A Flexible Statistical Growth Model for Describing Plant Disease Progress

CHOONG-HOE KIM

金忠會 : 植物病 進展의 한 柔軟的인 統計的 生長 모델

Korean J. Plant Prot. 26(1) : 31~36 (1987)

ABSTRACT A piecewise linear regression model able to describe disease progress curves with simplicity and flexibility was developed in this study. The model divides whole epidemic into several pieces of simple linear regression based on changes in pattern of disease progress in the epidemic and then incorporates the pieces of linear regression into a single mathematical function using indicator variables. When twelve epidemic data obtained from the field experiments were fitted to the piecewise linear regression model, logistic model and Gompertz model to compare statistical fit, goodness of fit was greatly improved with piecewise linear regression compared to other two models. Simplicity, flexibility, accuracy and ease in parameter estimation of the piecewise linear regression model were described with examples of real epidemic data. The result in this study suggests that piecewise linear regression model is an useful technique for modeling plant disease epidemic.

INTRODUCTION

For the purpose of analysis and comparison, progress of plant diseases has been often described into certain simple mathematical equations. Some researchers (1, 8) used biological growth functions to describe increase of disease in time. Logistic and Gompertz equations are the examples of such a growth function. The equations were linearized by appropriate transfermation and slopes of the lines were compared as a parameter of rate of disease increase in time (2). In this type of analysis of epidemic, problems lie in whether the such simple equations are really valid for depicting various patterns of epidemic progress that is the product of dynamic interaction of host, parasite and its environment. Although several other biological growth functions (2, 5) have been proposed as models for disease progression, researchers often find that their data were not fit with the proposed equations. In fact, several researchers suggested that no growth equation is suited for all the known epidemic patterns (2, 5). If the inappropriate model is chosen, inaccurate estimates of epidemic parameters result as indicated by Berger (1). For this reason, it is desirable to use a single mathematical function that is more flexible for describing various patterns of disease progress. Weibull probability density function was used to increase model flexibility (7), but have problems in difficulties of finding and measuring parameter values and interpretation of those parameters.

In this study, piecewise linear regression model was developed to describe plant disease progressions with increased flexibility, accuracy and ease in parameter estimation. The proposed model was evaluated and compared with logistic and Gompertz models with actual disease progress data.

DESCRIPTION OF THE MODEL

Although disease progress curve is typically sigmoid, pattern of disease increase deviates frequently from the sigmoid shape due to unexpected changes in cultural and meteological conditions or changes in host resistance associated with the growth stage of plants. Figure

Department of Plant Pathology, Agricultural Sciences Institute, Suwon 170, Korea (農業技術研究所, 病理科)

1 shows a real progress curve of rice blast disease in the field and illustrates this situation. Pattern of disease progress differed among certain ranges of epidemic periods. The epidemic in Figure 1 may be divided into three portions that have different rates of disease increase. Each portion of the epidemic may be described with a simple linear relationship, hence the entire epidemic could be described with three pieces of simple linear regression. Indicator variables may be used to incorporate the separate pieces of linear regression into a single mathematical equation, that is a piecewise linear regression.

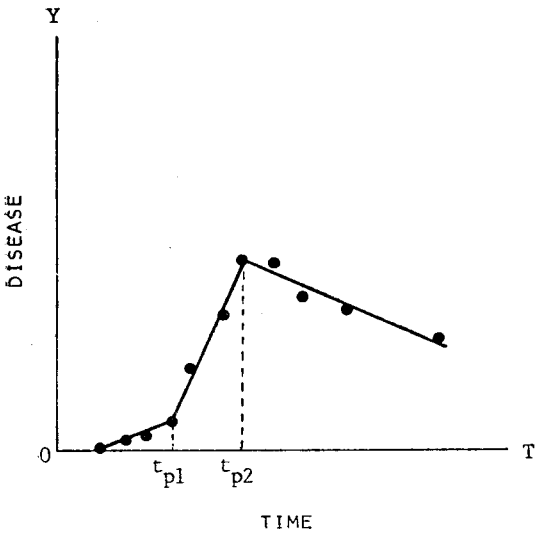

Figure. 1. Illustration of piecewise linear regression with disease progress data of rice blast obtained from the field experiments. t_{p1} and t_{p2} are the points of time when major changes in disease progression occurred.

General piecewise linear regression model for the illustration of Figure 1 may be expressed as follows:

$$Y_i=b_0+b_1t_{i1}+b_2(t_{i1}-t_{p1})t_{i2}+b_3(t_{i1}-t_{p2})t_{i3}+e_i \quad (1)$$

where Y_i is disease proportion or % disease; t_{i1} is time variable; t_{p1} and t_{p2} are the constants that are the points of time where slope change occurs; t_{i2} and t_{i3} are indicator variables defined as follows:

$$t_{i2}=\begin{cases}1 & \text{if } t_{i1}>t_{p1}\\ 0 & \text{otherwise}\end{cases}$$

$$t_{i3}=\begin{cases}1 & \text{if } t_{i1}>t_{p2}\\ 0 & \text{otherwise}\end{cases}$$

$$t_{p1}<t_{p2}$$

The equation (1) provides a three-piecewise linear regression and its response function is:

$$E(Y)=b_0+b_1t_1+b_2(t_1-t_{p1})t_2+b_3(t_1-t_{p2})t_3 \quad (2)$$

The first-piecewise linear regression is when $t_1\leq t_{p1}$ so that $t_2=0$, $t_3=0$. The equation (2) becomes:

$$E(Y)=b_0+b_1t_1 \quad (3)$$
$$t_1\leq t_{p1}$$

where b_1 is rate parameter of disease increase and b_0 is the Y intercept.

The second-piecewise linear regression is when $t_{p1}<t_1\leq t_{p2}$, so that $t_2=1$, $t_3=0$. The equation (2) becomes:

$$E(Y)=(b_0-b_2t_{p1})+(b_1+b_2)t_1 \quad (4)$$
$$t_{p1}<t_1\leq t_{p2}$$

The rate parameter in this time range is (b_1+b_2) and Y intercept is $(b_0-b_2t_{p1})$.

On the other hand, when $t_1>t_{p2}$ so that $t_1=t_2=1$, the equation (2) becomes:

$$E(Y)=(b_0-b_2t_{p1}-b_3t_{p2})+(b_1+b_2+b_3)t_1 \quad (5)$$
$$t_1>t_{p2}$$

The parameters of slope and Y intercept of the third-piecewise linear regression become $(b_1+b_2+b_3)$ and $(b_0-b_2t_{p1}-b_3t_{p2})$, respectively. Parameters for this example are shown in Figure 2.

In actual application of piecewise linear regression, number of pieces of linear regression is determined by the shape of disease progress curves. The relationships between number of pieces of linear regression (N_R), number of time points that slope changes (N_P), and number of indicator variables used (N_I) are:

$$N_R=N_P+1=N_I+1 \text{ or}$$
$$N_P=N_I=N_R-1$$

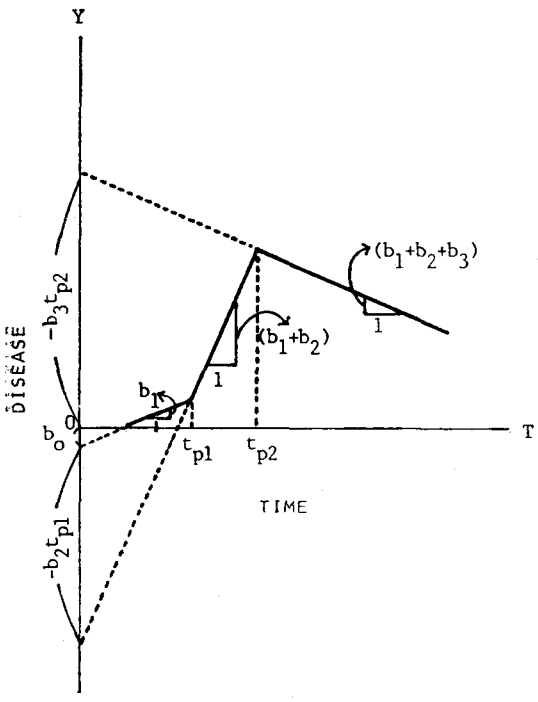

Figure. 2. Illustration of parameters of piecewise linear regrsssion model : $E(Y)=b_0+b_1t_1+b_2(t_1-t_{p1})t_2+b_3(t_1-t_{p2})t_3$, where t_1 is an independent variable (time), and t_2 and t_3 are the indicator variables associated with t_{p1} and t_{p2}.

Majority of epidemic progress pattern may be described with two- or three-piecewise linear regression model. Some more complex case of epidemic may need more than three-piecewise linear regression model for describing its disease progress pattern.

The extention of the model (2) to more than three-piecewise linear regression is straightforward. For instance, if the slope of the regression line sets to change at the three different points of time, t_{p1}, t_{p2} and t_{p3}, the response function of the model would be:

$$E(Y)=b_0+b_1t_1+b_2(t_1-t_{p1})t_2+b_3(t_1-t_{p2})t_3 + b_4(t_1-t_{p3})t_4 \quad (7)$$

$$t_{p1}<t_{p2}<t_{p3}$$

where: t_1=time

$$t_2=\begin{cases}1 & \text{if } t_1>t_{p1}\\ 0 & \text{otherwise}\end{cases}$$

$$t_3=\begin{cases}1 & \text{if } t_1>t_{p2}\\ 0 & \text{otherwise}\end{cases}$$

$$t_4=\begin{cases}1 & \text{if } t_1>t_{p3}\\ 0 & \text{otherwise}\end{cases}$$

To illustrate the use of model (2), consider the actual disease progress data in Table 1. From the plot of the data, it appeared that pattern of disease progression changes at 21 and 57 days after inoculation. So the three-piecewise linear regression model (2) is to be employed with $t_{p1}=21$ and $t_{p2}=57$. The fitting of regression model (2) with data of Table 1 becomes routine by following the standard linear regression procedure (6). The fitted response function is:

$$E(Y)=-30.19+2.18t_1-1.57(t_1-21)t_2-2.17(t_1-57)t_3$$

$$R^2=0.9434$$

Table 1. Disease progress data of rice blast epidemic observed in field condition and matrix of time(t) variable for an example of generating piecewise linear regression equation. t_2 and t_3 are indicator variables associated with points of time where major slope changes occurred (21 and 57 days after inoculation in this example)

Y_i (% disease)	t_i (days after inoculation)	t_1	$(t_1-21)t_2$	$(t_1-57)t_3$
0.4	14	14	0	0
14.7	21	21	0	0
16.9	29	29	8	0
30.0	36	36	15	0
32.7	43	43	22	0
31.3	50	50	29	0
36.3	57	57	36	0
16.2	71	71	50	14

where: $t_2 = \begin{matrix}1 \text{ if } t_1>21\\ 0 \text{ otherwise}\end{matrix}$

$t_3 = \begin{matrix}1 \text{ if } t_1>57\\ 0 \text{ otherwise}\end{matrix}$

From this fitted model, expected unit disease is estimated to increase by 2.18 for unit increase of time up to 21 days after inoculation, and by 0.61 (=2.18−1.57) up to 57 days after inoculation and to decline by 1.56 (=2.18−1.57−2.17), thereafter.

APPLICATION OF THE MODEL

Brief introduction of the logistic and Gompertz model. For the purpose of comparison of the piecewise linear regression with other biological growth equations, the logistic and Gompertz models were briefly explained in this section.

Logistic growth model for plant disease epidemic is expressed as:

$$y=1/(1+\exp(-\ln y_0/(1-y_0))+rt) \quad (8)$$

where: y=disease proportion; y_0=initial disease proportion; r=rate parameter; t=time. Equation (8) is derived from the differential growth function:

$$dy/dt=ry(1-y) \quad (9)$$

The rate parameter (r) of the equation (9) was termed as the apparent infection rate by Vanderplank (8). The equation (8) is linearized by transfermation and becomes following:

$$\text{logit } (y)=\text{logit } (y_0)+rt \quad (10)$$

The integrated logistic curve is sigmoid and symmetrical at its center of sigmoid curve (1, 8).

Equation of Gompertz model for biological growth is:

$$y=\exp\ (-b(\exp(-rt))) \quad (11)$$

where: b=position parameter; r=rate parameter; t=time. Equation (11) is derived by integrating from the differential equation:

$$dy/dt=ry\ \ln(1/y) \quad (12)$$

The equation (11) is linearized by transfermation and takes the following form:

$$-\ln(-\ln(y))=-\ln(b)+rt \quad (13)$$

The integrated curve of the equation (11) is sigmoid but asymmetrical and appreciably skewed to the right direction (1).

Comparison of epidemic. Piecewise linear regression, logistic model and Gompertz model were compared by determining model fitness with actual disease progress curves. Twelve epidemic curves were obtained from the field water-management experiments on rice blast disease conducted in Louisiana, USA in 1984 and 1985 (3, 4). For logistic and Gompertz models, disease proportions for 12 curves were treated with the appropriate transfermation (equations 10 and 12). For piecewise linear regression model, points of time at which disease progress pattern changes were determined from the plots of the data of 12 epidemics. Matrices of time (t) variables were then obtained with proper indicator variables. The piecewise linear regression was generated based on these matrices. The least square techniques was used to obtain parameter estimates. Statistical fitness of the three models was determined by comparing coefficient of determination with the linear regressions. Computing was done using CRISP (Crop Research Integrated Statistical Package) language from the VAX 11/785 main frame computer in the Computing Center of Rural Development Adminstration. The results are summarized in Table 2. Among the three models, piecewise linear regression model provided the best statistical fit for all 12 epidemics. Particularly for the last three epidemics listed in Table 2, statistical fit was greatly improved with the piecewise linear regression. Seven out of twelve epidemics were fitted well to the two-piecewise linear regression model and remaining six epidemics were described with three-piecewise linear regression model with

Table 2. Comparison of the statistical parameters of 12 rice blast epidemics fitted to piecewise linear regression model, to logistic model and to Gompertz model

Epidemic	Number of observation	Piecewise linear regression model[a]				Logistic model[b]		Gompertz model[c]	
		r_1	r_2	r_3	R^2	r	R^2	r	R^2
1	11	1.118	0.442	—	0.958	0.166	0.707	0.062	0.912
2	11	0.227	0.980	—	0.977	0.164	0.717	0.051	0.929
3	11	1.098	0.561	—	0.979	0.159	0.757	0.061	0.953
4	11	0.318	2.231	−1.279	0.973	0.136	0.693	0.050	0.934
5	11	0.138	2.478	−1.909	0.987	0.129	0.686	0.051	0.841
6	11	0.109	0.614	—	0.972	0.138	0.696	0.036	0.870
7	8	2.354	−1.383	—	0.987	0.084	0.856	0.055	0.956
8	8	3.513	−1.841	−1.353	0.989	0.076	0.806	0.050	0.904
9	8	4.318	−3.033	—	0.957	0.083	0.880	0.058	0.944
10	8	3.846	−3.415	−0.550	0.991	0.036	0.520	0.018	0.596
11	8	3.228	−2.726	—	0.948	0.067	0.498	0.027	0.558
12	8	2.185	−1.568	−2.166	0.943	0.051	0.356	0.017	0.358

[a] Parameters 'r_1, r_2 and r_3' are the partial regression coefficients of the model. R^2 is coefficient of determination of the model and is significant at P=.05.

[b] Rate parameter 'r' is apparent infection rate termed by Vanderplank (8). The model is significant at P=.05.

[c] Rate parameter 'r' is the slope of the linearized Gompertz equation. The model is statistically significant at P=.05.

good statistical fit. Gompertz model provided better fit compared to the logistic model for all 12 epidemics. Particularly, biological growth model becomes less fit when the disease increase slowed down rapidly as the seasons progressed. Coefficients of determination for 12 disease progress curves ranged from 97 to 99% with the piecewise linear regression model, 60 to 94% with the logistic model and 60 to 98% with Gompertz model, respectively.

DISCUSSION

The linearization of disease progress curves is necessary procedure in analysis of epidemic to compare epidemics, to predict disease progress and to estimate effects of various control measures being applied on plant diseases. Biological growth models for disease progress curves have often had problems in linearization eventhough various transfermations were applied. This was usually due to the fact that disease progress curves of many epidemics often deviate significantly from the standard curve of the model employed. Although disease progress curves were successfully linearized, transfermations may be prone to obscure the fundamental interconnection between the variables, and thus may distort important biological facts (6). When the logarithmic transfermation is employed for linearization of biological growth model, it needs to check whether the transformed error terms meet the condition of the untransformed error terms. This kind of residual analysis has not been done usually. If inadequate model is used for epidemic analysis, parameters obtained from the model become incorrect.

Piecewise linear regression model proposed in the study avoids the problems present in flexibility and transfermation that are described in other models. Piecewise linear regression model is relatively simple, easy to understand and it enables analysis of almost all shapes of disease progress curves without any transfermation. This model also allows calculation of the rate parameters separately at a

certain range of time and provides additional information on the disease progression. Unexpected changes occurred in the epidemic can be partly incorporated into the model using indicator variables by determining the points of time where the pattern of disease progression changes.

Most of biological growth models are purely phenological and deterministic and do not provide informations on the nature of host, pathogen and environment interactions, since all interaction factors are incorporated into a single rate parameter '*r*' (2). This limitations are less severe in the piecewise linear regression model because the model includes the time points when the significant changes in disease progress occur during the epidemic.

Through separating the entire epidemic into several pieces of linear regression, the proposed model greatly improved model fitness. The increased statistical fit of the model can provide more precise estimates of epidemic parameters. The increased preciseness and flexibility of the piecewise linear regression model can be effectively used as a powerful tool for the analysis of epidemic and simulation studies.

Between the logistic and Gompertz models, Gompertz model provided better statistical fit than the logistic model did in this study. This was expected because disease progress curves of the 12 epidemics were asymmetrical and skewed to the right. Better statistical fit with Gompertz model campared to the logistic was reported for epidemic progress of several plant diseases (1).

摘 要

植物病 進展曲線을 간편하고 융통성있게 기술하는 切片 1次 回歸모델이 本 硏究에서 提案되었다. 이 모델은 病進展狀況을 그 進展形態에 따라 少數의 1次 回歸式으로 나누고 指標變數를 使用하여 다시 한개로 묶어 作成된다. 圃場試驗에서 얻은 12個의 實際病進展狀況에 對한 切片 1次 回歸모델의 統計的 適合度는 旣存의 두모델(Logistic모델과 Gompertz모델)에 比하여 增進되었으며 이 모델이 가진 單純性, 융통성 및 母數豫測의 容易性이 論議되었다. 그 結果, 切片 1次 回歸모델은 植物病 進展을 記述하는 한 統計的 모델로써 有用하게 使用될 수 있으리라 생각된다.

LITERATURES CITED

1. Berger, R. D. 1981. Comparison of the Gompertz and logistic equations to describe plant disease progress. Phytopathology 71 : 716~719.
2. Hau, B., S. P. Eisensmith and J. Kranz. 1985. Construction of temporal models: II. Simulation of aerial epidemics. Pages 31~65 in: C. A. Gilligan, ed. Mathematical Modelling of Crop Disease. Academic Press. 255 pp.
3. Kim, Choong-Hoe. 1986. Effect of water-management on the etiology and epidemiology of rice blast caused by *Pyricularia oryzae* Cav.. Ph.D. Dissertation, Louisiana State Univ., Baton Rouge. 170 pp.
4. Kim, Choong-Hoe, M. C. Rush, D. R. Mac Kenzie and M. A. Marchetti. 1987. Effect of water-management on rice blast epidemiology and flooding dependent resistance. Phytopathology 77 : (in press).
5. Kranz, J. 1974. Comparison of epidemics. Annu. Rev. Phytopathol. 12 : 355~374.
6. Neter, J., and W. Wasserman. 1974. Applied Linear Statistical Models. Richard D. Irwin, Inc. 842 pp.
7. Pennypacker, S. P., Knoble, H. D., Antle, C. E., and Madden, L. V. 1980. A flexible model for studing plant disease progression. Phytopathology 70 : 232~235.
8. Vanderplank, J. E. 1963. Plant Disease: Epidemics and Control. Academic Press, N.Y. 349pp.