



Identifying risk factors for European stone fruit yellows from a survey

Gael Thebaud, Nicolas Sauvion, Joël Chadoeuf, Arnaud Dufils, Gérard
Labonne

► To cite this version:

Gael Thebaud, Nicolas Sauvion, Joël Chadoeuf, Arnaud Dufils, Gérard Labonne. Identifying risk factors for European stone fruit yellows from a survey. *Phytopathology*, 2006, 96 (8), pp.890-899. 10.1094/PHYTO-96-0890 . hal-01927468

HAL Id: hal-01927468

<https://hal.science/hal-01927468>

Submitted on 30 May 2020

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.

Identifying Risk Factors for European Stone Fruit Yellows from a Survey

Gaël Thébaud, Nicolas Sauvion, Joël Chadœuf, Arnaud Dufils, and Gérard Labonne

First, second, and fifth authors: Institut National de la Recherche Agronomique (INRA), UMR BGPI, CIRAD TA 41/K, Campus International de Baillarguet, 34398 Montpellier Cedex 5, France; first and third authors: INRA, Unité de Biométrie, Domaine Saint-Paul, Site Agroparc, 84914 Avignon Cedex 9, France; and fourth author: Station Expérimentale La Pugère, Chemin de la Barque, 13370 Mallemort, France.

Accepted for publication 20 March 2006.

ABSTRACT

Thébaud, G., Sauvion, N., Chadœuf, J., Dufils, A., and Labonne, G. 2006. Identifying risk factors for European stone fruit yellows from a survey. *Phytopathology* 96:890-899.

European stone fruit yellows (ESFY) is becoming a major economic problem for *Prunus* growers in Europe. The causal agent ("*Candidatus* Phytoplasma prunorum") and its vector (*Cacopsylla pruni*) have been identified, but the present knowledge of the risk factors for this disease relies, at best, on specific experiments. To assess the relative significance of several factors correlated with ESFY incidence in the field, an exhaustive survey was performed on apricot and Japanese plum orchards in the Crau plain (France). After a preliminary multivariate exploration of the data, we used a logistic regression model to analyze and predict the cumulative number of diseased trees on the basis of a set of quantitative (age, planting density, and area of the orchard) and categorical variables (species, cultivar, and rootstock). Because of the nature of the data, we

used an overdispersed binomial model and we developed a parametric bootstrap procedure based on the beta-binomial distribution to obtain confidence intervals. Our results indicated that the age, species, and cultivar of the scion were the major factors explaining the observed number of diseased trees. The planting density and the rootstocks used in the zone under study were less significant, and the area of the orchard had no effect. The residuals of the model showed that some explanatory variables had not been taken into account, because part of the remaining variability could be explained by a grower effect. The spatial distribution of the residuals suggested that one of the reasons for this grower effect was the correlation between orchards closer than 100 m, possibly caused by the flight behavior of infectious vectors.

Additional keywords: generalized linear model, Monte Carlo, over-dispersion.

A frequent goal of epidemiological studies is to highlight the factors that are highly correlated with disease incidence or severity, or even to point at causal factors explaining the emergence of a new disease. To these aims, the analysis of surveys can be seen as an introduction or an alternative to the experimental approach, in particular (i) for quarantine diseases, (ii) when many potential factors are considered, (iii) when the investigated factors are related to the agricultural landscape (e.g., size, shape, or density of the plots), (iv) when an immense number of replicates would be necessary to achieve enough statistical power, or (v) when a lack of knowledge about the biological system could question the epidemiological significance of the experiments. For example, if an unknown insect transmits a disease, the study of risk factors by the experimental inoculation of plant material may not reflect field conditions because of the vector's behavior, among other reasons. In contrast, analyzing survey data is a way to take advantage of many "natural experiments" that occur under field conditions, where more factors can be investigated simultaneously, the side effect being a lack of a priori control by an experimental design. Such observational studies are common in animal and human epidemiology, but are rarer in botanical epidemiology, probably because the networks for high-quality data collection on a large scale are infrequent for plant diseases, and because many risk

factors can be investigated directly through cost-effective experiments. Most of these surveys linked disease incidence with climatic parameters such as temperature and moisture. Soil structure or pH (41) and the quantity and proximity in space or time of putative sources of inoculum (12,38) also have been examined. Finally, only a few observational studies have characterized how disease incidence was influenced by human-driven factors such as agricultural practices, control methods, and the choice of the cultivated genotype (21,43,47). However, for some of these variables, prophylactic practices can be defined so that the growers can make preventive choices to reduce disease incidence without the economic and environmental costs associated with the use of pesticides. Our study focused on these human-driven factors because of their high practical impact on the management of European stone fruit yellows (ESFY).

ESFY is a systemic disease affecting the genus *Prunus* (32). It has been present in Europe since at least the beginning of the 20th century (7,19), but its prevalence has increased in recent decades (30) and ESFY is now a major economic problem on apricot (*Prunus armeniaca*) and Japanese plum (*P. salicina*) in Europe. We know that "*Candidatus* Phytoplasma prunorum" (42), the causal agent of the disease, is transmitted on the persistent mode by *Cacopsylla pruni* (5), but the risk factors of ESFY are still poorly understood. Standardized captures in the field (29) and experimental breeding (3,4) showed the vector's strong host preference for blackthorn (*P. spinosa*), myrobalan (*P. cerasifera*), Japanese plum, and plum trees (*P. domestica*). Experimental vector transmissions and graft inoculations (3,4) demonstrated that there is a wide range of susceptibility to "*Ca. P. prunorum*" within the genus *Prunus*, with cherry trees being highly resistant (27) and

Corresponding author: G. Thébaud; E-mail address: thebaud@ensam.inra.fr.

DOI: 10.1094/PHYTO-96-0890

This article is in the public domain and not copyrightable. It may be freely reprinted with customary crediting of the source. The American Phytopathological Society, 2006.

Japanese plum being highly susceptible (19). Field evaluations (6) and experimental inoculations (16,26,28) also repeatedly indicated a differential sensitivity to infection between cultivars and between rootstocks. In order to assess whether the choice of rootstock and cultivar is the only risk factor of ESFY, some other potential risk factors had to be investigated, as well as their potential interactions. Addressing this question with an experimental approach would require an immense number of trees and a lot of time because the annual incidence in apricot is quite low. Thus, we preferred the alternative of analyzing survey data generated by a prophylactic program undertaken in France. This program intends to reduce the number of secondary transmissions and the regional pool of inoculum, on the basis of experience with *Plum pox virus*, another vector-borne disease of *Prunus* spp. with a high economic impact.

In the Crau plain (France), a partnership was initiated to collect epidemiologically relevant data in addition to the data that were necessary for disease management. Several potential risk factors were recorded to explain the dependent variable, which was defined as the number of diseased trees among a known number of exposed trees in each orchard. In order to analyze such binomial data, a generalized linear model (GLM) with a logit link function (i.e., logistic regression) is often preferred to the more classical linear models because it appropriately takes into account the binomial nature of the dependent variable (8,18,25,35). However, few studies have analyzed survey data with GLMs (1,13,36,38, 43). In this article, we present a logistic regression model for identifying and quantifying risk factors of ESFY through the analysis of a regional survey.

MATERIALS AND METHODS

Data record and selection. The survey extended inside a square of ≈ 25 km per side in the Crau plain, an area of apricot production in southeastern France. Most of the data were collected in 2003 by well-trained technical staff and the database was updated in 2004. In each orchard, nine variables were recorded. The dependent variable was the cumulative number of diseased trees among a known number of exposed trees from the date of orchard planting to March 2004 (the term “incidence” will denote this variable in the rest of the article). Except for a few trees that were excluded from the analysis, ESFY was the only cause of tree death. The growers frequently removed the trees with typical symptoms and new trees were often replanted, but only the initial trees were considered in the analysis. Thus, the incidence was estimated on the basis of the number of trees that were dead or removed, or with the characteristic winter symptom of early leafing. Some of the potential risk factors that we recorded were related directly to the biological characteristics of the system (species and cultivar of the scion, rootstock, surface, planting density, and age of the orchard). We also recorded human factors such as the grower and the nurseries that provided the planting material (the abbreviated names of the variables are indicated in Table 1). The mean ESFY incidence in relation to the age of the orchards was

assessed in 517 apricot orchards. For more reliability in the estimated effects, we removed the plots with missing data and those with uncommon cultivars or rootstocks. The mean of the quantitative variables and the levels of the categorical variables in the final data subset (225 orchards and $\approx 69,000$ trees from 17 farms) are summarized in Table 1. The mean ESFY incidence in these orchards was 6.3%.

Method overview. Our general approach to the statistical analysis of this survey consisted of five successive steps: (i) a multivariate analysis was performed to remove overly correlated variables from the model, (ii) the most parsimonious overdispersed logistic regression model was built by a stepwise selection of the variables, (iii) the adequacy of this final model then was checked by an analysis of the residuals and by assessing its predictive power for an external data set, (iv) we subsequently evaluated the relative significance of the different variables on disease incidence, and then (v) a test was performed on the residuals to track the remaining spatial covariates. Unless otherwise stated, all analyses were performed with the R statistical software, version 2.0.1 (39).

Selection of the variables. In order to avoid overloading the model with redundant explanatory variables, a preliminary multivariate exploration of the data was undertaken. We first performed a normalized principal components analysis (23) on the quantitative variables (cumulative incidence [Y], area [AREA], age [AGE], and planting density [DENS]), and a multiple correspondence analysis (2) on the categorical variables (cultivar [CLV], rootstock [RST], origin of the cultivar [OCLV], and origin of the rootstock [ORST]). Then, a Hill and Smith analysis (22) was performed to mix these two analyses and, thus, analyze the relationships among all variables simultaneously. The ADE-4 software (46) was used for the computation and generation of factorial maps.

Logistic regression model. To estimate simultaneously the impact of potential factors on ESFY incidence, we built a logistic regression model (35). This GLM was dedicated to the analysis of proportions arising from binomial data: in each orchard, the numbers of exposed (n_i) and diseased trees (D_i) were known. The number of diseased trees in the i th orchard initially was assumed to have a binomial distribution with probability p_i for each tree to show symptoms, whereas a given combination of k explanatory factors defined this probability p_i . Thus, the model could be written $D_i|p_i \sim B(n_i, p_i)$, where p_i was defined by

$$\ln[p_i/(1 - p_i)] = a_0 + a_{1,i}(\text{Fact}_1) + a_{2,i}(\text{Fact}_2) + \dots + a_{k,i}(\text{Fact}_k)$$

The best-fitting model was obtained by a manual stepwise procedure for selecting significant variables and biologically meaningful second-order interactions. The R function *glm* was used for model fitting by iteratively reweighted least squares.

Modeling overdispersion. In the model described above, the observed number of diseased trees (D_i) should have a binomial variance: $\text{Var}(D_i) = n p_i (1 - p_i)$. However, in observational studies, the presence of overdispersion (extrabinomial variation, here) is very common (18,35), and it also was encountered in this study. Following Collett (8), we accounted for the extrabinomial variance with Williams’ iterative algorithm (48) because the number

TABLE 1. Summary of the variables used in the preliminary multivariate analysis

Variable	Full name	Mode	Levels (frequency) or mean (range) ^a
Y	Cumulative incidence	Dependent, quantitative	6.3 (0–29.7) %
GRW	Grower	Categorical	17 levels (2–55 orchards/grower)
CLV	Cultivar	Categorical	Goldrich (27), Early Blush (33), Hargrand (66), Orangered (99)
RST	Rootstock	Categorical	Manicot-GF 1236 (11), Montclar (28), myrobalan (36), GF 305 (61), peach (89)
OCLV	Origin of the cultivar	Categorical	8 levels (1–119 orchards/origin) + 71 NA
ORST	Origin of the rootstock	Categorical	9 levels (1–118 orchards/origin) + 71 NA
AGE	Age	Quantitative	8.3 (2–15) years
AREA	Area	Quantitative	0.73 (0.08–2.5) ha
DENS	Planting density	Quantitative	410 (238–571) trees/ha

^a NA = not available.

of trees was not identical in all the orchards. Overdispersed models were fitted with the *dispmod* R package.

Model assessment. The standard goodness-of-fit criterion for GLMs (the ratio between the residual deviance and the residual degrees of freedom) is meaningless for overdispersed models (8). Hence, several other procedures were carried out in order to check the model. First, we examined the standardized deviance residuals in order to look for outlying values or for correlation with the linear predictors or with the variables (included or not in the final model). Then, after weighting each point by the corresponding number of trees in the orchard, we compared the linear regression between observed and predicted values with their expected linear relationship. We also performed both an external validation and a robustness analysis by assessing the ability of the final model to provide an accurate estimate of incidence in an independent data set composed of the 57 orchards excluded from the initial data because their rootstocks were unknown or under-represented. To obtain predictions in these new orchards, the parameters associated with the rootstock were replaced by the weighted mean of rootstock effects obtained after fitting the final model.

A characteristic of the data set was the low number of expected diseased trees in many orchards. Thus, for this discrete variable, the validity of confidence intervals based on asymptotic theorems could be questioned. Consequently, we used a parametric bootstrap approach (17) to derive confidence intervals (i) for the predicted number of diseased trees in each orchard and (ii) for the estimated value of the a_k parameters. This bootstrap procedure was performed as follows. First, the fitted overdispersed binomial model provided estimates of disease incidence in each orchard (p_i) and of an overdispersion parameter ϕ . These parameters were used (as shown in the Appendix) to draw 1,000 independent realizations for each orchard from a beta-binomial distribution, corresponding to a binomial law with extra-binomial variance (8,10, 24). For the predicted number of diseased trees in each orchard, 95% confidence intervals then could be derived from the quantiles 2.5 and 97.5% of the simulated distributions. After refitting the model (with Williams' procedure) on the 1,000 simulated data sets, the subsequent 1,000 reestimated values of each a_k parameter were used similarly to define a 95% confidence interval around their mean value.

Influence of the risk factors. The relative significance of the different factors was evaluated through deviance analysis. We initially assessed the significance of each variable alone. Afterwards, to evaluate the significance of each variable adjusted for the other variables, we used the final weights of the best overdispersed model to fit reduced models in which each factor and its interactions with other factors were removed in turn. Then, a χ^2 test was used to compare the full model with each reduced model. For disease management, quantifying the impact of the factors on the predicted incidence could be as useful as ranking their significance. Therefore, the influence of the factors was assessed through the increase of the residual sum of squares (RSS) and it was visualized by predicting disease progress for different levels of the variables (default parameters to their mode or mean: CLV = Orangered, RST = Peach, DENS = 384 trees/ha).

Spatial analysis of the residuals. The coordinates of the orchards' centroids were obtained from aerial orthophotographs (BD ORTHO, Institut Géographique National, France). The residuals then could be plotted at the location of the corresponding orchard for a visual inspection of their spatial pattern. The spatial dependence of the residuals was investigated further with a non-parametric test relying on the empirical semivariogram (34). This function, γ , is based on the squared difference between the values of the residuals separated by a distance $h \pm \varepsilon$ (i.e., residuals with coordinates s within the distance class h):

$$\gamma_{(h)} = \frac{1}{2N_{(h)}} \sum_{N(h)} (r_{(s+h)} - r_{(s)})^2$$

where $N_{(h)}$ is the number of pairs of residuals r in the distance class h . We chose 32-m-wide distance classes ($\varepsilon = 16$) in order to have at least 50 pairs of points in each class. Under the hypothesis of spatial independence, the values of the residuals should be distributed at random among the locations of the orchards' centroids. Therefore, this hypothesis of independence was challenged by a random labeling test (14,40): the function $\gamma_{(h)}$ computed on the observed residuals was compared with 1,000 random reallocations of the values of these residuals. After ordering the 1,000 simulated values of $\gamma_{(h)}$, a bilateral P value was computed, following Manly (33), as twice the proportion of simulated values more extreme or equal to the observed value of $\gamma_{(h)}$. A 95% confidence envelope also was derived from the 25th and 975th values of $\gamma_{(h)}$. The hypothesis of spatial independence between residuals should be rejected at the 5% significance level when the observed variogram is outside this envelope. However, when the model is misspecified in some way (e.g., biased estimate or different variance for some levels, or significant variable not included), the grower's trend to plant similar orchards side by side could produce spatial correlation between residuals under the null hypothesis of independence; therefore, as a precaution, the random labeling procedure was adapted to perform the permutations inside groups with homogeneous characteristics (for more details, see Manly, pages 182–199 [33]). Splitting the initial data set on the basis of the grower, cultivar, and age of the orchards defined 76 highly homogeneous groups that were used for conditioning the simulations.

RESULTS

Influence of the species. The subset that was specially selected to analyze the species effect included 15 Japanese plum orchards and 10 apricot orchards with the same characteristics: all were 6 to 8 years old on myrobalan rootstocks. The age of the orchards had no significant influence on disease incidence (data not shown); thus, the resulting model for p_i was simply $\ln[p_i/(1 - p_i)] = a_{1,\text{species}}$. The estimated disease incidence (23.1% for the Japanese plum and 5.96% for the apricot plots) and the corresponding confidence intervals indicated that 6 to 8 years after planting, Japanese plum orchards were considerably more affected by ESFY than apricot orchards (Table 2). An analysis of deviance showed that, even after allowing for overdispersion ($\phi = 0.019$), this four-fold effect of the species was highly significant ($P = 1.7 \times 10^{-14}$,

TABLE 2. Estimated mean and confidence intervals for the effect of the species on European stone fruit yellows incidence (one-variable overdispersed generalized linear model)

Species	Logit scale		Response scale ^a	
	Estimate $\pm$ SE ^b	Estimate	Bootstrap confidence interval ^c	Asymptotic confidence interval ^d
<i>Prunus armeniaca</i>	-2.76 $\pm$ 0.209	0.060	0.038–0.085	0.039–0.089
<i>P. salicina</i>	-1.20 $\pm$ 0.097	0.231	0.197–0.265	0.197–0.268

^a Estimated disease incidence and the corresponding 95% confidence interval are back-transformed from the logit scale.

^b SE = standard error.

^c Parametric bootstrap procedure is described in the text.

^d From a t distribution on 23 degrees of freedom.

χ^2 test on 1 degree of freedom [df]). Therefore, in the rest of the study, the few orchards planted with Japanese plum trees were excluded, so that our model only analyzed the incidence of ESMY in apricot orchards.

Multivariate analysis. When initially included in the preliminary multivariate analysis, the grower (GRW) completely defined the first factorial plane and masked any correlation between the variables. Thus, it was treated as a supplementary variable in this

analysis (Fig. 1), which showed that the data set was quite structured because many categorical variables were interrelated. The projections of the orchards on the first factorial plane (Fig. 1A) could be subdivided into distinct groups (denoted I to V) sharing some specific combinations of the variables, as indicated by Figure 1K and by a visual comparison between maps in Figure 1A to J. The most significant multicorrelation involved sparse and older orchards with cultivar 1 (Early Blush), origin 6, and rootstocks 2

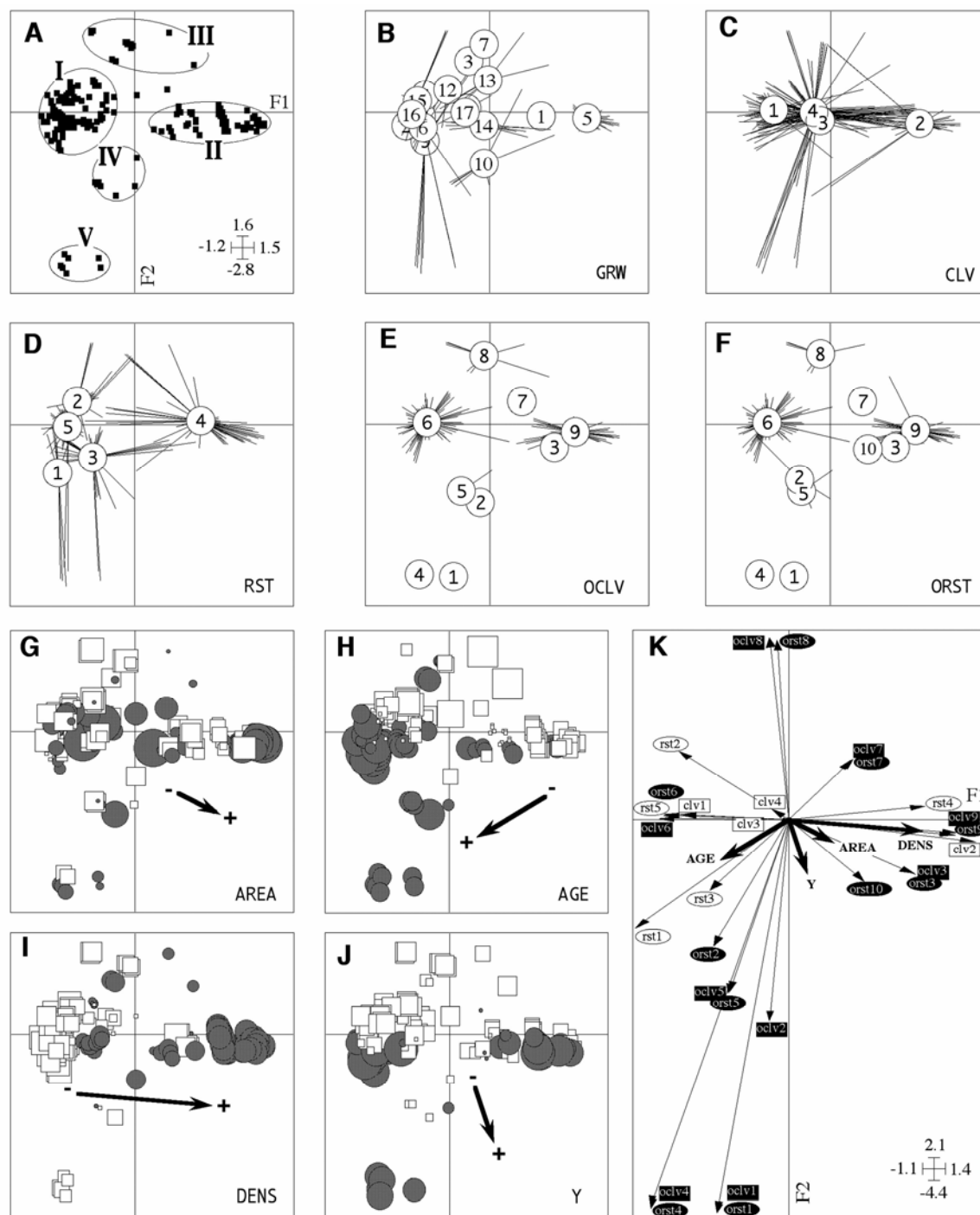


Fig. 1. Factorial maps from a multivariate Hill & Smith analysis coupling the categorical and quantitative variables that describe the 225 apricot orchards. **A**, Projections of the orchards on the first factorial plane (F1 $\times$ F2, F1 and F2 representing 13.8 and 8.5% of the total inertia, respectively). **B to F**, Categorical variables. For a given variable, each circle represents the mean position (barycenter) of one modality on F1 $\times$ F2, in connection with all the associated orchards. The supplementary variable grower (GRW) was not used to define F1 $\times$ F2 (see text). **G to J**, Quantitative variables. For a given variable, each orchard is represented by a gray circle (positive value) or a white square (negative value) with an area proportional to the absolute value of the normalized variable. **K**, Projection of the variables defining F1 $\times$ F2. Thin arrows: categorical variables; thick arrows: quantitative variables. Correlated variables have collinear vectors; the weight of the variables in the analysis increases with their distance from the center of the graph. CLV = cultivar, RST = rootstock, OCLV = origin of the cultivar, ORST = origin of the rootstock, AREA = area, AGE = age, DENS = planting density, and Y = cumulative incidence.

(Montclar) and 5 (GF 305) in group I; or dense and young orchards with cultivar 2 (Goldrich), rootstock 4 (peach), and origins 3 and 9 in group II. The only dependence between quantitative variables was the slight anticorrelation of AGE and DENS ($r = -0.29$). As expected, OCLV and ORST were almost completely correlated (Fig. 1E, F, and K). Additionally, these variables were strongly unbalanced and included 71 missing values; hence, they had to be removed. The supplementary variable GRW clearly was structured by the other variables (Fig. 1B) and, thus, strongly correlated to them. We were more interested in these underlying correlates; therefore, the descriptive variable GRW was not included in the model until the ultimate step of the analysis. Finally, Figure 1K showed that no explicative factor or variable was obviously correlated to the dependent variable Y, further highlighting the need for an explicative model.

GLM. An initial GLM was built with the five remaining variables and their biologically meaningful interactions. AREA then was dropped because it was not significant ($P = 0.24$, χ^2 test on 1 df). The deviance of the resulting model was much higher than twice the number of degrees of freedom, a value that is used as a rule of thumb for the detection of overdispersion (31). Because the analysis of the residuals indicated no obvious problem with the model, intraorchard dependence was the most probable cause of overdispersion. Thus, the model was corrected to allow extra-binomial variation ($\phi = 0.042$) so as not to overestimate the significance of the effects. The remaining four variables and four interactions were significant (all P values $< 4.3 \times 10^{-3}$); therefore, the final model for p_i was $\ln[p_i/(1 - p_i)] = a_0 + a_{1,i}(\text{CLV}) + a_{2,i}(\text{RST}) + a_3 \times \text{AGE}_i + a_4 \times \text{DENS}_i + a_{4,i}(\text{CLV}:\text{AGE}) \times \text{AGE}_i + a_{6,i}(\text{RST}:\text{AGE}) \times \text{AGE}_i + a_{7,i}(\text{RST}:\text{DENS}) \times \text{DENS}_i + a_8 \times \text{AGE}_i \times \text{DENS}_i$.

This model then was checked. One point appeared to be overly influential on the basis of Cook's distance (9). Because the analyses provided consistent results with or without this point, it was not discarded from the data set. The asymptotic results and parametric bootstrap distributions consistently identified only one slightly outlying value that was conserved because the associated information was accurate. A quantile-quantile plot indicated that the normal distribution roughly approximated the distribution of the standardized deviance residuals (Fig. 2A). There was no indication of inadequacy in the linear predictors because they were not correlated to the residuals. When plotted against the included variables, the residuals showed no particular trend other than a slight overestimation of ESFY incidence in young (2- to 4-year-old) orchards. The residuals also were distributed randomly with respect to the excluded variable AREA. In contrast, the residuals significantly differed with respect to the grower (Fig. 2C) and origin of the planting material (Fig. 2D): the observed mean of the residuals frequently was outside the range expected under the null hypothesis of independence. This fact indicates that the human factors not only summarized the other variables, but also had an additional significant effect that remained even after including some of the underlying factors in the model.

The weighted linear regression between observed and predicted incidence (Fig. 2B) had the following equation: Observed = $0.946 \times \text{Predicted} + 0.003$ ($R^2 = 0.49$ and standard error [SE] of 0.065 and 0.006, respectively). On the logit scale, an R^2 of 0.42 confirmed that the fit of the model was acceptable for an observational study; this model captured half of the variability of the data with a relatively small number of parameters (21 of the initial 224 df were used). The same procedure carried out on the external data

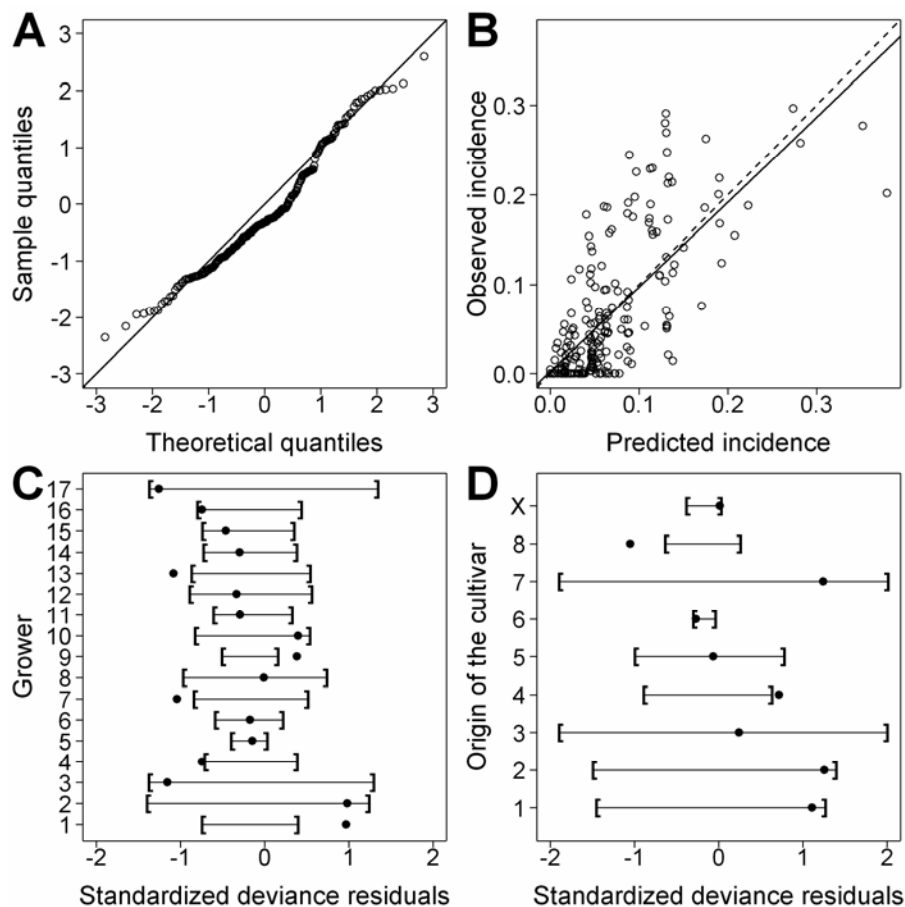


Fig. 2. Examination of the model fit. **A**, Normal quantile-quantile plot of the residuals. **B**, Linear regression of observed against predicted incidence (solid line) and expected line (dashed). **C** and **D**, Plot of the residuals by grower (GRW) and origin of the cultivar (OCLV), respectively. Points: mean of the observed residuals; brackets and solid line: 95% confidence interval for the expected mean under the null hypothesis of independence between residuals and GRW or OCLV, respectively. Under the null hypothesis, the points should lie between the brackets.

set resulted in a higher R^2 of 0.63, but the model tended to overestimate the incidence, as shown by the equation of the regression line: Observed = $0.870 \times \text{Predicted} - 0.016$ (SE = 0.088 and 0.009, respectively). Thus, despite the lack of information on the rootstock in the validation data set, the model still provided an acceptable prediction of ESFY incidence.

The extent of the confidence intervals for some parameters of the model (Table 3) confirmed that the four variables and four interactions included in the model only explained part of the variability of the observed incidence. For these parameters, the discrepancy between bootstrap and the asymptotic confidence intervals was sometimes quite high (e.g., for the rootstocks), indicating that the assumptions of the asymptotic results were not fulfilled. Thus, in the rest of the study, we used the more reliable and more conservative bootstrap intervals.

Influence of the risk factors. ESFY incidence in relation to the age of the orchards (Fig. 3A) was summarized by the following equation: $Y = 1/(1 + e^{-0.198t + 4.69})$. However, this binomial overdispersed model was not satisfactory ($R^2 = 0.1$), thus indicating a

major role for the other variables. The analysis of deviance for the one-variable models (Table 4) showed that the grower was the best single explanatory factor, much better than the cultivar or the other variables. However, because of the observed multicorrelation, it was statistically more accurate to assess the role of each variable after adjusting for the effect of the others. The corresponding deviance analysis (Table 5) showed that tree age was the most significant variable, followed by both density and cultivar, and then the rootstock, with the area of the orchard having no influence on disease incidence. These results were robust to some alterations of the model, because congruent conclusions (except that CLV became much more significant than DENS) were drawn when the same procedure was performed after fitting a simpler model without the interaction terms (data not shown). The influence of the main factors on the predicted incidence, assessed on the basis of the RSS, ranked in the same order as their significance. The interaction between the density and the age of the orchard was more significant than any other interaction. No conclusion could be drawn from the model concerning the effect of

TABLE 3. Estimates and confidence intervals for the parameters of the final overdispersed generalized linear model^a

Parameter ^b	Estimate $\pm$ SE ^c	Bootstrap confidence interval ^d		Asymptotic confidence interval ^e	
Intercept	-24.20 $\pm$ 6.36	-94.9	-16.8	-36.7	-11.7
CLV Goldrich	1.31 $\pm$ 1.37	-1.06	4.34	-1.39	4.00
CLV Hargrand	-1.24 $\pm$ 1.20	-3.45	1.46	-3.61	1.13
CLV Orangered	0.85 $\pm$ 1.17	-1.31	3.50	-1.45	3.15
RST Montclar	-0.12 $\pm$ 6.66	-10.5	67.7	-13.3	13.0
RST myrobalan	9.87 $\pm$ 6.36	2.21	80.2	-2.66	22.4
RST peach	11.50 $\pm$ 6.14	4.24	82.5	-0.61	23.6
RST GF305	11.50 $\pm$ 6.10	4.48	82.5	-0.57	23.5
AGE	1.54 $\pm$ 0.39	0.95	8.30	0.76	2.31
DENS ($\times 10^{-2}$)	3.67 $\pm$ 1.53	0.46	8.42	0.65	6.69
CLV Goldrich:AGE ($\times 10^{-1}$)	-1.59 $\pm$ 1.57	-5.28	1.12	-4.67	1.50
CLV Hargrand:AGE ($\times 10^{-1}$)	1.50 $\pm$ 1.21	-1.12	3.76	-0.89	3.89
CLV Orangered:AGE ($\times 10^{-1}$)	-0.08 $\pm$ 1.19	-2.71	2.20	-2.43	2.27
RST Montclar:AGE	-0.21 $\pm$ 0.34	-7.00	0.32	-0.88	0.46
RST myrobalan:AGE	-0.55 $\pm$ 0.34	-7.43	-0.07	-1.22	0.12
RST peach:AGE	-0.57 $\pm$ 0.33	-7.41	-0.09	-1.23	0.09
RST GF305:AGE	-0.59 $\pm$ 0.33	-7.40	-0.14	-1.24	0.06
RST Montclar:DENS ($\times 10^{-2}$)	0.24 $\pm$ 1.54	-4.14	3.53	-2.79	3.27
RST myrobalan:DENS ($\times 10^{-2}$)	-1.36 $\pm$ 1.48	-5.79	1.67	-4.28	1.56
RST peach:DENS ($\times 10^{-2}$)	-1.67 $\pm$ 1.47	-6.16	1.45	-4.56	1.23
RST GF305:DENS ($\times 10^{-2}$)	-1.46 $\pm$ 1.48	-6.04	1.51	-4.37	1.46
AGE:DENS ($\times 10^{-3}$)	-2.08 $\pm$ 0.53	-3.18	-1.08	-3.12	-1.04

^a All the values are on the logit scale.

^b CLV = cultivar, RST = rootstock, AGE = age, and DENS = planting density.

^c SE = standard error.

^d Parametric bootstrap procedure is described in the text.

^e From a t distribution on 203 degrees of freedom.

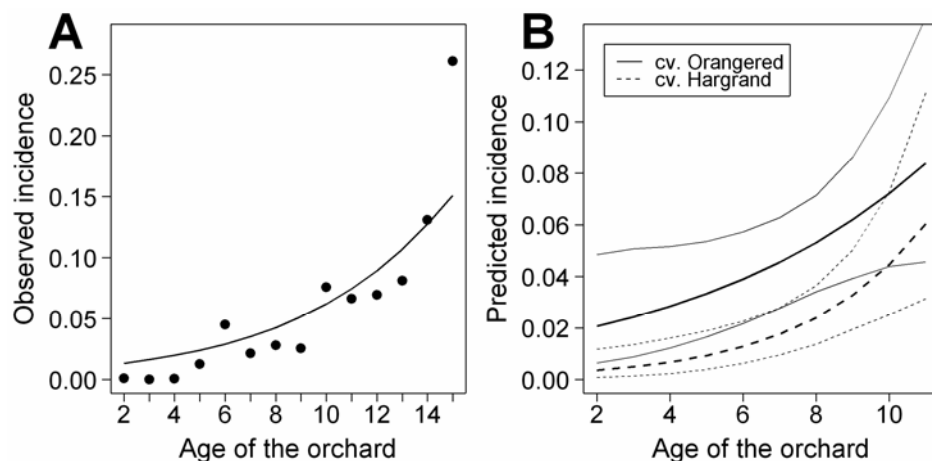


Fig. 3. European stone fruit yellows incidence in relation to the age of the orchards. **A**, Observed in the initial data set (517 orchards) and **B**, predicted for cvs. Orangered and Hargrand from the final data subset (including 225 orchards); mean prediction (bold lines) with the associated 95% bootstrap confidence envelopes (thin lines).

different levels of the factors, because the interactions were significant. The predictions for specific values generally were inconclusive as well because of the wide confidence intervals. However, the analysis of the model without interaction terms indicated that the incidence was significantly higher on GF 305 rootstock and on the cv. Orangered. This was further confirmed by the predicted ESFY progress in cvs. Orangered and Hargrand with the complete model (Fig. 3B).

Spatial analysis of the residuals. A simple map of the standardized deviance residuals clearly pointed out the spatial correlation between residuals. However, this correlation could have been at least partly explained by the spatial factor GRW (Fig. 2C), which had been discarded from the model. Thus, the mean of each level of this factor was deducted from the corresponding residuals. Then, the grower-adjusted residuals were used to compute an empirical variogram and its confidence envelopes obtained by random labeling within homogeneous subgroups (Fig. 4). The first values of $\gamma(h)$ (up to a distance of 100 m) were below or near the 95% confidence envelope (P values ranging from <0.002 to 0.062), indicating that close orchards tended to behave similarly. Thus, the spatial dependence between close orchards was highly significant, even after having adjusted for the effect of the other variables, including the grower and the location of simi-

lar plots. Given the several conservative corrections that were applied and the conservative definition of the homogeneous groups, this range of dependence may be slightly underestimated.

DISCUSSION

Risk factors. The aim of this study was to provide insight into the risk factors of ESFY on the basis of a regional survey. Some of these factors are related to the genetics or to the structure of the orchards, others are linked with unidentified human factors, and some have a spatial component. Our results demonstrated that the age of the orchard had the expected effect on the cumulative disease incidence. Nevertheless, several other factors were influential enough to be detected and ranked. Among the other factors, the species of the scion (*P. armeniaca* or *P. salicina*) was the most influential factor, with the incidence in 6- to 8-year-old Japanese plum trees being fourfold that of apricot trees of the same age when the two species are planted in the same zone. In fact, the progress of ESFY is so fast in Japanese plum orchards that it restricts the cultivation of this species in Europe (15). The difference between the two species probably is caused by a strong host preference of the vector *C. pruni*, because its relative density on these two hosts (29) and its better fecundity and life expectancy in experimental conditions (4) demonstrate that Japanese plum trees are more suitable hosts for the vector. The results (Tables 4 and 5) show that the scion cultivar was the second influential factor on disease incidence, with cv. Orangered more heavily infected than the other cultivars. This could be the result of a high attraction to the vector, or the consequence of a high level of susceptibility or sensitivity to the pathogen. Thus, to reduce the cost of ESFY, growers should take into account the relative disease risk of the planted scion. The planting density stood at a surprising third rank, which points out original speculative explanations: denser orchards could be more attractive, could influence the mobility of the vector, or could speed up symptom expression in plants that are under more severe stress. The rootstock played an unexpected minor role in the system, which was confirmed by the relatively good prediction obtained even when it is unknown or different from the subset used to build the model. This apparent discrep-

TABLE 4. Analysis of deviance for each one-variable generalized linear model (without overdispersion)

Variable ^a	Δdf^b	Deviance	AIC ^c	P value ^d
None	...	5,761	6,514	...
GRW	16	4,273	5,057	$<10^{-40}$
CLV	3	4,970	5,728	$<10^{-40}$
OCLV	8	5,298	6,066	$<10^{-40}$
ORST	9	5,338	6,108	$<10^{-40}$
AREA	1	5,449	6,203	$<10^{-40}$
RST	4	5,613	6,374	5.8×10^{-31}
AGE	1	5,643	6,397	1.7×10^{-27}
DENS	1	5,756	6,510	2.1×10^{-2}

^a GRW = grower, CLV = cultivar, OCLV = origin of the cultivar, ORST = origin of the rootstock, AREA = area, RST = rootstock, AGE = age, and DENS = planting density.

^b Difference between the number of degrees of freedom (df) in the null model and the number of df in each one-variable model.

^c Models are sorted by increasing Akaike information criterion (AIC) ($AIC = -2 \times \log\text{-likelihood} + 2 \times df$); a smaller value of AIC indicates a more parsimonious fit.

^d From a χ^2 distribution on Δdf .

TABLE 5. Analysis of deviance for each variable adjusted for the other effects (final model) in the generalized linear model

Variable ^a	RSS ^b	Δdf^c	Deviance	P value ^d
Full model	0.62	...	203.3	...
Main effect				
AGE	1.07	9	360.5	2.77×10^{-27}
DENS	0.79	6	243.4	4.18×10^{-7}
CLV	0.76	6	242.1	7.67×10^{-7}
RST	0.76	12	250.5	4.19×10^{-6}
Interaction only				
AGE:DENS	0.67	1	219.4	5.99×10^{-5}
RST:AGE	0.74	4	223.2	5.13×10^{-4}
CLV:AGE	0.71	3	218.5	1.64×10^{-3}
RST:DENS	0.70	4	218.5	4.27×10^{-3}

^a Main effect = main effect and its interactions, AGE = age, DENS = planting density, CLV = cultivar, and RST = rootstock.

^b Residual sum of squares: $RSS = \sum_{i=1}^{225} (p_{i,observed} - p_{i,estimated})^2$, where p_i denotes the incidence in the i th orchard. Higher values of RSS indicate a higher influence of the factor on the predicted incidence.

^c Difference between the number of degrees of freedom (df) in each reduced model and the number of df in the full overdispersed model.

^d From a χ^2 distribution on Δdf .

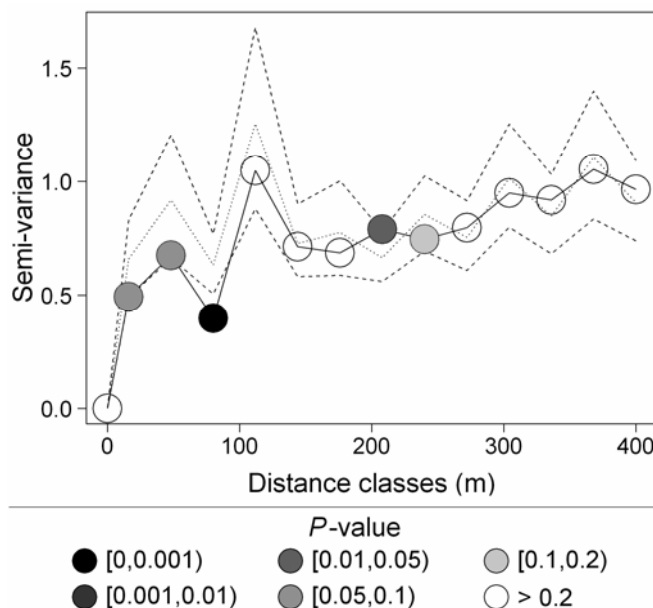


Fig. 4. Bilateral random labeling test ($\alpha = 5\%$) of spatial independence between the standardized deviance residuals. Solid line: variogram of the observed residuals; dotted line: mean of the simulated values; dashed lines: 95% confidence envelope. After a correction for the grower effect, the residuals were randomized conditional on the similarity between the orchards (see text).

ancy with previous observations indicating a major role of the rootstock in disease progress (28,37) may be caused by the excessive homogeneity of the rootstocks in the zone under study (80% of the rootstocks being peach cultivars). However, a competing explanatory hypothesis is that a faster visual detection of diseased trees (enabled by acute symptoms) has a decreasing influence on the cumulative incidence as the orchards grow old. This also could explain the significant interaction between the age of the orchard and the rootstock.

Both human variables had a significant impact on ESFY incidence, but the growers had much more influence than the nurseries (Table 4; Fig. 2C and D). The grower was the best informative one-variable model (Table 4). On the one hand, a large part of this high significance was the result of the correlation between the grower and many other variables (Fig. 1), with the grower being a good summary of several variables. On the other hand, the analysis of the residuals unequivocally demonstrated the existence of a grower effect not explained by the other variables (Fig. 2C). This result indicates that at least one grower-specific risk factor, though significant, was not included in the model. In practice, a more in-depth analysis of the differences in the agricultural practices of the growers with extremely low and extremely high incidence could point to the interesting factors. The level of prophylaxis, the insecticide protection, and the location of *P. spinosa* hedges are among the additional factors that could be investigated. Concerning the apparent nursery effect, we cannot completely rule out the possibility that some nurseries have differential levels of exposure to infectious vectors. However, it is more probably a case of confounding that results from the strong correlation between growers and nurseries (Fig. 1B and E). The difference in incidence between nurseries (Fig. 2D) is considerably reduced by the use of grower-adjusted residuals, whereas the grower effect (Fig. 2C) is only slightly attenuated by the use of the symmetrical nursery-adjusted residuals (data not shown).

The spatial dependence was significant up to 100 m, and cannot be explained by an underlying grower effect because we used grower-adjusted residuals. Neither could it be the biologically uninteresting result of a spatial proximity between orchards with similar characteristics, because we simulated the null hypothesis conditional on such similarity. Several hypotheses can be proposed to account for the remaining spatial dependence. It might be indirectly caused by underlying physical spatial factors that have not been recorded, such as the impact of soil characteristics on symptom expression. The spatial dependence also can originate from some properties of the vectorial transmission of ESFY. The most obvious explanation comes from the presence in the data of cultivar mixtures within some plots (with very close centroids, all closer than 41 m), where the vector could transmit the phytoplasma equivalently to one cultivar or the other. However, because this is not sufficient to generate dependence up to 100 m, other hypotheses are required, but they are more speculative. The population density of *C. pruni* could be higher in some places, thereby forming small patches with a range of action limited to the adjacent orchards. The other processes that can play a role are either multiple primary or secondary transmissions occurring across orchards, mainly in a radius of 100 m (which, in the study area, often corresponds to the nearest orchard only). For testing hypotheses related to these processes of disease spread, it would be interesting to analyze in detail the individual trees and the spatiotemporal pattern of diseased trees (45).

This kind of model-based analysis of a case study can raise questions related to its generality. The results showed that our model was more efficient for identifying and ranking the risk factors of ESFY than for predicting disease progress, because of the significant unexplained variability. However, the model is suitable for an approximate evaluation of disease progress in the Crau plain as far as it is not used to extrapolate to other levels of the factors (except for other rootstocks). This allows participating

growers to integrate the cost of disease control in the evaluation of the profitability of their orchards. Of course, outside the Crau plain, it would be inappropriate to make predictions with this model. The results on the risk factors clearly show that, even if the conclusions concerning particular levels of these factors may be of only local interest, the respective significance of the different factors is expected to be quite general. In this regard, the demonstration of a substantial effect of the grower should be highlighted because it indicates that, in addition to the choice of the cultivar, some agricultural practices can reduce or increase the incidence of ESFY.

Validity of the GLM. The data, by some of their features (dependence and small binomial coefficients), did not strictly correspond to the assumptions of the logistic regression model. These problems frequently arise in the analysis of surveys with GLMs; therefore, we discuss how they have been detected and handled. The spatial correlation between residuals can be seen as an unwelcome characteristic that does not meet the assumption of independence between observations, which underlie most of the common statistical models. It also can be seen as a biologically meaningful feature that may motivate further investigation. Whatever the purpose, simple and versatile nonparametric tests based on random labeling can be included in the final steps of the analysis of the data to track spatial dependence in the residuals (33,41). In this large-scale study, a multiscale exploration of spatial dependence using a variogram allowed the detection of a significant spatial correlation at a short distance (up to 100 m) between the residuals of the model. Taking this spatial correlation into account in the analysis is an interesting prospect for this study; however, building a statistically irreproachable model to this aim would require developing cutting-edge spatial statistics methods that extend far beyond the scope of this article. Moreover, the hierarchy of the factors is expected to be quite robust to the incorporation of spatial effects in the model, and the observed *P* values (Table 5) are so low that all the corresponding effects still would be significant after an improbable 10-fold correction. In addition to the dependence between orchards, the assumption of independence can be challenged within the orchards, leading to extrabinomial variation. This phenomenon can be suspected when the fit of a binomial model is unsatisfactory though nothing in the residuals indicates an incorrect specification of the model (35). In this study, both sources of dependence were identified (i.e., within and between orchards). A previous analysis of the spatial pattern of diseased trees within an orchard also indicated dependence between transmission events (44). These results may be partly explained by short-distance secondary transmissions. Moreover, because the phytoplasma is persistently transmitted by *C. pruni* (5), multiple transmissions also may account for the observed dependence within (and to a lesser extent, between) orchards. Finally, it should be mentioned that the data experience a slight dependent censoring, because some orchards have been removed in the past when the growers considered that they were too heavily infected. Thus, the increase in incidence as a function of time underestimates the real disease progress.

When some predicted values (p_i) are close to zero or when the number of individuals in some orchards (n_i) is very low, the assumptions for the asymptotic results are not met. This situation can arise quite frequently in epidemiological surveys. In such circumstances, parametric bootstrap procedures could be used more frequently when satisfactory resampling models are available. Here, we expressed the parameters of the beta-binomial distribution as a function of ϕ and p_i (the overdispersion parameter and estimated proportions, respectively). To our knowledge, it is a new result that allows the computing of parametric bootstrap confidence intervals for overdispersed logistic models. In practice, bootstrap and asymptotic methods gave consistent results in the identification of outliers and similar confidence intervals for the species effect (Table 2), as well as for half of the parameters of

the final model (Table 3). For the other parameters, the bootstrap intervals were wider and, hence, more conservative.

Concluding remarks. Performing a large-scale disease assessment for a control program (e.g., 11,12,20) can provide reliable information for epidemiological studies. In such situations, a close collaboration with plant protection services from the initial steps is a prerequisite to ensure that some epidemiologically important data which are inexpensively and easily obtained are not omitted because of their lack of relevance to the short-term control program. Our results highlight the need for a rigorous experimental or field evaluation of the sensitivity of different cultivars to ESFY as a basis for future genetic improvement. In addition, more detailed investigations of the grower-specific agricultural practices possibly would identify other risk factors, and the analysis of the spatiotemporal point pattern formed by the diseased trees probably would provide insight into the transmission behavior of *C. pruni*.

APPENDIX

Here, we show how the parameters of a beta-binomial model can be derived from an overdispersed binomial model fitted by the Williams procedure (48), which provides estimates of both p_i and the overdispersion parameter ϕ .

Let D_i be a binomial random variable: $D_i \sim B(n_i, p_i)$. The variance of D_i is: $\text{Var}(D_i) = n_i p_i (1 - p_i)$. As explained by Collett, pages 192–195 (8), the variance of an overdispersed binomial variable D_i can be written: $\text{Var}(D_i) = n_i p_i (1 - p_i) [1 + (n_i - 1)\phi]$. This is, in particular, the variance of a beta-binomial random variable D_i in which $D_i | P_i \sim B(n_i, p_i)$, P_i having a beta distribution with $E(P_i) = p_i$ and $\text{Var}(P_i) = \phi p_i (1 - p_i)$. The mean and variance of a random variable P_i with a beta distribution (with shape parameters α_i and β_i) are $E(P_i) = \alpha_i / (\alpha_i + \beta_i) = p_i$, and $\text{Var}(P_i) = p_i (1 - p_i) / (\alpha_i + \beta_i + 1)$; therefore, we obtain by identification and resolution of the subsequent two-parameter equation:

$$\alpha_i = p_i \left(\frac{1}{\phi} - 1 \right) \text{ and } \beta_i = (1 - p_i) \left(\frac{1}{\phi} - 1 \right)$$

Thus, the parameters p_i and ϕ estimated by Williams' algorithm define the parameters α_i and β_i of the beta-binomial distribution. These parameters then can be used in a parametric bootstrap procedure to derive confidence intervals for the predicted incidence in the orchards (p_i) and for the parameters (a_k) of the logistic regression model.

ACKNOWLEDGMENTS

This study was partly funded by the Conseil Régional de PACA. We thank the growers of Crau, I. Ricavy and Y. Fradin (experimental station La Pugère), and the Chambre d'Agriculture des Bouches-du-Rhône for data collection and for creating and actualizing the database; E. Klein and C. Bruchou for helpful discussions; J.-N. Bacro and S. Dallot for critically reviewing the manuscript; and M. Hilf for improving the English.

LITERATURE CITED

- Aegerter, B. J., Nuñez, J. J., and Davis, R. M. 2003. Environmental factors affecting rose downy mildew and development of a forecasting model for a nursery production system. *Plant Dis.* 87:732-738.
- Burt, C. 1950. The factorial analysis of qualitative data. *Br. J. Stat. Psychol.* 3:166-185.
- Carraro, L., Ferrini, F., Ermacora, P., and Loi, N. 2002. Role of wild *Prunus* species in the epidemiology of European stone fruit yellows. *Plant Pathol.* 51:513-517.
- Carraro, L., Ferrini, F., Ermacora, P., and Loi, N. 2004. Transmission of European stone fruit yellows phytoplasma to *Prunus* species by using vector and graft transmission. *Acta Hortic.* 657:449-453.
- Carraro, L., Loi, N., and Ermacora, P. 2001. Transmission characteristics of the European stone fruit yellows phytoplasma and its vector *Cacopsylla pruni*. *Eur. J. Plant Pathol.* 107:695-700.
- Carraro, L., Osler, R., Refatti, E., and Favali, M. A. 1992. Natural diffusion and experimental transmission of plum leptonecrosis. *Acta Hortic.* 309:285-290.
- Chabrolin, C. 1924. Quelques maladies des arbres fruitiers de la vallée du Rhône. *Ann. Epiphyties* 10:263-338.
- Collett, D. 1991. *Modelling Binary Data*, 1st ed. Chapman and Hall, London.
- Cook, R. D., and Weisberg, S. 1982. *Residuals and Influence in Regression*. Chapman and Hall, London.
- Crowder, M. J. 1978. Beta-binomial ANOVA for proportions. *Appl. Stat.* 27:34-37.
- Dallot, S., Gottwald, T., Labonne, G., and Quiot, J.-B. 2003. Spatial pattern analysis of Sharka disease (*Plum pox virus* strain M) in peach orchards of southern France. *Phytopathology* 93:1543-1552.
- Dallot, S., Gottwald, T., Labonne, G., and Quiot, J.-B. 2004. Factors affecting the spread of *Plum pox virus* strain M in peach orchards subjected to roguing in France. *Phytopathology* 94:1390-1398.
- De Wolf, E. D., Madden, L. V., and Lipps, P. E. 2003. Risk assessment models for wheat Fusarium head blight epidemics based on within-season weather data. *Phytopathology* 93:428-435.
- Diggle, P. J. 1983. *Statistical Analysis of Spatial Point Patterns*. Mathematics in Biology. Academic Press, London.
- Duval, H. 1999. Prunes japonaises: Un défi à relever. *Arboricult. Fruit.* 524:35-41.
- Duval, H., Jullian, J.-P., and Lemaire, J.-M. 1999. Enroulement chlorotique de l'abricotier: Evaluation de la sensibilité des principales variétés de prunier japonais... et du nouveau sur le vecteur. *Phytoma* 516:38-40.
- Efron, B., and Tibshirani, R. J. 1993. *An Introduction to the Bootstrap*. Chapman and Hall, New York.
- Garrett, K. A., Madden, L. V., Hughes, G., and Pfender, W. F. 2004. New applications of statistical tools in plant pathology. *Phytopathology* 94:999-1003.
- Goidanich, G. 1933. Un deperimento dei susini. *Boll. Reg. Staz. Patol. Veg. Roma* 13:160-173.
- Gottwald, T. R., Sun, X., Riley, T., Graham, J. H., Ferrandino, F., and Taylor, E. L. 2002. Geo-referenced spatiotemporal analysis of the urban citrus canker epidemic in Florida. *Phytopathology* 92:361-377.
- Hamel, C., Vujanovic, V., Nakano-Hylander, A., Jeannotte, R., and St-Arnaud, M. 2005. Factors associated with Fusarium crown and root rot of asparagus outbreaks in Quebec. *Phytopathology* 95:867-873.
- Hill, M. O., and Smith, A. J. E. 1976. Principal component analysis of taxonomic data with multi-state discrete characters. *Taxon* 25:249-255.
- Hotelling, H. 1933. Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.* 24:417-441.
- Hughes, G., and Madden, L. V. 1993. Using the beta-binomial distribution to describe aggregated patterns of disease incidence. *Phytopathology* 83:759-763.
- Hughes, G., and Madden, L. V. 1995. Some methods allowing for aggregated patterns of disease incidence in the analysis of data from designed experiments. *Plant Pathol.* 44:927-943.
- Jarausch, W., Eyquard, J. P., Lansac, M., Mohs, M., and Dosba, F. 2000. Susceptibility and tolerance of new French *Prunus domestica* cultivars to European stone fruit yellows phytoplasmas. *J. Phytopathol.* 148:489-493.
- Jarausch, W., Eyquard, J. P., Mazy, K., Lansac, M., and Dosba, F. 1999. High level of resistance of sweet cherry (*Prunus avium* L.) towards European stone fruit yellows phytoplasmas. *Adv. Hortic. Sci.* 13:108-112.
- Kison, H., and Seemüller, E. 2001. Differences in strain virulence of the European stone fruit yellows phytoplasma and susceptibility of stone fruit trees on various rootstocks to this pathogen. *J. Phytopathol.* 149:533-541.
- Labonne, G., and Lichou, J. 2004. Data on the life cycle of *Cacopsylla pruni*, Psyllidae vector of European stone fruit yellows (ESFY) phytoplasma, in France. *Acta Hortic.* 657:465-470.
- Lemaire, J.-M., Jullian, J.-P., Audergon, J.-M., and Castelain, C. 1998. Enroulement chlorotique de l'abricotier: Symptomatologie et gamme d'hôtes. *Arboricult. Fruit.* 520:21-24.
- Lindsey, J. K. 1999. On the use of corrections for overdispersion. *Appl. Stat.* 48:553-561.
- Lorenz, K.-H., Dosba, F., Poggi-Pollini, C., Llácer, G., and Seemüller, E. 1994. Phytoplasma diseases of *Prunus* species in Europe are caused by genetically similar organisms. *Z. Pflanzenkrankh. Pflanzenschutz* 101:567-575.
- Manly, B. F. J. 1991. *Randomization and Monte Carlo Methods in Biology*. Chapman and Hall, London.
- Matheron, G. 1965. *Les Variables Régionalisées et leur Estimation*. Masson, Paris.
- McCullagh, P., and Nelder, J. 1989. *Generalized Linear Models*. Chapman and Hall, London.
- Mila, A. L., Carriquiry, A. L., and Yang, X. B. 2004. Logistic regression modeling of prevalence of soybean Sclerotinia stem rot in the north-

- central region of the United States. *Phytopathology* 94:102-110.
37. Morvan, G. 1977. Apricot chlorotic leaf roll. *EPPO Bull.* 7:37-55.
 38. Nicot, P. C., and Rouse, D. I. 1987. Relationship between soil inoculum density of *Verticillium dahliae* and systemic colonization of potato stems in commercial fields over time. *Phytopathology* 77:1346-1355.
 39. R Development Core Team. 2004. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna.
 40. Ribeiro, P. J. J., Christensen, O. F., and Diggle, P. J. 2003. *geoR* and *geoRglm*: Software for model-based geostatistics. Third International Workshop on Distributed Statistical Computing, Vienna.
 41. Rosso, P. H., and Hansen, E. M. 2003. Predicting Swiss needle cast disease distribution and severity in young Douglas-fir plantations in coastal Oregon. *Phytopathology* 93:790-798.
 42. Seemüller, E., and Schneider, B. 2004. 'Candidatus Phytoplasma mali', 'Candidatus Phytoplasma pyri' and 'Candidatus phytoplasma prunorum', the causal agents of apple proliferation, pear decline and European stone fruit yellows, respectively. *Int. J. Syst. Bacteriol.* 54:1217-1226.
 43. Shtienberg, D. 1996. Variables associated with intensity of *Alternaria* leaf spot in Pima cotton. *Phytopathology* 86:123-128.
 44. Thébaud, G., Labonne, G., Castelain, C., and Chadœuf, J. 2004. Spatio-temporal analysis of disease spread provides insights into the epidemiology of European stone fruit yellows. *Acta Hort.* 657:471-476.
 45. Thébaud, G., Peyrard, N., Dallot, S., Calonnec, A., and Labonne, G. 2005. Investigating disease spread between two assessment dates with permutation tests on a lattice. *Phytopathology* 95:1453-1461.
 46. Thioulouse, J., Chessel, D., Dolédec, S., and Olivier, J. M. 1997. ADE-4: A multivariate analysis and graphical display software. *Stat. Comput.* 7:75-83.
 47. Welham, S. J., Turner, J. A., Gladders, P., Fitt, B. D. L., Evans, N., and Baierl, A. 2004. Predicting light leaf spot (*Pyrenopeziza brassicae*) risk on winter oilseed rape (*Brassica napus*) in England and Wales, using survey, weather and crop information. *Plant Pathol.* 53:713-724.
 48. Williams, D. A. 1982. Extra-binomial variation in logistic linear models. *Appl. Stat.* 31:144-148.