Natural Sciences

Original article

Assessment of the level of adjustment of three epidemiological models in the analysis of epidemics with incidences less than 100% such as the lethal wilt of oil palm (*Elaeis guineensis* Jacq.)

Evaluación del nivel de ajuste de tres modelos epidemiológicos en el análisis de epidemias con incidencias inferiores al 100% como la marchitez letal de la palma de aceite (*Elaeis guineensis* Jacq.)

 Juan Manuel López-Vásquez[1],*, **Jairo Castaño-Zapata[2]**

[1] Programa de Plagas y Enfermedades, Centro de Investigaciones en Palma de Aceite - Cenipalma, Santa Marta (Magdalena), Colombia
[2] Facultad de Ciencias Agropecuarias, Universidad de Caldas, Manizales, Colombia

Abstract

The production of oil palm is a major agricultural activity in Colombia. Lethal wilt (LW) of the oil palm is one of the most devastating diseases in the Eastern zone of the country. Several epidemiological models used in epidemic analyses assume that there is a constant area where the host will become diseased at the end of the epidemic (maximum incidence of disease = 100%). Based on the analysis of three different epidemics, we demonstrated the error in the application of the model that best fits the observed data when the maximum incidence of the disease (*Kmax*) is below the assumed. We assessed the fit of the monomolecular, logistic, and Gompertz models at different final incidence values of the disease including the maximum observed ($y_1 + 0.1$). We analyzed the data with linear regression and residuals variance and distribution. We measured the relative quality level of fit of the model for each *Kmax* by determining coefficients (R^2) and the Akaike and Bayesian information criteria (AIC & BIC). The monomolecular model showed a tendency to increase the level of adjustment when *Kmax* assumed values were close to 1 while the logistic and Gompertz models remained stable regardless of the evaluated *Kmax* values. The consequences of assuming a *Kmax* with values equal to 1 reflected not only the erroneous estimation of parameters such as y_0 and *r* but also gave rise to a misinterpretation of the temporal behavior of the epidemic.

Keywords: Lethal wilt; *Elaeis guineensis*; Epidemiology; Monomolecular; Logistic; Gompertz.

Resumen

La producción de palma de aceite es muy importante en Colombia. La marchitez letal de la palma de aceite es una enfermedad muy limitante en la zona oriental colombiana. Algunos modelos empleados en el análisis de estas epidemias se basan en el supuesto de que existe un área constante donde el hospedante llega a enfermarse al final de ellas (incidencia máxima de la enfermedad = 100 %). El análisis de tres epidemias diferentes demostró el error que se comete en la aplicación del modelo que mejor se ajusta a los datos observados, con una máxima incidencia de la enfermedad (*Kmax*) por debajo de la asumida. El ajuste de los modelos monomolecular, logístico y Gompertz se hizo con diferentes incidencias, incluida la máxima observada ($y_1 + 0,1$). Los datos se sometieron a análisis de regresión lineal y de varianza y distribución de los residuales. La calidad del ajuste se midió mediante los coeficientes de determinación (R^2), el análisis de regresión lineal y los criterios de información Akaike (CIA) y bayesiano (CIB). Se evidenció la sensibilidad del modelo monomolecular cuando la incidencia máxima de la enfermedad se asumió por encima del valor real, contrario a lo observado

en los modelos logístico y Gompertz, cuyos valores se mantuvieron estables independientemente del aumento de la *Kmax* evaluada. Las consecuencias de asumir una *Kmax* con valores igual a 1 no solamente se reflejan en la estimación errónea de parámetros como el y_0 y el *r*, sino que también conduce a una mala interpretación del comportamiento de la epidemia.

Palabras clave: Marchitez letal; *Elaeis guineensis*; Epidemiología; Modelo monomolecular; Modelo logístico; Gompertz.

Introduction

The current lack of understanding of the causal agent of LW leads to uncertainty about the influence of various factors on the dynamics of the disease, which makes it difficult to develop a management plan. In this context, it has been necessary to largely resort to plantations history and basic management practices such as the reduction of the inoculum starting from the early elimination of diseased palms, which implies a strict follow-up, to the appearance of cases through weekly phytosanitary censuses (**Cenipalma,** 2019). Thus, those who have to deal with the disease have extensive records of historical data. However, there are few methodological tools that allow converting these data into information and, from there, into knowledge about the behavior of the disease on a time scale.

An epidemic in a crop can be regarded as a change in the incidence of the disease in a host population over time and space. The characterization of the progression of those changes over time can provide valuable epidemiological insights. This characterization is important to understand how diseases develop in plant populations and how management measures affect epidemics (**Campbell & Madden,** 1990; **Van der Plank,** 1963). The graphical representation of the incidence of a disease (*y*) versus time (*t*) is known as disease progression or development curve. For many purposes, this is the main description of an epidemic and focuses on the interactions occurring among the host, the pathogen, and the biological and physical environment in the development of the disease whose main objective is to understand the complex interactions occurring. This information can be used, for example, to predict the incidence of the disease at a given time, quantify the effects of management strategies on epidemics, or, ultimately, to develop a theoretical basis for determining whether an epidemic might occur, and, if it does, to identify the factors affecting both the disease development rate (*r*) and its final incidence (y_1) (**Madden,** *et al.***,** 2007; **Xu,** 2006).

In principle, plant disease epidemics can be classified into two basic types, monocyclic and polycyclic, depending on the source of the inoculum that comes into contact with the host over the course of the disease and the number of infection cycles per crop cycle (**Madden,** *et al.***,** 2007). Therefore, the classification of an epidemic as monocyclic or polycyclic depends on the type of behavior over time as measured by the level of adjustment of the different growth models to the values observed in the field. Thus, a monocyclic epidemic can be described quite well using a monomolecular model while a polycyclic epidemic can be described with a logistic or Gompertz model (**Arneson,** 2001). Epidemiological processes are characterized by using growth models with varying degrees of complexity. The selection of an appropriate model allows the characterization of an entire epidemic with few parameters. It is, therefore, relatively easy to compare epidemics and assess the effects of biological and environmental factors, as well as the possible managements for implementing more effective, efficient, and sustainable strategies in integrated disease management (**Forrest,** 2007; **Madden,** 1986). The resulting disease development curves are graphical representations of the development dynamics of the epidemics and their parameters can be fitted to the obtained data using any standard statistical package (**Van Maanen & Xu,** 2003; **Neher,** *et al.,* 1997). In this line, multiple fully functional applications have been developed in different programming languages such as Java, with the EpiModel software (**Nutter & Parker,** 1997) and Excel spreadsheets (**Bowen,** 2015), and R packages such as epiphy (**Gigot,** 2018) and epifitter (**Alves & Del Ponte,** 2021).

Most of the mathematical models used for determining disease progression curves can be estimated using only two parameters such as y_0 (initial disease or constant of integration) and r (rate of disease development). Using these two, two or more epidemics can be compared and a considerable understanding of disease dynamics is achieved. However, not all epidemics can be described by two-parameter models (**Madden, *et al.***, 2007). The values observed over time in the development of a disease do not often fit any specific model, which most probably reflects the fact that the biological basis of the models is only a simple approximation to a much more complex pathosystem and, consequently, the model lacks the necessary flexibility to be adapted to various aspects of the development of the disease (**Park & Lim,** 1985). The classical growth-curve models for epidemics analysis include several implicit assumptions, which may lead to misinterpretations about the progression of the disease if they are not considered (**Neher & Campbell,** 1992). Many of those models, such as the monomolecular, logistic, and Gompertz, assume a constant host area that becomes diseased by the end of the epidemic implying a maximum disease incidence of 100%. Consequently, the models used to describe epidemics assume an asymptote or a maximum disease incidence (*Kmax*) equal to 1 (**Campbell & Madden,** 1990; **Seem,** 1988). However, this assumption is not valid for many diseases and the mathematical description of the epidemic is thus inaccurate, imprecise, and inappropriate. Besides, in some epidemics, such as rusts, powdery mildews, and viral diseases caused by obligate parasites, it is only possible to visualize a maximum disease intensity between 25 and 40%, far below the expected 100% (**Jeger,** 1982; **Campbell & Madden,** 1990; **Gilligan,** 1990; **Kranz,** 2003; **Madden, *et al.*,** 2007). This has resulted in the ongoing use of growth models for epidemiological analysis that do not consider the maximum disease intensity. Consequently, important parameters, such as the growth or disease development rate, are not properly estimated, or, simply, a given disease is analyzed in a model with different behavior.

Several researchers have reported this type of error in the description of epidemics. For example, **Turner, *et al.*** (1969), **Kiyosawa** (1972), and **Analytis** (1973, 1979) agreed that the value of the asymptote or the *Kmax* could affect the calculation of r in generalized Logistic models in the description of epidemics. **Kushalappa & Ludwing** (1982) indicate that empirically determining the value of the asymptote (*Kmax*) greatly improves the goodness of fit of disease development models. **Park & Lim** (1985) mathematically illustrate how disease development rates (r) are underestimated when calculated with traditional models under the assumption that the maximum disease incidence is 1 (*Kmax = 1*) when really the maximum incidence is below this value (*Kmax <1*). **Neher & Campbell** (1992) quantified the magnitude of underestimation in r when values above a real incidence are assumed in models such as the monomolecular, logistic, and Gompertz. Since *Kmax* affects rates of disease development, **Neher & Campbell** (1992) suggested for *Kmax* <1 shape parameters m with the values 0, 1, and 2 for the monomolecular, Gompertz and logistic functions, respectively, to use weighted mean absolute rates $\rho = r\,K/(2m + 2)$ with $K = Kmax$ and the rate disease progress r calculated by the function used to fit the curve. This approach introduces the shape parameter m as a measurement term and descriptor. Parameter K can be dropped from the equation (if different with the same m) and then rK could be calculated as an overall (mean) measure of the absolute rate of disease development (**Neher & Campbell,** 1992). Although this approach manages to correct the underestimation of parameters such as the rate (r) of development of the disease, it fails to improve the classification of the model that best fits, which still leads to generating erroneous or biased interpretations of pathosystems with monocyclic or polycyclic behaviors completely different from the real one. The objective of the following study is to demonstrate the error committed in the classification of the model that best fits the data when *Kmax* is lower based on the temporal analysis of three epidemics caused by lethal wilt (LW) in oil palm (*Elaeis guineensis* Jacq.) plantations.

Methodology

Plot selection

We obtained the historical disease monitoring records for three plots sowed with susceptible cultivars of *E. guineensis* between January 2015 and December 2017 (36 months) at the *Luker Agrícola* (4°35'N, 72°49'W, 198 m. a.s.l.) and *Palmar de Oriente* plantations (4°29'N, 72°50'W, 195 m. a.s.l.) located in the municipality of Villanueva (Casanare), Colombia, an area with a high incidence of the disease. The three plots we selected satisfied the following characteristics (listed from highest to lowest importance): (*i*) final incidence with the highest record (y_1); (*ii*) known initial incidence (y_0), and (*iii*) records with the longest follow-up time (*t*). All historical disease-monitoring records were done palm by palm by personnel trained in the detection of plants with initial symptoms of the disease at 30-day intervals. The total population of plants evaluated from plots 1, 2, and 3 corresponded to 1,315 (8.5 ha), 2,842 (17.8 ha), and 2,835 (17.7 ha), respectively.

Models used when Kmax is a parameter

Given that the monomolecular, logistic, and Gompertz models generally assume that the maximum disease incidence is 1, we modified the mathematical equations of the traditional models in terms of proportion to the differential, nonlinear, and linearized versions in which the maximum disease incidence was a parameter (**Table 1**).

Assessment of goodness of fit of the models

We assessed the goodness of fit of the monomolecular, logistic, and Gompertz models at different final incidences including the maximum observed ($y_1 + 0.1$), which were 0.43, 0.28, and 0.25, respectively (**Table 2**). Then, we subjected the data to linear regression analysis, analysis of variance, and distribution of residuals using *R* v1.1.463 (**R Core Team,** 2018) and the statistical package '*nlme*' v3.1-144 (**Pinheiro, *et al.,*** 2004). We calculated the F-test on the estimated parameters as the ratio between each of them and the asymptotic standard error to determine whether the parameters contributed to fit the model. Higher F-values are associated with higher contributions of the estimated parameters to fit the model (**Hosmer & Lemeshow,** 2000). The asymptotic standard error is used to assess the goodness of fit of the models to the real points of the epidemic using the least-squares regression associated with the coefficients estimated by each model. It is accepted that small standard errors indicate more accurate estimates (**McCullagh & Nelder,** 1992).

Comparison of the goodness of fit of the models

We used the statistical package '*AICcmodavg*' v2.2-2 (**Mazerolle,** 2019) to compare the goodness of fit of the models for each *Kmax* while the relative goodness of fit of the three models was determined from the estimated coefficients of determination (R^2) and the implementation of the Akaike (AIC) and Bayesian (BIC) information criteria.

The R^2 of the linear regression analyses determines the ability of the model to replicate the results and the proportion of variation of the results that the model can explain. A coefficient value closer to 1 means less variation in the estimate. The AIC uses a relative

Table 1. Differential, nonlinear, and linearized equations of the models used to analyze the disease progression data when the maximum disease incidence (*Kmax*) is a parameter

Model	*dy/dt*	*y =*	**Linearized form**
Monomolecular	$r_M * (K - y)$	$K\{1 - [((K - y_0)/K)\ exp(-r_M t)]\}$	$ln[K/(K - y)] = ln[K/(K - y_0)] + r_M t$
Logistic	$r_L y(1 - y/K)$	$K/[1 + exp\ (-\{ln[y_0/(K - y_0)] + r_L t\})]$	$ln[y/(K - y)] = ln\ [y_0/(K - y_0)] + r_L t$
Gompertz	$r_G y[ln(K) - ln(y)]$	$K\{exp[ln(y_0/K)\ exp(-r_G t)]\}$	$-ln[-ln(y/K)] = -ln[-ln(y_0/K)] + r_G t$

Source: Campbell, 1998

K: Kmax: maximum disease incidence (y) or asymptote of the disease progression curve; *Y:* disease in proportion at the time of observation; y_0: disease in proportion at the first observation; *r*:* disease growth rate of a sᴘecific model; *t:* considered time interval

Table 2. Parameters of maximum disease incidence (*Kmax*) assessed in three plots of the *Luker Agrícola* and *Palmar de Oriente* plantation

Plot	Maximum observed disease incidence (y_1 + 0.1)							
1	*1.00*	*0.90*	*0.80*	*0.70*	*0.60*	*0.50*	*0.45*	*0.43* *(y_1 + 0.1)*
2	*1.00*	*0.90*	*0.80*	*0.70*	*0.60*	*0.50*	*0.40*	*0.28* *(y_1 + 0.1)*
3	*1.00*	*0.90*	*0.80*	*0.70*	*0.60*	*0.50*	*0.40*	*0.25* *(y_1 + 0.1)*

y_1+0.1: maximum incidence observed per plot in proportion

value between the goodness of fit of the model and its complexity calculated by the formula $AIC = -2 \log(]L(\theta|y)) + 2K$, where $\log(L(\theta|y))$ is the numerical value of the log-likelihood function at its maximum point and K is the number of estimated parameters in the function. The BIC applies a higher penalty for the number of parameters in the model calculated by the formula $BIC = -2 \log(L(\theta|y)) + K \log n$; as in the previous formula, $\log(L(\theta|y))$ is the numerical value of the log-likelihood function at its maximum point, K is the number of estimated parameters in the function, and n is the number of parameters in the model. The farther the AIC and BIC values are from zero, the higher the goodness of fit of the model to the observed values (**Burnham & Anderson,** 2002).

Results and discussion

After 36 months of observations in the three epidemics evaluated, 15, 22, and 18 plants died on average per month, respectively, for a total of 2,001 plants, i.e., 29% of the 6,992 plants evaluated. We assessed all models based on their goodness of fit measured by linear regression tests, analysis of variance, and 95% confidence intervals. **Table 3** shows the F-test values of the least-squares method for the analysis of variance, the coefficient estimates, the asymptotic standard error, and the 95% confidence interval for the coefficients of the three models estimated at different maximum incidences of LW in the three analyzed epidemics, respectively (**Table S1, S2, and S3,** https://www.raccefyn.co/index.php/raccefyn/article/view/1571/3188).

The F-test values indicated that the estimated parameters contributed significantly to the fit of the disease progression curves in the three epidemics, regardless of the maximum disease incidence. We also found that the estimates calculated by the three models contributed significantly to the fit of the coefficients y_0 and r by confirming that the 95% confidence intervals did not contain a value equal to zero.

A more detailed analysis of the F-test and the asymptotic standard error per plot revealed variations between the models as the maximum incidence parameter was adjusted to a real value of the disease. In general, the monomolecular model showed higher F-test values and lower asymptotic standard errors when we assessed maximum asymptotes of 1.0, 0.9, 0.8, 0.7, 0.6, and 0.5 in the epidemics. However, there were evident changes as the maximum incidence parameter (*Kmax*) was adjusted to the real observed value changing from the monomolecular to the Gompertz model in the analyzed epidemics with maximum asymptotes of 0.45 and 0.43, 0.40 and 0.28, and 0.25 (**Table 3**) corresponding to plots 1, 2, and 3, respectively.

Figures 1, 2, and **3** represent the variations when the parameter *Kmax* changed from a theoretical value (*Kmax = 1*) to a real value and its effect on the estimates of the LW development rate. The R^2 and the AIC and BIC applied to assess the relative goodness of fit of the model to the data observed showed the same trend evident in the F-test, the asymptotic standard error, and the graphic representations of the goodness of fit of the models (**Figure 4**).

Table 3. Linear regression analysis of the three growth models assessed at different asymptotes or maximum disease incidence (*Kmax*) of the records corresponding to plots 1, 2 and 3

Kmax	Model	F-test values	Coefficients	Estimates	Asymptotic standard error	95% confidence interval	
						Lower limit	Upper limit
Plot 1							
1.00	Monomolecular	1431.289674**	$y_0$1	**0.002964367**	**0.00937395**	**-0.016085792**	**0.022014526**
			r_L2	**0.016714745**	**0.00044181**	**0.015816878**	**0.017612612**
	Logistic	137.0899723**	y_0	-3.071511341	0.170775253	-3.418568411	-2.724454272
			r_L	0.094241267	0.008048932	0.077883868	0.110598665
	Gompertz	287.6131306**	y_0	-1.166184266	0.053913631	-1.275749946	-1.056618586
			r_G	0.043093946	0.002541042	0.037929927	0.048257966
0.43	Monomolecular	365.7086095**	y_0	-0.57180627	0.116947313	-0.809471805	-0.334140735
			r_M	0.105407418	0.005511929	0.094205831	0.116609004
	Logistic	968.1729275**	y_0	-2.802311908	0.12473964	-3.055813357	-2.54881046
			r_L	0.182933939	0.005879194	0.170985979	0.1948819
	Gompertz	1054.969355**	y_0	**-1.484957753**	**0.089138158**	**-1.666108285**	**-1.303807221**
			r_G	**0.136457343**	**0.004201235**	**0.127919406**	**0.14499528**
Plot 2							
1.00	Monomolecular	1011.878966**	$y_0$1	**-0.053902304**	**0.006740292**	**-0.067600226**	**-0.040204383**
			r_M2	**0.010105466**	**0.000317682**	**0.009459859**	**0.010751072**
	Logistic	213.335392**	y_0	-5.280915097	0.207018695	-5.701627704	-4.86020249
			r_L	0.142512958	0.009757148	0.122684048	0.162341869
	Gompertz	706.3612045**	y_0	-1.748302626	0.036880859	-1.823253548	-1.673351703
			r_G	0.04619849	0.001738258	0.042665923	0.049731056
0.28	Monomolecular	97.77702297**	y_0	-0.75450794	0.19278383	-1.146291821	-0.362724059
			r_M	0.089846744	0.009086234	0.071381294	0.108312194
	Logistic	775.7687896**	y_0	-4.708555057	0.169305304	-5.052624831	-4.364485283
			r_L	0.222254237	0.007979651	0.206037635	0.238470838
	Gompertz	317.5515829**	y_0	**-2.192142651**	**0.16195157**	**-2.521267839**	**-1.863017462**
			r_G	**0.136020897**	**0.007633057**	**0.12050866**	**0.151533135**
Plot 3							
1.00	Monomolecular	1876.458036**	$y_0$1	**-0.035405942**	**0.004207057**	**-0.043955711**	**-0.026856173**
			r_M2	**0.008589368**	**0.000198286**	**0.008186402**	**0.008992333**
	Logistic	111.6098686**	y_0	-5.296409667	0.282329724	-5.870172699	-4.722646636
			r_L	0.140579254	0.013306687	0.113536813	0.167621696
	Gompertz	305.9106851**	y_0	-1.707792134	0.052170333	-1.813815007	-1.601769261
			r_G	0.043006516	0.002458878	0.038009475	0.048003557
0.25	Monomolecular	171.1757723**	y_0	-0.588192091	0.129213662	-0.850785846	-0.325598336
			r_M	0.079678824	0.006090063	0.067302327	0.092055321
	Logistic	385.4504086**	y_0	-4.462901455	0.228748933	-4.927775217	-3.998027693
			r_L	0.211668711	0.010781332	0.189758408	0.233579013
	Gompertz	700.4618084**	y_0	**-1.968300076**	**0.099877178**	**-2.171274923**	**-1.765325229**
			r_G	**0.124586741**	**0.004707384**	**0.115020187**	**0.134153296**

Values in bold indicate coefficients with the highest contribution to the model fit according to the standard asymptotic error test.

** Level of significance of 1%; [1]y_0: disease in proportion in the first observation; [2]r*: disease growth rate of a specific model
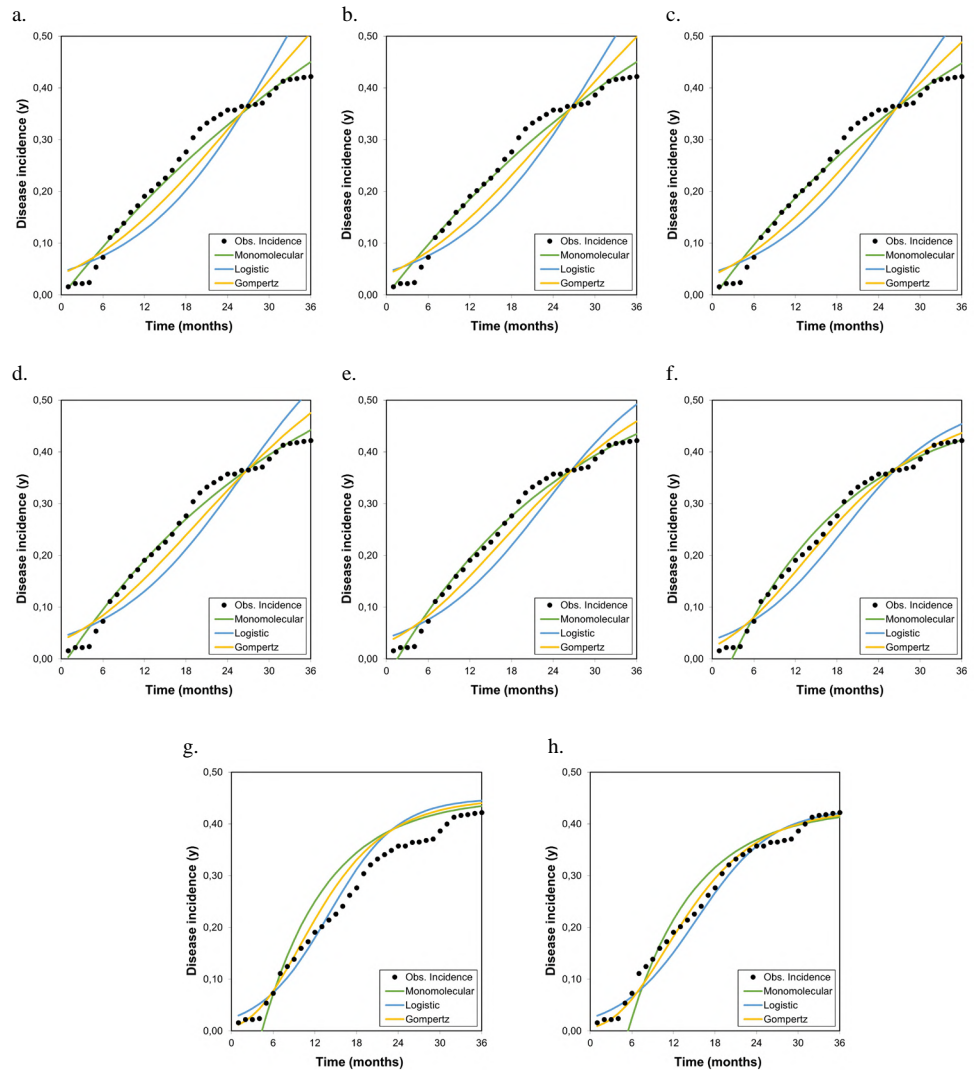
Figure 1. Graphical representation of the goodness of fit of the models to the disease progression curves at different asymptotes or maximum disease incidences (*Kmax*) corresponding to the records collected from plot 1
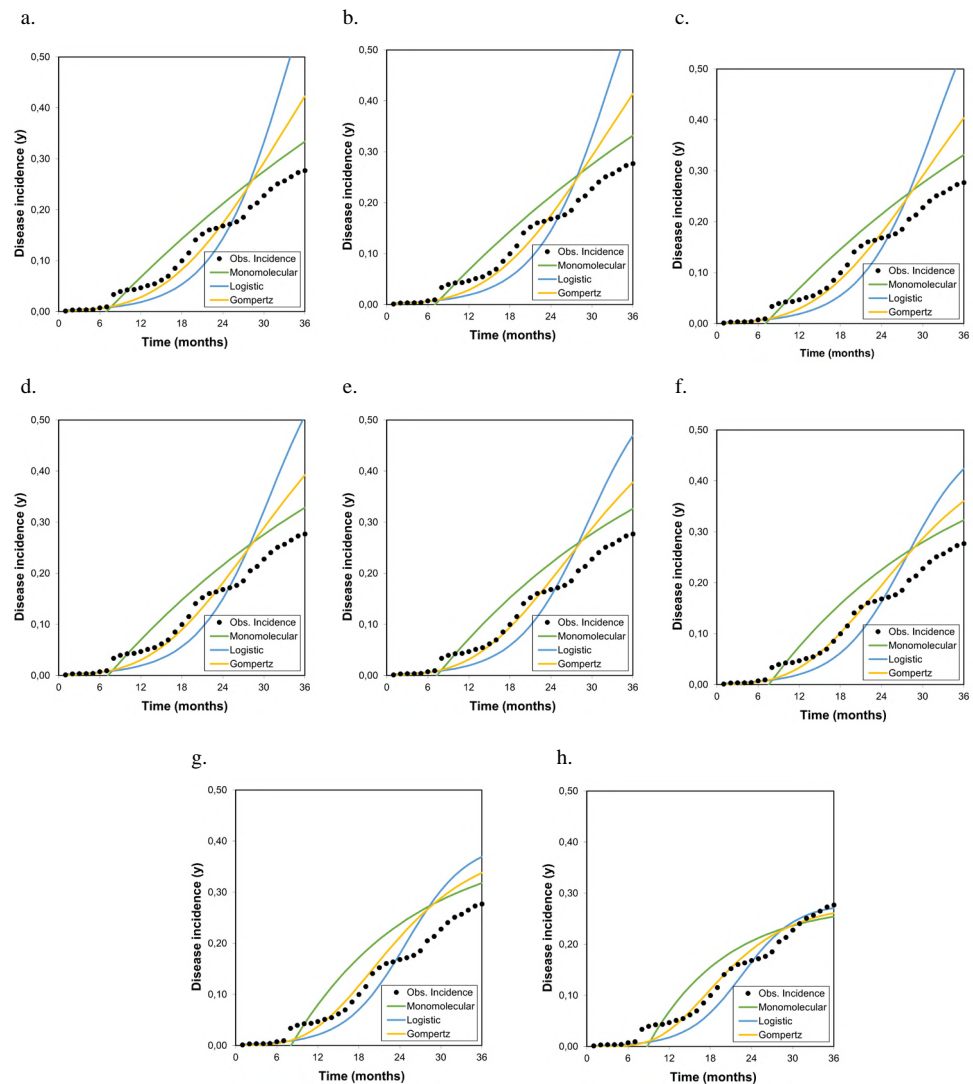
r_M, r_L, and r_G represent the growth or disease development rates in the monomolecular, logistic, and Gompertz models, respectively.

As the *Kmax* parameter approached the theoretical maximum incidence (*Kmax = 1*), the model that best fitted the observed values in the three analyzed epidemics was the monomolecular. However, its goodness of fit in each of the tests did not faithfully represent the behavior of the disease when the actual maximum incidence was much lower than 1. When the *Kmax* was assumed as the real one, i.e., 0.43, 0.28, and 0.25 for epidemics 1, 2, and 3, respectively, the goodness of fit measured by the R^2 from the linear regression analyses and the AIC and BIC showed that the model that best fitted the observed values was the Gompertz.

These results demonstrate that the monomolecular model has high sensitivity when assuming a theoretical asymptote equal to 1 (*Kmax = 1*), which was further supported by the values obtained in the R^2, AIC, and BIC analyses. When the asymptote was below 0.50, the goodness of fit gradually decreased and other models with better fits acquire major importance as is the case of the Gompertz model given that with the maximum

observed incidences (0.43, 0.28, and 0.25 for plots 1, 2, and 3, respectively), the LW development curves were fitted to the development curves observed in the three epidemics (**Figures 1, 2,** and **3**). It was clear from these Figures that below 0.50, the best settings of LW development rate were achieved when the models were fitted to the maximum disease incidences with low values. Consequently, the underestimation of highly relevant epidemiological parameters, such as the rate (r) of disease development, is extremely likely not only in sensitive models such as the monomolecular one but also in the logistic and Gompertz models when the maximum disease incidence is assumed to be higher than the real value.

Park & Lim (1985) initially observed that assuming an asymptote equal to 1 in logistic functions can lead not only to underestimate the rate (r) of disease development but also to incorrectly classify its values when fitting a data set with real asymptotes below 1 differing from each other. **Neher & Campbell** (1992) found that the greatest underestimation of the


Figure 2. Graphical representation of the goodness of fit of the models to the disease progression curves at different asymptotes or maximum disease incidences (*Kmax*) corresponding to the records collected from plot 2

r_M, r_L, and r_G represent the growth or disease development rates in the monomolecular, logistic, and Gompertz models, respectively.
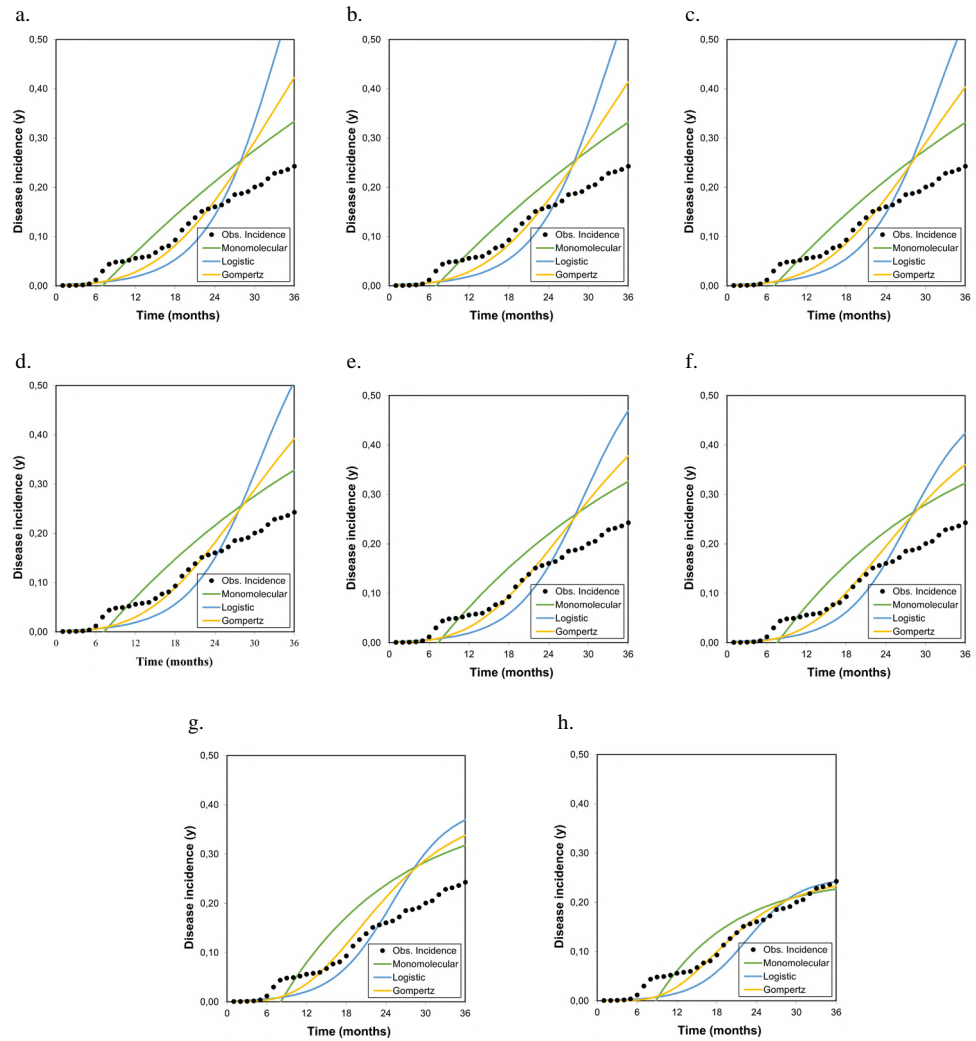
Figure 3. Graphical representation of the goodness of fit of the models to the disease progression curves at different asymptotes or maximum disease incidences (*Kmax*) corresponding to the records collected from plot 3

r_M, r_L, and r_G represent the growth or disease development rates in the monomolecular, logistic, and Gompertz models, respectively.

rate (***r***) of disease development occurred most notably when they used the monomolecular model and the ***r*** was higher when *Kmax* diminished considerably, regardless of the rate of disease development or growth rate, which agrees with our results. Furthermore, these authors concluded that the extent of disease development rate underestimation depended on the model, the disease growth rate itself, and the proximity of the real value of *Kmax* to 1 because underestimations were more pronounced with lower real maximum incidence values, higher disease growth rates, and prolonged epidemics.

Recently, **López-Vásquez, *et al.*** (2021) conducted a study to identify suitable tools for the epidemiological analysis of the LW pathosystem and found that the exponential, monomolecular, logistic, Gompertz, and Richards models using a *Kmax = 1* did not meet the assumptions of normality and constant variance, which resulted in a poor fit of the models to the observed values, both in linear and nonlinear models, thereby considering the area under the disease progression curve (AUDPC) as a descriptive alternative for this disease. The variation of the models best fitted to the observed data in the three epidemics
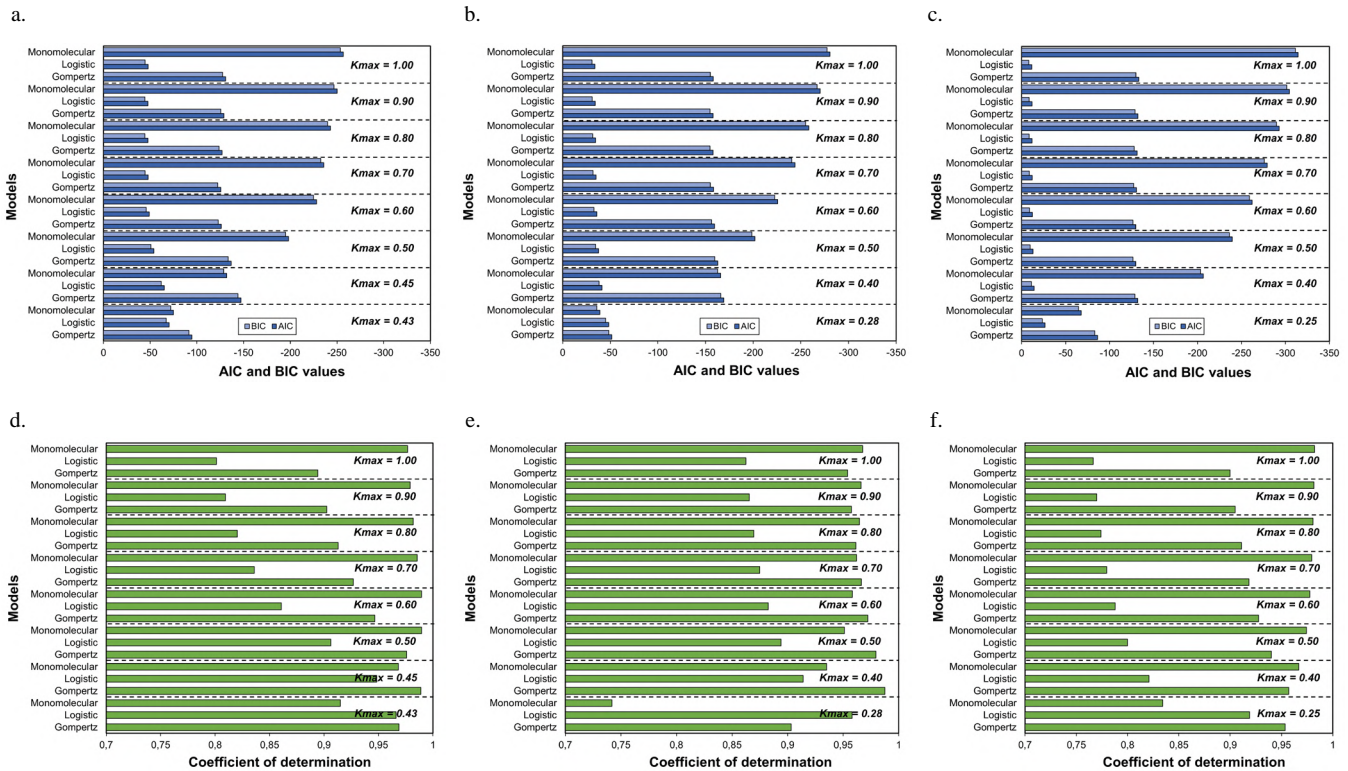
Figure 4. Goodness of fit measured by the Akaike (AIC) and Bayesian (BIC) information criteria (A, B, and C) and the coefficients of determination (R^2) of the linear regression analyses (D, E, and F) for each of the three growth models analyzed at different asymptotes or maximum disease incidences (*Kmax*). **A** and **D.** Plot 1; **B** and **E.** Plot 2; **C** and **F.** Plot 3

we evaluated was influenced by the change in the maximum incidence. This variation in the goodness of fit associated with the maximum incidences was more evident in the monomolecular model than in the logistics and Gompertz models as the *Kmax* parameter approached the absolute value of 1. However, this trend changed when it was adjusted to the maximum real disease incidence while the Gompertz and logistic models had a better fit of the data to the observed values in the three epidemics.

Our results demonstrated the sensitivity of the monomolecular model when the maximum disease incidence was assumed as higher than the real value. Besides, this model was the only one that tended to an increase in the goodness of fit when the maximum disease incidence values were approaching 1, as opposed to the logistic and Gompertz models that had stable values regardless of the increase in the maximum incidence. All epidemiological procedures used in this study showed that assuming 100% of the maximum disease incidence (*Kmax = 1*) when the final disease incidence was below this value had a significant effect on the epidemiological analysis producing inaccuracies not only by underestimating the rate of disease development but also in the selection of the best-fit model, which consequently led to errors in the mathematical description of the epidemic.

Our results also confirmed the empirical estimations made by **Analytis** (1973) and **Park & Lim** (1985) to conclude that calculations with *Kmax = 1* affected the value of *r* in the case of an actual asymptote of disease intensity below 100%. Hence, *K* = 100% as disease intensity should not be indiscriminately used in equations to compute rates. In fact, **Madden, *et al.*** (2007) raised the possibility of including *Kmax* as an additional parameter in the characterization of the dynamics of the disease given the difficulty of the objective description of epidemics with only two parameters (y_0: initial disease or constant

of integration and *r*: rate of disease development). Currently, R packages such as epifitter (**Alves & Del Ponte,** 2021) have a function that fits modifications of the two-parameter models to account for this additional parameter as the maximum asymptote.

On the other hand, if the epidemic had been characterized with the type of model that best fitted the data when *Kmax = 1*, the LW epidemic would have been classified as having a monocyclic behavior given that the best settings were obtained with the monomolecular model. However, when we adjusted the *Kmax* to the actual *Kmax < 1*, the LW epidemic changed from a monocyclic to a polycyclic behavior because the best-adjusted models were those of the Gompertz type. These types of results have epidemiological implications in result analyses, as well as technical implications in the strategic management of the disease because a significant part of the comprehensive management strategy of the disease is based on the type of behavior of the epidemic. It is important to note that given that WL is currently considered an official control disease in Colombia (Resolution 004170; **ICA,** 2014), all the plots we selected were continuously subjected to phytosanitary measures for foci management and containment. This kind of anthropogenic event explains to a certain extent why this epidemic had such low final incidences in a given time period. Additionally, we should keep in mind that time (**t**) is a continuous parameter and, therefore, a 36-month-period cannot be considered the total length of an epidemic in a crop such as oil palm with an eventual economic life span of up to 25 years. This could be different in a semi-annual crop that can be harvested in less than 12 months. Therefore, it is possible that a longer time (more than 36 months) of evaluation may result in additional increases in disease incidence given the availability of healthy tissue, which could reach 100 % if no management practices are performed.

Conclusions

One of the main purposes of epidemiological analyses using models is to achieve a better understanding of how diseases develop in plant populations over time and how other factors influence their development. When selecting an appropriate model, one must have the ability to understand and explain the factors associated with the estimation of the parameters that describe the epidemic. However, the models often lack the necessary flexibility to adapt to the various changes occurring during the development of a disease. In this sense, the goodness of fit of a model depends on the increase in the number of estimated parameters and, consequently, on the flexibility of the selected model.

Although the true asymptote for the development of any disease is a function of the amount of susceptible tissue available, we should be aware of the risk of assuming a theoretical value of the maximum disease incidence in the host in a generalized and arbitrary manner. There are many epidemics with a maximum incidence below 100% due to unfavorable environments, insufficient populations of insect vectors or propagules, levels of host resistance, disease reduction practices focused on the early elimination of diseased plants, or other unspecified physical or biological factors.

Our results demonstrate the importance of including *Kmax* in the description of the temporal behavior of an epidemic. The consequences of using incorrect values reflect not only in the erroneous estimation of parameters, such as y_0 and *r* but also in the risk of incorrectly interpreting the temporal behavior of the epidemic because of a false fit of the selected model. Assuming the parameter *Kmax* as unique and absolute may lead to inaccuracies in the type of epidemiological analysis and, consequently, to a biased description of reality.

Acknowledgments

Supplementary information

Table S1: Linear regression analysis of the three growth models assessed at different asymptotes or maximum disease incidence (*Kmax*) of the records, corresponding to plot 1. See the table S1 in https://www.raccefyn.co/index.php/raccefyn/article/view/1571/3188
Table S2: Linear regression analysis of the three growth models assessed at different asymptotes or maximum disease incidence (*Kmax*) of the records, corresponding to plot 2. See the table S1 in https://www.raccefyn.co/index.php/raccefyn/article/view/1571/3188
Table S3: Linear regression analysis of the three growth models assessed at different asymptotes or maximum disease incidence (*Kmax*) of the records, corresponding to plot 3. See the table S1 in https://www.raccefyn.co/index.php/raccefyn/article/view/1571/3188

Author contributions

JML conceived the idea, conducted data analysis, and wrote the first draft of the manuscript; JCZ supervised the project and revised/conceptualized the versions of the manuscript. All authors provided feedback and approved the final manuscript.

Conflicts of interest

On behalf of all authors, the corresponding author declares there is no conflict of interest.

Funding

References

Analytis, S. (1973). Zur methodik der analyse von epidemien dargestellt am apfelschorf (*Venturia inaequalis* (Cooke) Aderh.). Acta Phytomed. **1:** 1-75.

Analytis, S. (1979). Die transformation von befallswerten in der quantitativen phytopathologie. Phytopathol. Z. **96:** 156-171.

Alves, K.S. & Del Ponte, E.M. (2021). Analysis and simulation of plant disease progress curves in R: introducing the epifitter package. Phytopathol. Res. **3:** 22. 13 p.

Arneson, P.A. (2001). Plant disease epidemiology: temporal aspects. The Plant Health Instructor. Accessed on: November 19, 2021. Retrieved from:

https://www.apsnet.org/edcenter/disimpactmngmnt/topc/EpidemiologyTemporal/Pages/default.aspx

Bowen K. L. (2015). Models of disease progress, pp. 9-15. In: Stevenson KL, Jeger MJ, editors. Exercises in Plant Disease Epidemiology, 2nd ed. The American Phytopathological Society (APS Press). St Paul.

Burnham, K. P. & Anderson, D. (2002). Model selection and multimodal interference. 2nd ed. Springer-Verlag, New York. 23 pp.

Campbell, C. L. (1998). Disease progress in time: modelling and data analysis, pp. 181-206. In: D. Gareth Jones (ed.), The epidemiology of plant disease. Kluwer Publishers, Dordrecht.

Campbell, C. L. & Madden, L.V. (1990). Introduction to plant disease epidemiology. Wiley, New York. 532 pp.

Forrest, F. W. (2007). The role of plant disease epidemiology in developing successful integrated disease management programs, pp. 45-79. In: A. Ciancio and K. G. Mukerji (eds.), General Concepts in Integrated Pest and Disease Management. vol. 1, Springer, Dordrecht, The Netherlands.

Gigot, C. (2018). Epiphy: analysis of plant disease epidemics. Accessed on: November 19, 2021. Retrieved from: https://CRAN.R-project.org/package=epiphy

Gilligan, C. A. (1990). Comparison of disease progress curves. New Phytol. **115** (2): 223-242.

Hosmer, D. W. & Lemeshow, S. (2000). Applied logistic regression, 2nd ed. John Wiley & Sons, Inc. 392 pp.

ICA. (2014). Resolución ICA 004170: Por medio de la cual se declaran las plagas de control oficial en el cultivo de palma de aceite en el territorio nacional y se establecen las medidas fitosanitarias para su manejo y control. Accessed on: November 19, 2021. Retrieved from: https://www.ica.gov.co/getattachment/0e5ecaac-ccb6-444f-96bb-3d488048b51f/2014R4170.aspx

Jeger, M. J. (1982). Using growth curve relative rates to model disease progress of apple powdery mildew. Prot. Ecol**. 4:** 49-58.

Kiyosawa, S. (1972). Mathematical studies on the curve of disease increase. Ann. Phytopathol. Soc. Jpn. **38:** 30-40.

Kushalappa, A. C. & Ludwing, A. (1982). Calculation of apparent infection rate in plant disease: Development of a method to correct for host growth. Phytopathology. **72:** 1373-1377.

Kranz, J. (2003). Comparison of temporal aspects of epidemics: the disease progress curves, pp 93-133. In: Comparative Epidemiology of Plant Diseases. Springer, Berlin, Heidelberg.

López-Vásquez, J. M., Mesa-Fuquen, E., Acosta, M. C. Tovar, J. P. (2021). Descripción del comportamiento de la enfermedad marchitez letal de la palma de aceite a partir del análisis epidemiológico temporal. Revista Palmas. **42** (2): 69-80.

Madden, L. V. (1986). Statistical analysis and comparison of disease progress curves, pp. 55-84. In: K. Leonard and W. E. Fry, (eds.), Plant Disease Epidemiology. Macmillan Publishing Co., New York.

Madden, L. V., Hughes, G., van den Bosch, F. (2007). The study of plant disease epidemics. APS Press, St. Paul, MN. 421 pp.

Mazerolle, M. J. (2019). *AICcmodavg*: model selection and multimodel inference based on (Q) AIC(c). R package version 2.2-2. The R Foundation for Statistical Computing, Vienna, Austria.

McCullagh, P. & Nelder, J. A. (1992). Generalized linear model. Chapman & Hall. London. 526 pp.

Neher, D. & Campbell, C. L. (1992). Underestimation of disease progress rates with the Logistic, Monomolecular, and Gompertz models when maximum disease intensity is less than 100 percent. Phytopathology. **82** (8): 811-814.

Neher, D. A., Reynolds, K. L. Campbell, C.L. (1997). Analysis of disease progress curves using linear models. Chapter 7. In: L. J. Francl and D. A. Neher (eds.), Exercises in Plant Disease Epidemiology. APS Press. St. Paul, Minnesota. 242 pp.

Nutter, F. W., Eggenberger, S. K., Littlejohn, K. J. (2017). Visualizing, describing, and modeling disease progress curves using Epimodel, pp 21-30. In: Stevenson KL, Jeger MJ, editors. Exercises in Plant Disease Epidemiology, 2nd ed. The American Phytopathological Society (APS Press). St Paul. 276 pp.

Park, E. W. & Lim, S. M. (1985). Empirical estimation of the asymptotes of disease progress curves and the use of the Richards generalized rate parameters for describing disease progress. Phytopathology. **75:** 786-791.

Pinheiro, J. C., Bates, D., DebRoy, S., Sarkar, D. (2004). *nlme*: linear and nonlinear mixed effects models. R package version **3:** 1-52. The R Foundation for Statistical Computing, Vienna, Austria.

R Core Team. (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rocha, P. J., Tovar, J. P., Gutiérrez, D. F., Mosquera, M. (2007). Marchitez letal en palma de aceite. Boletín Técnico **N°22. Bogotá, Colombia.**

Seem, R. C. (1988). Measurement and analysis of the effects of crop development on epidemics, pp. 51-68. In: J. Kranz and J. Rotem, (eds.), Experimental Techniques in Plant Disease Epidemiology. Springer-Verlag, Berlin. 317 pp.

Turner, M. G., Blumenstein, B. A., Sebaugh, J. L. (1969). A generalization of the logistic law of growth. Biometrics. **25:** 557-580.

Van Maanen, A. & Xu X. (2003). Modelling plant disease epidemics. Europ. J. Plant. Pathol. **109** (7): 669-682.

Van der Plank, J. E. (1963). Plant disease: epidemics and control. Academic Press, New York, USA. 349 pp.

Xu, X. (2006). Modelling and interpreting disease progress in time, pp. 215-238. In: Cooke, B. M., Gareth Jones, D., and Kaye, B. (eds.) The Epidemiology of Plant Disease. Springer, Dordrecht, The Netherlands. 576 pp.