.4). Thus, although some locations tended to show either low or high canker probability over years, there was considerable year-to-year variation.

**RHCD.** The linear predictor  $\beta'Z$  finally considered in the Cox's regression model to estimate the hazard function and survival time to crown decline for nondeclining trees located in  $x$  was:

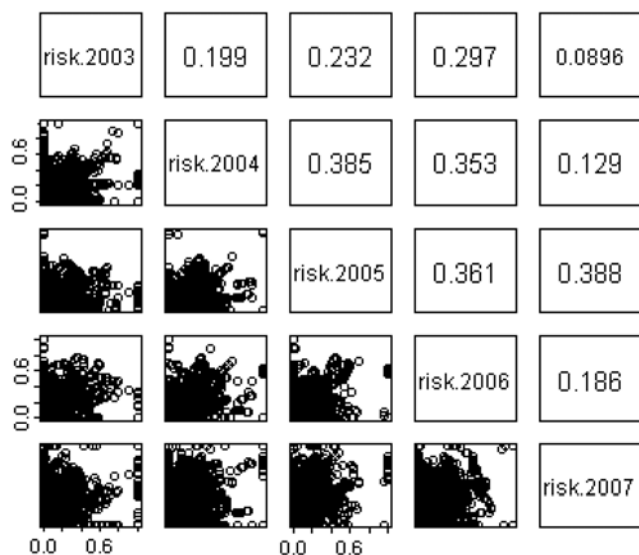
$$\beta'Z = \beta_1 c_t(x) + \beta_2 c_{t-1}(x) + \beta_3 \log(g) + \beta_4 c_t(x) \times \log(g) + \beta_5 c_t(x) \times c_{t-1}(x) \quad (4)$$

where  $g$  = the basal area and  $c_t(x)$  = the local probability of canker in year  $t$  estimated using the equation 3. Type of trees (*cop*; coppiced or not) and distance to river bank ( $d$ ) were non-significant in the different models tested (not shown).

Parameter estimates for the model (equation 4) are given in Table 1. All coefficients were statistically significant except the interaction between  $\log(g)$  and  $c_t$ , which was only marginally significant. The tests for Cox's proportional hazards assumption are given in Table 2. We noted that tests were not significant for



**Fig. 1.** Plot of local probability of canker in 2007 along of the Sarre River over the survey area and its scale. The graph follows the contours of the river from upstream (at the bottom of each graph) to downstream.



**Fig. 2.** Each graph in the lower of panel corresponds to graphs of probabilities of canker in a year  $t$  according to probability of canker in year  $t - 1$ , whereas values in the panel upper correspond to Pearson's coefficient of correlation between probabilities of canker for the two previous years.

TABLE 1. Estimates for Cox model hazard function

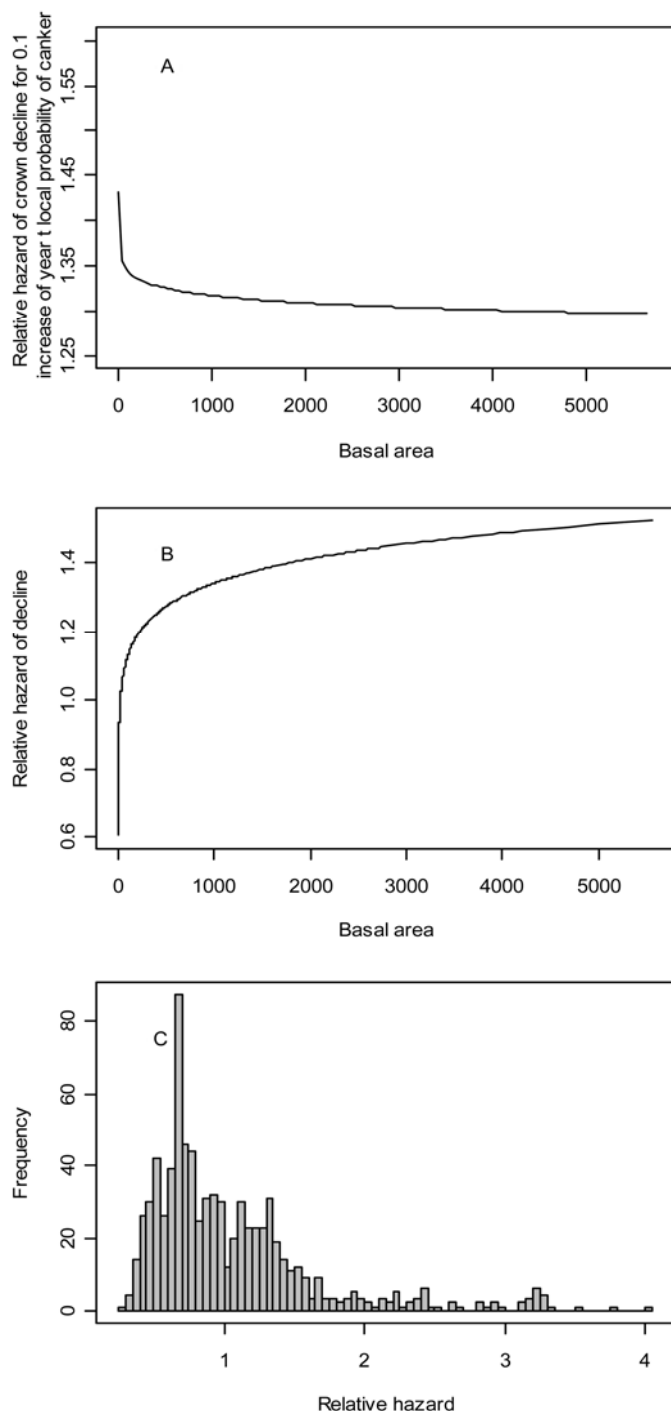
Covariate	$\hat{\beta}_i$	$P$	Lower 0.95	Upper 0.95
$c_t$	3.38	0.0	2.94	3.82
$c_{t-1}$	1.14	48.10 <sup>-06</sup>	0.58	1.68
$\log(g)$	0.09	25.10 <sup>-08</sup>	0.05	0.13
$c_t \log(g)$	-0.09	27.10 <sup>-03</sup>	-0.18	-0.01
$c_t c_{t-1}$	-2.24	13.10 <sup>-06</sup>	-3.21	-1.23

TABLE 2.  $\chi^2$  test for Cox's proportional hazards assumptions

Covariate	$\chi^2$	$P$
$c_t$	1.5281	0.216
$c_{t-1}$	0.3408	0.559
$\log(g)$	0.0752	0.784
$c_t \log(g)$	0.2011	0.654
$c_t c_{t-1}$	0.1473	0.701
Global	6.8511	0.232

covariates as well as for the global model, indicating that the assumption of proportional hazards was fulfilled.

Altogether, the hazard function of decline in year  $t$  depended on local probability of cankers in year  $t$  and  $t - 1$  and on the logarithm of trunk basal area. The exponents of coefficients  $\hat{\beta}_i$  are interpretable as multiplicative effects on the hazard. Local probability of cankers in year  $t$  had an impact on RHCD, especially for small seedlings (Fig. 3A). For large trees, local probability of cankers in year  $t$  still had a larger impact on crown decline hazard than local probability of cankers in year  $(t - 1)$  (they have risk factors of 1.3 and 1.12, respectively). Moreover,



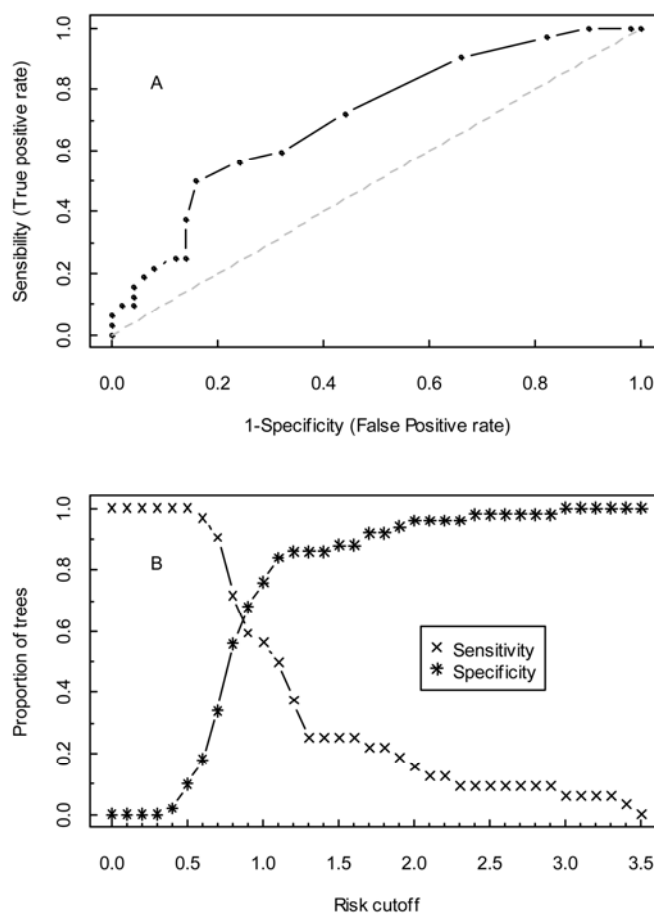
**Fig. 3.** Relative hazard of crown decline. **A**, Evolution of the relative hazard for a 0.1 increase in current year local probability of canker with increasing basal area. **B**, Evolution of the relative hazard with increasing basal area. **C**, Histogram of relative hazard for trees presenting no symptoms of *Phytophthora alni* in 2006.

seedlings (trees with small basal area) seem to have less risk of crown decline than old trees (Fig. 3B).

The RHCD estimated for all trees with no canker and no crown decline (rate 1) in 2006 was 0.1 to 4 (Fig. 3C). Although most healthy trees had RHCDs close to 1, a significant number had high risk (1.5 to 4) and may be suspected to be already root infected. Thus, determining a threshold to discriminate between infected and uninfected trees would be very useful.

**Decision threshold of RHCD to discriminate asymptomatic trees between infected to uninfected.** *P. alni* was detected on 29% of T1 trees (trees with healthy crown and no canker) in 2007 and 55% in 2008, indicating the presence of many infected, asymptomatic trees. In T1, the probability of *P. alni* detection was correlated with RHCD and the ROC curve was positively bowed away from the “no discrimination” line (Fig. 4A). This suggests that RHCD is an informative indicator. The best relative hazard threshold to separate the T1 group into two subgroups—T1a, the uninfected trees and T1b, the putatively infected asymptomatic trees—is 0.9 (Fig. 4B). Repeating the analysis in 2007 and 2008 independently yielded the same threshold value of 0.9.

The proportions of trees with detection of *P. alni* for the 5 groups (T1a, T1b, T2, T3, and T4) is shown in Figure 5A. The logarithm of tree trunk diameter included as covariate in the logistic regression model to account for tree size was not significantly correlated to the probability to detect *P. alni* ( $\chi^2 = 1.69$ ,  $P = 0.194$ ). The probability of detection for *P. alni* was marginally higher in 2008 compared with 2007 ( $\chi^2 = 3.75$ ,  $P = 0.053$ ) (Fig. 5A). By contrast, detection greatly depended on the crown status ( $\chi^2 = 11.07$ ,  $df = 4$ ,  $P = 0.026$ ) (Fig. 5A).



**Fig. 4.** Model validation and decision threshold assessment. **A**, Receiver operation characteristic curve derived from the relative hazard of crown decline (solid line) compared with the no discrimination line (straight line). **B**, Sensitivity and specificity analysis to check distinction threshold for healthy and asymptomatic infectious.

**Quantity of inoculum and tree infectious level.** We observed that *P. alni* quantification at the base of trees (through spot counts) greatly depended on the crown status of the tree (Kruskal-Wallis  $\chi^2 = 16.95$ ,  $df = 4$ ,  $P = 0.001$ ) (Fig. 5B). Also, depending on the year, *P. alni* inoculum in the soil was highest at the base of T1b or T2 trees (i.e., trees with relatively healthy crowns). At the base of those trees there was both a large frequency of *P. alni* detection and a high inoculum when detection was positive (i.e., high number of lesions on bait leaves) (Fig. 5). Trees in the group T3 (severe crown decline) seemed to produce less inoculum, especially in 2007. Far less inoculum was present at the base of dead trees and asymptomatic trees with a low predicted RHCD (Fig. 5B, T4 and T1a, respectively). The number of spots induced on bait leaves when *P. alni* was detected was also positively correlated to the log of tree trunk diameter ( $\chi^2 = 21.60$ ,  $df = 1$ ,  $P < 0.001$ ) and was significantly higher in 2008 compared with 2007 ( $\chi^2 = 107.97$ ,  $df = 1$ ,  $P < 0.001$ ).

## DISCUSSION

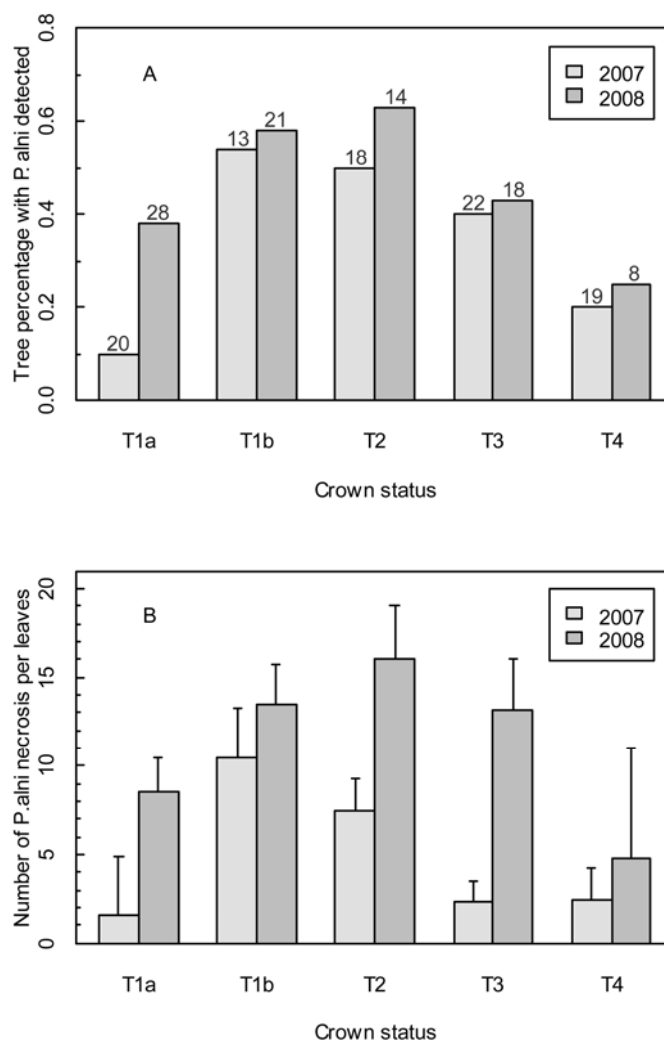
In this study of *P. alni*-induced alder decline, we showed that apparent symptoms of the disease are largely insufficient to identify infectious trees. Indeed, trees that showed no symptoms, either crown decline or cankers, were among the main inoculum producers. An estimated crown decline hazard function enabled the discrimination between healthy but infected trees and asymptomatic trees using a threshold of 0.9. Ignoring the presence of asymptomatic infected trees can result in a strong bias both in disease prevalence estimates and in the epidemic dispersal analysis (those trees being the best source of inoculum). For example, in 2007, while 67.1% of the trees showed symptoms of the disease, either cankers or crown decline, an additional 11.3% could be predicted to be infected, based on the value of the decline relative hazard function  $>0.9$ .

The approach used in this work enabled us to identify main producers of *P. alni* inoculum more easily. For root disease, host infection status is generally assessed by biological methods to check the presence of the responsible agent (11). This may quickly become heavy work when the number of surveyed plants increases ( $>705$  asymptomatic alder trees in 2007 in our survey). Thus, it is interesting to have a decision criterion for an early discrimination between potentially healthy and infected hosts. The use of decision model for disease diagnosis, prediction, or management options assessment has been growing since the middle of the 1990s in plant epidemiology and crop protection. Several studies were carried out to design and evaluate risk indicators usually derived from classic regression or discriminant analysis and used to assess a process difficult to measure directly (19,25,26,43). The novelty here is the use of the hazard function to identify infectious asymptomatic trees in a soilborne disease. There are only a few examples of the application of survival analysis to plant pathological data. Scherm and Ojiambo (32) provide a complete discussion on the use of survival analysis in botanical applications. It is important to note that there is a distinction between diagnostic and decision-making as practiced here and as generally practiced by plant pathologists. Plant pathologists often make decisions in the context of crop protection (19,43). Thus, although they diagnose at the level of the individual plant, they treat at the level of field or population for a given crop. Here, we used an approach analogous to the clinical approach, where diagnosis and decision are made at the same level (individual tree). The infectious status of each asymptomatic tree was first checked and the decision was further made about their individual contribution to the disease dispersal. Finally, although the statistical approach used provides tools to improve detection of infectious trees, it has some limits. This is particularly true for the use of Cox's hazard function (6,18) because, here, the risk of crown decline was interpreted as a risk of

infection but could not be used to strictly monitor the duration of the incubation period.

In this work, hazard function for crown decline of asymptomatic trees was estimated to depend on tree size (represented by basal area) and local probability of cankers. Increase in RHCD with increasing basal area may only indicate an increase in risk for old trees. However, this may also be related to the time during which the plant was exposed to infection because seedlings (trees with a small basal area) had only spent a few months in the survey and were less exposed to infection. Indeed, seedlings exposed to high local probability of canker in the current year have a high RHCD.

Although one of the symptoms of the disease is cankers, the approach used considers only crown decline as a critical stage of disease progression. Cankered trees are not considered as a source of inoculum if they show no root infection, as may occur when direct trunk infection via lenticels occurs during flooding. Moreover, fine root destruction is deemed the most important cause of tree death due to *P. alni* infection. Nevertheless, canker occurrence is a good marker of local pressure of infection and was used to estimate the tree infection risk. It should also be a good marker of disease dispersal because canker formation is a quick response to *P. alni* infection, in contrast to crown decline.



**Fig. 5.** Presence of *Phytophthora alni* in soil at the base of alder trees according to their crown status. **A**, Frequency of detection in soil. **B**, Amount of inoculum present in soil. Trees with no symptoms (T1a and T1b, no canker, healthy crown rated as 1 and relative hazard to crown decline respectively less and higher than 0.9); trees with crown rated as 2 (T2), 3 (T3), or 4 (T4, dead trees).

The correlation observed between RHCD and probability of detection of *P. alni* is a posteriori validation of a model used to describe tree crown decline and a use of local probability of canker as a surrogate of root infection. However, very frequent detection of *P. alni* at the base of healthy trees, even those with low RHCD, showed that most trees in the stand were in contact with pathogen inoculum. It appears to be the amount of inoculum to which trees are exposed and the frequency of exposure that matters more for disease induction.

Tree capacity to produce inoculum depends on the size of the living root system, which depends on tree size and health status. This result suggests that a tree's contribution to dispersal of *P. alni* should be weighted as a function of its size and crown status. Then, with regard to the tree's ability to produce inoculum, three groups can be separated: (i) healthy trees with RHCD <0.9 and dead trees do not produce inoculum, (ii) trees with a declining crown rated as 2 and 3 are infectious, and (iii) asymptomatic infected trees with RHCD >0.9 are the major inoculum producer in the population.

Little detection of *P. alni* could be observed on recently dead (1 to 3 years) trees, and these trees were no longer significant inoculum producers. This confirms the limited survival of *P. alni* in soil in the absence of a living host, as has already been mentioned by Jung and Blaschke (23). This is in contrast to the usually good survival in soil of *Phytophthora* spp. (10) and can be related to the absence of resistance spores in *P. alni*. Indeed, this *Phytophthora* sp. does not produce chlamydospores and the oospores produced show limited ability to germinate, resulting from the hybrid status of the pathogen (7).

Although the AUC and, therefore, the performance of the RHCD test do not appear faultless, this work provides tools that notably improve our ability to detect the main sources of inoculum in the *P. alni*-induced alder decline. Also, it revises extensively our understanding of infection dynamics and disease dispersal, because the asymptomatic infected individuals are the main inoculum producers (much more than symptomatic trees). Together with this result, the approach described above proposes a new view of plant disease epidemiology to deal with cryptic diseases, and might be extended in the monitoring of tree root diseases as well as in soilborne disease ecology.

## APPENDIX

**Estimation of local probability of canker.** The observed spatial point pattern of trees over the survey area was interpreted as a realization of a nonhomogeneous bidimensional point process which may be characterized by its intensity  $\lambda(x)$  (the density of sampled trees per unit of surface) (12). When considering point locations, one of the simplest density estimation, the "naïve estimate," counts the number of data points in a defined search area surrounding a sample point and then divides this by the search area surface to give a spatial density in number per square meter. Mathematical expression of the naïve estimator is given by:

$$\lambda(x) = \lim_{B_\tau(x) \rightarrow 0} \frac{E[N(B_\tau(x))]}{|B_\tau(x)|} \quad (5)$$

where  $N(B_\tau(x))$  is the number of trees in the area  $B_\tau(x)$  and  $|B_\tau(x)|$  is the surface of this area. In two dimensions, this naïve estimator density is calculated at a number of sample points across the area of interest to describe a continuous intensity surface.

To make the naïve estimator less influenced by the edge of the search area, because data points near the edge have a small weight, it can be improved by assigning a weight to each data point found in the search area based on some function of its distance from the sample point. The weight calculations are performed using a number of different functions given the generic

name "kernel function" and, hence, the method is known as a "kernel estimator." The kernel estimator of  $\lambda$  for a bidimensional case (35,41) is defined by:

$$\hat{\lambda}(x, \tau) = \frac{1}{\tau^2} \sum_{i=1}^n K\left(\frac{x_i - x}{\tau}\right) \quad (6)$$

with  $x$ , the bidimensional coordinates of the sample point;  $x_i$ , the bidimensional coordinates of the  $i$ th point (tree location in our case) found in the search area;  $n$ , the total number of points in the search area;  $\tau$ , the bandwidth (also named smoothing parameter); and  $K$ , the kernel, such as

$$\int_{\mathbb{R}^2} K(x) dx = 1 \text{ and } K_\tau(x) = \tau^{-1/2} K\left(\tau^{-1/2} x\right)$$

The  $B_\tau(x)$  of formula 5 is, in this case, a disc of center  $x$  and radius  $\tau$ . In our study, quartic kernel defined by

$$K(u) = \frac{15}{16} (1 - u^2)^2 1_{-1 < u < 1}$$

was used. The selection of bandwidth  $\tau$  was made in two steps. Least square cross-validation (9,40) was used to estimate the values of  $\tau$  on several portions of the surveyed area, and the one which was large enough to avoid situations without trees in the discussion was chosen. The same bandwidth  $\tau$  of 20 m was used in all estimations.

Using the previous technique, the local probability of canker in year  $t$ , defined as the probability that a tree which is located in a point  $x$  became cankered in year  $t$  given it was without cankers in year  $t - 1$ , was estimated by comparing the intensity of new cankered trees to the intensity of trees without canker the previous year (population at risk the previous year) through the estimation of the ratio:

$$P(I_t(x) = 1 | I_{t-1}(x) = 0) = \frac{\lambda_{nc}(x, t)}{\lambda_{wc}(x, t-1)} \quad (7)$$

with  $I_t(x)$ , the variable which is 1 if the located tree in  $x$  is cankered in year  $t$  and 0 otherwise;  $\lambda_{nc}(x, t)$ , the intensity of the process describing the distribution of trees that got cankered for the first time during year  $t$ ; and  $\lambda_{wc}(x, t-1)$ , the intensity of the process describing the distribution of trees without cankers in year  $t - 1$ . Indeed, if  $\lambda_p(x, t-1)$  is the intensity of the process describing the distribution of present trees in year  $t - 1$ , then:

$$P(I_t(x) = 1, I_{t-1}(x) = 0) = \frac{\lambda_{nc}(x, t)}{\lambda_p(x, t-1)} \quad (8)$$

and

$$P(I_{t-1}(x) = 0) = \frac{\lambda_{wc}(x, t-1)}{\lambda_p(x, t-1)}$$

hence, the estimated risk of canker is:

$$c_t(x) = \hat{P}(I_t(x) = 1 | I_{t-1}(x) = 0) = \frac{\hat{P}(I_t(x) = 1, I_{t-1}(x) = 0)}{\hat{P}(I_{t-1}(x) = 0)} = \frac{\hat{\lambda}_{nc}(x, t)}{\hat{\lambda}_{wc}(x, t-1)} \quad (9)$$

To obtain an estimation of a continuous risk of canker, as displayed by Figure 1,  $c_t(x)$  was calculated at a number of sample points across the survey area of interest.

## ACKNOWLEDGMENTS

We thank O. Caël, B. Bertin, and B. Fabre for their participation in the survey of the studied Sarre area for disease prevalence; P. Boizeau for help with part of the statistical analysis; J. Chadoeuf for very useful comments on the manuscript; S. Adu-Bredu for improving the English version; and the editor for useful comments.

## LITERATURE CITED

1. Bendz-Hellgren, M., and Stenlid, J. 1997. Decreased volume growth of *Picea abies* in response to *Heterobasidion annosum* infection. Can. J. For.

- Res. 27:1519-1524.
2. Bloomberg, W. J., and Morrison, D. J. 1989. Relationship of growth reduction in Douglas-fir to infection by *Armillaria* root disease in southeastern British Columbia. *Phytopathology* 79:482-487.
3. Brasier, C. M., Kirk, S. A., Delcán, J., Cooke, D. E., Jung, T., and Man In't Veld, W. A. 2004. *Phytophthora alni* sp. Nov. and its variants: designation of emerging heteroploid hybrid pathogens spreading on *Alnus* trees. *Mycol. Res.* 108:1172-1184.
4. Brasier, C. M., Rose, J., and Gibbs, J. N. 1995. An unusual *Phytophthora* associated with widespread alder mortality in Britain. *Plant Pathol.* 44:999-1007.
5. Claessens, H. 2003. The alder population in Europe. Pages 5-14 in: *Phytophthora Disease of Alder in Europe*. J. N. Gibbs, C. van Dijk, and J. Webber, eds. *For. Res. Bull.* 126.
6. Cox, D. R. 1972. Regression models and life tables. *J. R. Stat. Soc. B. Stat. Met.* 34:187-202.
7. Delcán, J., and Brasier, C. 2001. Oospore viability and variation in zoospore and hyphal tip derivatives of the hybrid alder *Phytophthoras*. *For. Pathol.* 31:65-83.
8. Diggle, P. J., Gomez-Rubio, V., Brown, P. E., Chetwynd, A. G., and Gooding, S. 2007. Second-order analysis of inhomogeneous spatial point processes using case-control data. *Biometrics* 63:550-557.
9. Duong, T., and Hazelton, M. L. 2005. Cross-validation bandwidth matrices for multivariate kernel density estimation. *Scand. J. Stat.* 15:485-506.
10. Erwin, D. C., and Ribeiro, O. K. 1996. *Phytophthora Diseases Worldwide*. American Phytopathological Society Press, St. Paul, MN.
11. Fillhart, C. R., Bachand, G. D., and Castello, D. J. 1998. Detection of infectious tobamoviruses in forest soils. *Appl. Environ. Microbiol.* 64:1430-1435.
12. Gatrell, A. C., Bailey, T. C., Diggle, P. J., and Rowlingson, B. S. 1996. Spatial point pattern analysis and its application in geographical epidemiology. *Trans. Inst. Br. Geogr.* 21:256-274.
13. Gibbs, J. N. 1995. *Phytophthora* root disease of alder in Britain. *EPPO Bull.* 25:661-664.
14. Godfrey, K. R. 1983. *Compartmental Models and their Applications*. Academic Press, London.
15. Gonçalves, R. C., Alfenas, A. C., Maffia, L. A., and Crous, P. W. 2001. Evaluation of bioassays to quantify *Cylindrocladium* inocula in soil. *Mycoscience* 42:261-264.
16. Grambsch, P., and Therneau, T. 1994. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika* 81:515-26.
17. Hanley, J. A., and McNeil, B. J. 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143:29-36.
18. Hosmer, D. W., Jr., and Lemeshow S. 2008. *Applied Survival Analysis: Regression Modeling of Time to Event Data*, second ed. John Wiley & Sons, Inc., New York.
19. Hughes, G., McRoberts, N., and Burnett, F. J. 1999. Decision-making and diagnosis in disease management. *Plant Pathol.* 48:147-153.
20. Ios, R., Andrieux, A., Marçais, B., and Frey, P. 2006. Genetic characterization of the natural hybrid species *Phytophthora alni* as inferred from nuclear and mitochondrial DNA analyses. *Fungal Genet. Biol.* 43:511-529.
21. Ios, R., Husson, C., Andrieux, A., and Frey, P. 2005. SCAR-based PCR primers to detect the hybrid pathogen *Phytophthora alni* and its sub-species causing alder disease in Europe. *Eur. J. Plant Pathol.* 112:323-335.
22. Jacquez, J. A. 1996. *Compartmental Analysis in Biology and Medicine*, third ed. Biomedware, Ann Arbor, MI.
23. Jung, T., and Blaschke, M. 2004. *Phytophthora* root and collar rot of alders in Bavaria: distribution, modes of spread and possible management strategies. *Plant Pathol.* 53:197-208.
24. Lonsdale D. 2003. *Phytophthora* disease of alder: sources of inoculum, infection and host colonisation. Pages 65-72 in: *Phytophthora Disease of Alder in Europe*. J. N. Gibbs, C. van Dijk, and J. Webber, eds. *For. Res. Bull.* 126.
25. Makowski, D., Denis, J.-B., Ruck, L., and Penaud, A. 2008. A Bayesian approach to assess the accuracy of a diagnostic test based on plant disease measurement. *Crop Prot.* 27:1187-1193.
26. Makowski, D., Taverne, M., Bolomier, J., and Ducarne, M. 2005. Comparison of risk indicators for *Sclerotinia* control in oilseed rape. *Crop Prot.* 24:527-531.
27. Marçais, B., Caël, O., and Delatour, C. 2000. Relationship between presence of basidiomes, above-ground symptoms and root infection by *Collybia fusipes* in oaks. *Eur. J. Plant Pathol.* 30:7-17.
28. McCullagh, P., and Nelder, J. 1989. *Generalized Linear Models*, second ed. Chapman and Hall, London.
29. Ojiambo, P. S., and Scherm, H. 2005. Survival analysis of time to abscission of blueberry leaves affected by *Septoria* leaf spot. *Phytopathology* 95:108-113.
30. Samuelsen, S. O., Wisloff, T. F., and Skrondal, A. 2005. A simple correction for ties when censoring times depend on covariates. *Stat. Med.* 24:3111-3121.
31. Santini, A., Barzanti, G. P., and Capretti, P. 2003. Susceptibility of some mesophilic hardwoods to alder *Phytophthora*. *J. Phytopathol.* 151:406-410.
32. Scherm, H., and Ojiambo, P. S. (2004) Applications of survival analysis in botanical epidemiology. *Phytopathology* 94:1022-1026
33. Schoenfeld, D. 1982. Partial residuals for the proportional hazards regression model. *Biometrika* 69:239-241.
34. Schwarz, G. 1978. Estimating the dimension of a model. *Ann. Stat.* 6:461-464.
35. Silverman, B. W. 1986. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London.
36. Stoyan, D., and Penttinen, A. 2000. Recent application of point process methods in forestry statistics. *Stat. Sci.* 15:61-78.
37. Streito, J. C. 2003. *Phytophthora* disease of alder: identification and distribution. Pages 25-38 in: *Phytophthora Disease of Alder in Europe*. J. N. Gibbs, C. van Dijk, and J. Webber, eds. *For. Res. Bull.* 126.
38. Streito, J. C., Legrand, P., Tabary, F., and Jarnouen de Villartay, G. 2002. *Phytophthora* disease of alder (*Alnus glutinosa*) in France: investigation between 1995 and 1999. *For. Pathol.* 32:179-191.
39. Thoirain, B., Husson, C., and Marçais, B., 2007. Risk factors for the *Phytophthora*-induced decline of alder in northeastern France. *Phytopathology* 97:99-105.
40. Wand, M. P., and Jones, M. C. 1993. Comparison of smoothing parameterizations in bivariate kernel density estimation. *J. Am. Stat. Assoc.* 88:520-528.
41. Wand, M. P., and Jones, M. C. 1995. *Kernel Smoothing*. Chapman and Hall, London.
42. Windham, A. S., and Lucas, L. T. 1987. A quantitative baiting technique for selective isolation of *Rhizoctonia zeae* from soil. *Phytopathology* 77:712-714.
43. Yuen, J., Twengström, E., and Sigvald, R. 1996. Calibration and verification of risk algorithms using logistic regression. *Eur. J. Plant Pathol.* 102:847-854.