Comparison of logistic regression and growth function models for the analysis of the incidence of virus infection

M. Bergua, M. Luis-Arteaga and F. Escriu*

Centro de Investigación y Tecnología Agroalimentaria de Aragón.
Avda. de Montañana, 930. 50059 Zaragoza. Spain

Abstract

A logistic regression model was compared to logistic, Gompertz and log-logistic growth functions for analyzing a set of data describing the incidence of *Alfalfa mosaic virus* infection in lucerne fields aged from one to five years, and located in three different ecological areas of the Ebro Valley, Northeast Spain. Models were fitted in the form of generalized linear models, and none of them explained well the high variability of the field data, although they were useful to analyze the interdependence among epidemiological factors associated with estimated parameters in the models. The logistic regression model proved more sensitive than classical growth function models to detect significant differences in parameters such as the rate of incidence increase with age of lucerne field or the initial amount of disease, and to detect differences associated to explanatory variables such as the ecological area. Results indicate that logistic regression may be a method well suited to statistical analyses in plant epidemiology.

Additional key words: epidemiology, Gompertz, logistic, log-logistic, logistic regression, model.

Resumen

Comparación de modelos de regresión logística y modelos de función de crecimiento para el análisis de la incidencia de infección por virus

Se ha comparado un modelo de regresión logística con modelos de función de crecimiento logístico, de Gompertz y log-logístico para el análisis de un conjunto de datos que describen la incidencia de infección del virus del mosaico de la alfalfa en campos de alfalfa con edades comprendidas entre uno y cinco años, localizados en tres áreas ecológicas distintas del Valle del Ebro en España. Ninguno de los modelos, que se ajustaron como modelos lineales generalizados, explicaron suficientemente la alta variabilidad existente en los datos de campo, pero fueron útiles para analizar la interdependencia entre determinados factores epidemiológicos asociados con los parámetros estimados por los modelos. El modelo de regresión logística proporcionó mayor sensibilidad que los modelos clásicos de crecimiento para detectar diferencias significativas en los parámetros, como el índice de aumento de la incidencia con la edad del alfalfar o la cantidad inicial de infección, y diferencias asociadas con las variables explicativas del modelo, como el área ecológica. Los resultados indican que la regresión logística puede ser un método adecuado de análisis estadístico para estudios de epidemiología vegetal.

Palabras clave adicionales: epidemiología, Gompertz, log-logístico, logístico, modelo, regresión logística.

Introduction[1]

Knowledge of epidemiological features of plant diseases provides useful information for understanding the biology of their causal agents, and is the basis for the establishment, planning and monitoring of effective disease management strategies (Jeger, 2004). The first and more common approach to the epidemiological study of an epidemic is the analysis of its temporal dynamics through the description and interpretation of the disease progress curve (DPC). DPC measures the change with time in the amount of disease in the

[1] Abbreviations used: AMV (*Alfalfa mosaic virus*), df (degrees of freedom), DPC (disease progress curve), MSE (mean squares of error), STE (standard error).

population of host plants, and may be considered as the epidemic «signature» in the sense that it integrates all host, pathogen and environmental factors occurring during the epidemic (Campbell and Madden, 1990), which determine the final amount of disease. Different mathematical models have been used for the analysis and prediction of DPCs, including those derived from population growth functions (Campbell and Madden, 1990) or theoretical ecology and epidemiology (Segarra *et al.*, 2001; Gilligan, 2002), and those based on computer simulations (Lannou *et al.,* 1994; Bertschinger, 1997). Comparative analyses of epidemics by model fitting may help to identify what factors are primary determinants of model parameters describing the epidemics, such as the rate of disease change or the initial and final amount of disease. Among the simplest growth models, the most widely used in plant epidemiology are the logistic, the Gompertz and the log-logistic. These models are usually fitted to data by least squares regression of their linearized form. Logistic regression represents an alternative fitting method for the logistic transformation, in which linear regression is performed by the maximum-likelihood method. In fact, least-squares regression might be considered as a particular case of maximum-likelihood regression when the dependent variable follows a normal distribution (Sokal and Rohlf, 1995). Thus, logistic regression should be a more general and better approach for fitting bivariate, binomially distributed variables, such as count data of disease infection. Whereas logistic regression has been extensively used in case-control analysis in clinical epidemiology (Petrie and Sabin, 2000; Agresti, 2002), it is infrequent in plant epidemiology studies (Mila *et al.*, 2003; Musaka *et al.*, 2003; Weiland *et al.*, 2003; Thebaud *et al.*, 2006; Harikrishnan and Del Río, 2007).

In this paper, we discuss the results of fitting logistic, Gompertz and log-logistic models, or a logistic regression model, to data sets describing *Alfalfa mosaic virus* (AMV) infection in one to five years-aged lucerne fields in the Ebro Valley, Northeast Spain. AMV (*Alfamovirus*, *Bromoviridae*) is worldwide distributed in lucerne crops, reaching considerable incidence (Gibbs, 1962; Forster *et al.*, 1985; Hajimorad and Francki, 1988). AMV is transmitted by more than 15 aphid species, and it is also seed-transmitted in lucerne, with seed transmission rates varying from 1 to 50% depending on virus strain, cultivar and age of the infected plant (Frosheiser, 1974; Pathipanawat *et al.*, 1997). This high transmissibility, together with its broad host range, with more than 400 host species including many

vegetable crops, are probably the main reasons for its wide geographical spreading and high incidence. Fitting logistic, Gompertz and log-logistic growth models to AMV incidence data by least squares regression provides similar conclusions regarding the comparison of the rate of incidence increase among epidemics. However, logistic regression seems a more sensitive method for identifying differences in DPC-description parameters.

Material and Methods

Origin of data

Data used in this work are incidence of AMV in 26 different fields of lucerne of ecotype 'Aragon', aged from one to five years, sampled in three ecological areas (named A, B and C) in the Ebro Valley (Northeast Spain) during the summer of 2005 and 2006 (Table 1). Areas A and B were located near Sariñena (Huesca), and area C in Tauste (Zaragoza). In each area, one or two fields of each age were sampled in a random systematic manner taking about 100 samples per field and analyzing them for the presence of AMV infection by ELISA with antisera raised against AMV (Loewe).

Statistical models

To analyze incidence of AMV in lucerne fields, logistic, Gompertz and log-logistic models were fitted to incidence data by least-squares regression. The logistic model was also fitted by maximum-likelihood regression. Models were fitted as generalized linear models (Agresti, 2002) and took the following mathematical expression: $T_M = k + a(Z) + b \cdot E + c(Z) \cdot E$. The dependent variable in the models T_M derived from data (Y, expressed as frequency) of AMV incidence (Table 1) after their transformation according to each model: $T_M = Ln \ [Y/(1-Y)]$ for logistic and log-logistic least-squares and logistic regression (M = L, LL, LR, respectively), and $T_{M=G} = -Ln \ [-Ln \ (Y)]$ for the Gompertz least-squares model. The dependent variable was fitted as a function of two explanatory variables, Z and E, with Z explaining the effect on AMV incidence of the ecological area ($Z = A$, B, C, see Table 1), and E explaining the effect on AMV incidence of the age of the lucerne field ($E = Age$ for all the models except log-logistic, in which $E = Ln \ (Age)$, with $Age = 1$ to 5 years,

Table 1. Incidence[1] of *Alfalfa mosaic virus* (AMV) in commercial lucerne fields aged from one to five years and located in different ecological areas of the Ebro Valley (Northeast Spain)

Ecological area	Age of lucerne field (years)				
	1	2	3	4	5
A	0.0	28.0	27.0	19.0	—
	—	18.0	75.7	47.0	52.0
B	0.0	9.0	40.4	66.6	—
	4.0	34.3	41.0	—	59.5
C	25.0	67.0	83.0	96.0	98.0
	0.0	10.0	96.0	23.0	79.0

[1] Data are percentages of AMV infection in 26 different fields of lucerne, one to five years old, located in three ecological areas of the Ebro Valley (see Materials and Methods). In each area, one or two fields of each age were sampled during the summer of 2005 (first row of data) and 2006 (second row). A dash indicates the lack of data, i.e. only one field of the corresponding age was sampled.

see Table 1). The term $k + a(Z)$, which is named Y_0, represents an intercept that does not depend on age effect, i.e., it quantifies the initial amount of disease of an hypothetical DPC, with k constant and $a(Z)$ accounting for differences in ecological area. The term $b + c(Z)$, which is named r_M (M = L, G, LL or LR), represents the rate of increase of AMV incidence with age, with b constant and $c(Z)$ accounting for differences in ecological area. The term $c(Z) \cdot E$ represents the interaction between the explanatory variables Z and E. Thus, models also take the form $T_M = Y_0 + r_M \cdot E$, which is their more generally known linearized form.

Statistical parameters of goodness of fit for least-squares regression models, i.e. the coefficient of determination (R^2) and the mean squares of error (MSE), were recalculated to give R^{2*} and MSE^* after back-transformation of the predicted T_M and the predicted mean of its distribution, in order to allow goodness of fit comparisons among models. Similarly, the initial amount of disease Y_0^* for the hypothetical DPC was calculated by back-transformation from the intercept parameter Y_0 in the models. The rate of increase of AMV incidence for each model (r_M) was recalculated to give the weighted mean absolute rate of incidence increase using the expression: $r_R = r_M / (2m + 2)$, with $m = 2$ for logistic, log-logistic and logistic regression models, and $m = 1$ for Gompertz model (Richards, 1959). This new parameter allowed comparisons among models.

Results

Incidence data of AMV infection of 26 commercial lucerne fields with ages ranging from one to five years

are presented in Table 1. Fields were located in three different ecological areas referred to as A, B and C. Models describing hypothetical DPCs for increase of AMV incidence with age were fitted to transformed incidence data (T_M) in the form of generalized linear models with two explanatory variables accounting for ecological area (Z) and age (E) of fields. Transformations of incidence or incidence and age variables varied depending on the model fitted in each case (see Materials and Methods).

Growth function models

Linearized forms of logistic, Gompertz and log-logistic growth models were fitted by least squares regression. Criteria for comparing models are presented in Table 2. The three models significantly explained the variation of the data in terms of explanatory variables ($P < 0.01$). Gompertz and log-logistic models yielded quite similar results, while the logistic model explained a smaller proportion of the variation, as indicated by the back-transformed coefficient of determination (R^{2*}). Based on both R^{2*} and MSE^*, the Gompertz model showed the best fit to data, slightly better than the log-logistic model. Distribution of standarized residuals was also similar for the three models (not shown). Parameter values and their standard errors were calculated in the three models for each level (A, B, C) of the explanatory variable Z, which affects the component $a(Z)$ of the intercept Y_0, and the component $c(Z)$ of the rate of increase of incidence with age, r_M (Table 3, case I). Models directly provide values of the intercept, $Y_0 = k$, and the rate of increase of incidence, $r_M = b$, for

Table 2. Summary statistics for the goodness of fit of growth function models fitted by least-squares regression to data of *Alfalfa mosaic virus* (AMV) incidence

Model	R^2 (%)[1]	R^{2*} (%)[2]	MSE*[2]	d.f.[3]	F[3]	P[3]
Logistic	58.61	49.40	0.069	5	5.66	0.0021
Gompertz	59.69	59.22	0.054	5	5.92	0.0016
Log-logistic	67.49	57.53	0.058	5	8.31	0.0002

[1] Coefficient of determination for transformed dependent variable T_M. [2] Coefficient of determination and mean-squares of error recalculated after back-transformation of the dependent variable T_M. [3] Hypothesis test for the models: d.f., degrees of freedom; F, Fisher statistic; P, probability of the significance test.

$Z = A$, i.e. a(Z) and c(Z) equal to zero. These parameters are calculated as $Y_0 = k + a(Z)$ and $r_M = b + c(Z)$ for $Z = B$ or $Z = C$. The parameters describing DPCs, i.e. the initial amount of disease (Y_0^*) and the weighted mean absolute rate of increase of incidence with age (r_R), were calculated after back-transformation of Y_0 or Richards' correction of r_M, respectively, for comparison among models (Table 3, case I, see Materials and Methods).

The advantage of using general linear models with explanatory variables Z and E is that they provide the

Table 3. Estimated parameters and their standard errors of growth function models fitted by least-squares regression to data of *Alfalfa mosaic virus* (AMV) incidence

Model parameters[2]		Model[1]: $T_M = k + a(Z) + b \cdot E + c(Z) \cdot E$					
		Logistic		Gompertz		Log-logistic	
		Coefficient[3]	STE[3]	Coefficient[3]	STE[3]	Coefficient[3]	STE[3]
I) k		−4.053*	1.628	−1.341	0.979	−4.043**	1.246
a(Z)	$Z = A$	0.0	—	0.0	—	0.0	—
	$Z = B$	−0.595	2.129	−0.441	1.280	0.047	1.581
	$Z = C$	0.455	2.077	−0.160	1.249	1.058	1.570
b		0.971	0.502	0.395	0.302	2.917*	1.125
c(Z)	$Z = A$	0.0	—	0.0	—	0.0	—
	$Z = B$	0.270	0.686	0.188	0.413	0.253	1.492
	$Z = C$	0.363	0.635	0.476	0.382	0.625	1.415
r_R	$Z = A$	0.162	—	0.099	—	0.486	—
	$Z = B$	0.207	—	0.146	—	0.528	—
	$Z = C$	0.222	—	0.218	—	0.590	—
Y_0^*	$Z = A$	0.017	—	0.022	—	0.017	—
	$Z = B$	0.009	—	0.003	—	0.018	—
	$Z = C$	0.027	—	0.011	—	0.048	—
II) r_R	$Z = A,B,C$	0.202	—	0.165	—	0.544	—
Y_0^*	$Z = A$	0.008	—	0.000	—	0.012	—
	$Z = B$	0.010	—	0.001	—	0.017	—
	$Z = C$	0.038	—	0.093	—	0.062	—

[1] T_M: transformed dependent variable fitted by least-squares regression. Transformation was different for each model fitted: $T_L = T_{LL} = Ln[Y/(1–Y)]$; $T_G = –Ln[–Ln(Y)]$, with M = L for logistic, G for Gompertz and LL for log-logistic models. Z: explanatory variable for the ecological area of lucerne field. E: explanatory variable for the age of lucerne field: $E = Age$ for logistic and Gompertz models, $E = Ln (Age)$ for log-logistic model, with $Age = 1$ to 5 years. [2] I) Interaction $Z \times E$ is present in the models, II) Interaction $Z \times E$ is dropped from the models. k, a(Z), b and c(Z) are intercept and coefficients for explanatory variables Z and E (see Materials and Methods for details). r_R: weighted mean absolute rate of increase of incidence with age after Richards' correction for comparison among models: $r_R = r_M / (2m + 2)$, with $r_M = b + c(Z)$, $m = 2$ for logistic and log-logistic models, $m = 1$ for Gompertz model. Y_0^*: initial amount of disease after back-transformation of $Y_0 = k + a(Z)$. [3] Significance t-test for coefficient estimate: significance at 95% (*) or 99% (**) confidence level and standard error (STE) are indicated.

means to test hypotheses about the effects of these variables and their interaction. Analysis of variance of explanatory variables in the three models indicated that the variable age (E) had a significant effect on AMV incidence ($P \leq 0.0004$), and that the interaction $Z \times E$ was not significant ($0.4515 \leq P \leq 0.9012$), i.e., the effect of age (E) on AMV incidence did not change for A, B, or C ecological areas under study. Thus, the interaction $Z \times E$ was dropped from the models, so that DPCs for the three areas had the same weighted mean absolute rate of increase of incidence with age (Table 3, case II). With the same rate of increase of incidence (r_R), the three areas were compared for the initial amount of disease (Y_0^*), which did not differ for the logistic model, while the Gompertz and log-logistic models detected differences between areas A and B, on one side, and area C, on the other ($P \leq 0.0189$).

Logistic regression models

Logistic regression models were step-wise fitted to the AMV incidence data (Y) by the maximum-likelihood method. Fitting started with model I, which included no explanatory variables, and proceeded by introducing explanatory variables for ecological area (Z) and/or age of lucerne field (E) at subsequent steps (models II to IV). Finally, the complete model V included both explanatory variables (E and Z) and their interaction ($Z \times E$) (Table 4). Maximum-likelihood ratio statistics (G) tests the goodness of fit of each model. The decreasing values of G indicated a better fit as new explanatory variables were added to the models. However, the G value for the complete model V was still significant, which means a lack of fit, i.e., additional explanatory variables should be included in order to better explain the variation within data. Significance of the expla-

natory variables in the models could be analyzed by computing the loss of fit resulting from dropping each variable from the model. Thus, the effect of ecological area, represented by the term a(Z), was tested by subtracting the G-goodness of fit value of model IV from that of model III, i.e. $G = 152.3844$ with two degrees of freedom (d.f.), which is significant ($P < 10^{-33}$). Similarly, the effect of age, represented by the term b · E, was tested by the difference between the G values of models II and IV, and the effect of $Z \times E$ interaction, represented by the term c(Z) · E, was tested by the difference between the G values of models IV and V, which is $G = 488.0101$ with one d.f. for age and $G = 19.4428$ with two d.f. for $Z \times E$ interaction ($P < 10^{-107}$ and $P < 10^{-5}$, respectively). Thus, the age of the lucerne field was the main factor affecting AMV incidence, the ecological area was the second more important factor, and the interaction between both effects was the third one. These results indicate that the increase of incidence with age differed depending on the ecological area.

Parameter estimates and their standard errors for model V, which includes the three effects indicated above, are presented in Table 5. The intercept and the rate of increase of incidence with age are $Y_0 = k$ and $r_{LR} = b$, respectively, for $Z = A$. Those for $Z = B$ or $Z = C$ are computed as $Y_0 = k + a(Z)$ and $r_{LR} = b + c(Z)$, respectively. Coefficient c(Z) for areas B and C was significant, which indicates that the effect of age in these areas differed from that in area A. The effect of age did not differ between areas B and C ($P = 0.8408$, Wald), i.e., the $Z \times E$ interaction was significant because of area A. Thus, the model was reduced to describe DPCs with a common rate of increase of AMV incidence with age (r_R) for areas B and C, different from that of area A. Intercepts for areas B and C were significantly different ($P < 0.0001$) and are presented in Table 5 in terms of the initial amount of disease (Y_0^*).

Table 4. Step-wise fitting of logistic regression models by maximum-likelihood method to data of *Alfalfa mosaic virus* (AMV) incidence

Model[1]	d.f.[2]	G[2]	P[2]
I.　Logit (Y) = k	25	1,263.7955	< 0.0001
II.　Logit (Y) = k + a(Z)	23	1,103.9491	< 0.0001
III.　Logit (Y) = k + b · E	24	768.3234	< 0.0001
IV.　Logit (Y) = k + a(Z) + b · E	22	615.9390	< 0.0001
V.　Logit (Y) = k + a(Z) + b · E + c(Z) · E	20	596.4962	< 0.0001

[1] Y: frequency of AMV-infected plants: Logit (Y) = Ln[$Y/(1-Y)$]. k, a(Z), b and c(Z): intercept and coefficients for explanatory variables Z and E (see Materials and Methods for details). [2] Hypothesis testing for the models: d.f., degrees of freedom; G, maximum-likelihood ratio statistic for goodness of fit; P, probability of significance test.

Table 5. Parameter estimates and their standard errors of a logistic regression model fitted by maximum-likelihood method to data of A*lfalfa mosaic virus* (AMV) incidence

Model[(1)]: Logit $(Y) = k + a(Z) + b \cdot E + c(Z) \cdot E$			
Model parameters[(1,2)]		Coefficient[(3)]	STE[(3)]
k		–2.219**	0.229
a(Z)	Z = A	0.0	—
	Z = B	–0.987**	0.326
	Z = C	0.107	0.290
b		0.489	0.067
c(Z)	Z = A	0.0	—
	Z = B	0.372**	0.098
	Z = C	0.353**	0.089
r_R	Z = A	0.081	—
	Z = B, C	0.142	—
Y_0*	Z = A	0.098	—
	Z = B	0.040	—
	Z = C	0.106	—

[(1)] Y: frequency of AMV-infected plants: Logit $(Y) = \text{Ln}[Y/(1 - Y)]$. k, a($Z$), b and c($Z$): intercept and coefficients for explanatory variables Z and E (see Materials and Methods for details). [(2)] r_R: weighted mean absolute rate of increase of incidence with age after Richards' correction: $r_R = r_{LR} / (2m + 2)$, with $m = 2$. Y_0*: initial amount of disease after back-transformation of $Y_0 = k + a(Z)$. [(3)] Significance Wald test for coefficient estimates (** significance at 99% confidence level); STE, standard error of coefficient estimates.

Discussion

Generalized linear models allow modelling of a random categorical or continuous dependent variable in terms of categorical or continuous explanatory variables. This is usually performed by maximum-likelihood regression, which maximizes the likelihood for the distribution of the dependent variable in the fitting process, so that this distribution is not necessarily restricted to normality, contrary to least-squares regression (Agresti, 2002). This is the case of logistic regression, a method specially suited for bivariate, binomial distributions. This type of statistical distributions are frequent in human, animal or plant epidemiology, in which response dependent variables such as presence or absence of disease need to be explained by categorical variables such as breed of an animal or strain of a pathogen, and/or continuous variables such as age of a patient or time from infection. Logistic regression has been extensively used in medicine, for example in clinical studies (Petrie and Sabin, 2000). However, in spite of its potential for many types of studies, its use is not generalized in plant epidemiology (Mila *et al.*, 2003; Musaka *et al.*, 2003; Weiland *et al.*, 2003; Thebaud *et al.*, 2006; Harikrishnan and Del Río, 2007). We have compared logistic regression with three functions frequently used in plant epidemiology: the logistic, Gompertz and log-logistic growth functions (Campbell and Madden, 1990; Jeger, 2004). The comparison was done for the analysis of data on AMV incidence in lucerne fields from different ecological areas and with different ages. The three growth models, fitted by least-squares regression, showed a poor fit to the data, as indicated by their back-transformed coefficient of determination (R^{2*}), which explained about 50-60% of the data variability according to the model. The maximum-likelihood fitted logistic regression model V, did neither show a good fitting to the data. The poor fit of all models is probably due to the high variability contained within data, in which lucerne fields of the same ecological area and age presented quite different AMV incidences. The purpose of the temporal analysis of a particular data set is what determines the degree of precision needed for such an analysis (Campbell and Madden, 1990). The goodness of fit obtained for our data is probably not enough accurate for predictive purposes or for a highly precise description of the DPC. However, it is sufficiently good to represent the observed field variability and to analyze the level of interdependence of certain epidemiological factors in AMV epidemics. In fact, models that differed in goodness of fit, as the logistic and Gompertz models, yielded similar conclusions regarding comparisons of rates of increase of AMV incidence with age of lucerne fields among different ecological areas. Other comparisons, however, would require a better fitting of models to the data for the detection of significant differences. For example, logistic, Gompertz and log-logistic models were fitted assuming the same rate of increase of incidence for all the ecological areas, as no significant differences were found for rates of increase. Thus, the initial amount of disease in the different areas could be compared, and only the best fitted models (Gompertz and log-logistic) detected significant differences between areas A and B, on one side, and area C. Logistic regression also provided a useful methodology for this type of comparisons. Model V in Table 4, which showed a clear lack of fit, was sensitive enough for detecting significant differences in the rate of increase of incidence with age between ecological area A and areas B and C, contrary to logistic, Gompertz or log-logistic growth

models. This may be attributed to the only difference between the logistic-regression model V (Table 4) and the logistic growth model (Table 3): the statistical method for fitting the distribution of the dependent variable, maximum-likelihood regression for model V and least-squares regression for the logistic growth model. As was the case for the Gompertz and log-logistic models, logistic-regression model V detected significant differences in the initial amount of disease between ecological areas B and C, once it was assumed that the rate of increase of incidence was the same for these areas.

In summary, four different generalized linear regression models were fitted to field data of AMV incidence in lucerne fields located in different ecological areas and with different ages, resulting in roughly comparable poor fits. Logistic regression was useful for the analysis of epidemiological factors associated to estimated parameters in the model, and provided more sensitive analyses than traditional least-squares regression of growth functions. Thus, logistic regression should be considered a good method with application to a wide variety of statistical analyses in the field of plant epidemiology.

Acknowledgements

References

AGRESTI A., 2002. Categorical data analysis. John Wiley & Sons Inc, New York, USA. 710 pp.

BERTSCHINGER L., 1997. Modelling plant virus disease epidemics: development and use of simulation models. Int Potato Cent, Lima, Peru. 32 pp.

CAMPBELL C.L., MADDEN L.V., 1990. Introduction to plant disease epidemiology. John Wiley & Sons Inc, New York, USA. 532 pp.

FORSTER R.L.S., MORRIS-KRSINICH B.A.M., MUSGRAVE D.R., 1985. Incidence of alfalfa mosaic virus, lucerne Australian latent virus, and lucerne transient streak virus in lucerne crops in the North Island of New Zealand. New Zeal J Agr Res 28, 279-282.

FROSHEISER F.I., 1974. Alfalfa mosaic virus transmission to seed through alfalfa gametes and longevity in alfalfa seed. Phytopathology 64, 102-105.

GIBBS A.J., 1962. Lucerne mosaic virus in British lucerne crops. Plant Pathol 11, 167-171.

GILLIGAN C.A., 2002. An epidemiological framework for disease management. Adv Bot Res 38, 1-64.

HAJIMORAD M.R., FRANCKI R.I.B., 1988. Alfalfa mosaic virus isolates from lucerne in South Australia: biological variability and antigenic similarity. Ann Appl Biol 113, 45-54.

HARIKRISHNAN R., DEL RÍO L.E., 2007. A logistic-regression model for predicting white mold incidence of dry bean in North Dakota. Phytopathology 97, S161-S161.

JEGER M.J., 2004. Analysis of disease progress as a basis for evaluating disease management practices. Annu Rev Phytopathol 42, 61-82.

LANNOU C., DEVALLAVIELLEPOPE C., GOYEAU H., 1994. Host mixture efficacy in disease-control-effects of lesion growth analysed through computer-simulated epidemics. Plant Pathol 43, 651-62.

MILA A.L., CARRIQUIRY A.L., ZHAO J., YANG X.B., 2003. Impact of management practices on prevalence of soybean *Sclerotinia* stem rot in the north-central United States and on farmers' decisions under uncertainty. Plant Dis 87, 1048-1058.

MUKASA S.B., RUBAIHAYO P.R., VALKONEN J.P.T., 2003. Incidence of viruses and viruslike diseases of sweetpotato in Uganda. Plant Dis 87, 329-335.

PATHIPANAWAT W., JONES R.A.C., SIVASITHAMPARAM K., 1997. Factors influencing transmission of alfalfa mosaic virus through seed of annual medics (*Medicago* spp.) and the genetic control of seed transmission rate. Aust J Agr Res 48, 989-997.

PETRIE A., SABIN C., 2000. Medical statistics at a glance. Blackwell Science Ltd, Oxford, UK. 139 pp.

RICHARDS F.J., 1959. A flexible growth function for empirical use. J Exp Bot 10, 290-300.

SEGARRA J., JEGER M.J., VAN DEN BOSCH F., 2001. Epidemic dynamics and patterns of plant diseases. Phytopathology 91, 1001-1010.

SOKAL R.R., ROHLF F.J., 1995. Biometry. The principles and practice of statistics in biological research. W.H. Freeman and Co, New York, USA. 887 pp.

THEBAUD G., SAUVION N., CHADOEUF J., DUFILS A., LABONNE G., 2006. Identifying risk factors for European stone fruit yellows from a survey. Phytopathology 96, 890-899.

WEILAND J.E., STANOSZ J.C., STANOSZ G.R., 2003. Prediction of long-term canker disease damage from the responses of juvenile poplar clones to inoculation with *Septoria musiva*. Plant Dis 87, 1507-1514.