Comparison of the Gompertz and Logistic Equations to Describe Plant Disease Progress

R. D. Berger

Professor, Department of Plant Pathology, University of Florida, Gainesville 32611.
Appreciation is expressed to J. Cornell, Department of Statistics, and J. W. Mishoe, Department of Agricultural Engineering for assistance with the statistical and mathematical interpretations.
Journal Series Paper 2471 of the Florida Agricultural Experiment Station.
Accepted for publication 13 August 1980.

ABSTRACT

Berger, R. D. 1981. Comparison of the Gompertz and logistic equations to describe plant disease progress. Phytopathology 71:716-719.

The Gompertz transformation effectively linearized 113 disease progress curves of nine pathosystems. The Gompertz model avoided the curvilinearity commonly associated with logistically transformed values. Estimation of epidemic rate, projection of future disease severity, and determination of initial disease were more accurate with the Gompertz than with the logistic model. Since many pathosystems have asymmetrical disease progress curves, transformations other than the logistic may be more appropriate to estimate epidemic parameters.

Additional key words: simple linear regression, curve fitting, epidemic analysis.

Thorough description and accurate analysis of the dynamic process of plant disease increase in time is needed to compare epidemics. Typically, the sigmoidal progress curves of compound interest diseases are linearized with the logistic transformation to aid in interpretation (1,10,17,21,24,27). The daily increase of disease is commonly skewed to the right so the logistic equation for linearization may be inadequate or inappropriate. Nevertheless, research workers continue to use it despite the caution urged by Kranz (18): "Do not apply a transformation model blindly to any disease, check suitability first by verification of the underlying distribution." If the logistic transformation is used for a skewed distribution of values, erroneous interpretation of the epidemic parameters may result. Alternate transformations for growth curves are available (17). Disease progress curves of apple scab were fit poorly with the logistic model, but Analytis (1) obtained better fits with the Bertalanffy, Gompertz, and Mitscherlich transformations. Berger and Mishoe (12) also obtained good statistical fit for these three transformations applied to progress curves of several plant pathosystems. Plaut (23) obtained better statistical fit with the Gompertz model compared to the logistic for progress curves of several plant diseases. Griggs et al (15) applied polynomial curve fitting to compare epidemics of *Cronartium fusiforme* but this procedure is cumbersome (23). Another transformation that can be used to describe disease progress is the Weibull probability density function. One of the advantages of the Weibull model is its flexibility (22).

In this paper, I report an extension of the utility of the Gompertz transformation in analysis of plant disease progress curves. An abstract of this work has been published (11).

DESCRIPTIONS OF THE MODELS

The logistic model. As a population approaches its uppermost limit, the growth is slowed by the feedback information of limits on the system. Verhulst (as cited in references 5 and 17) used this feedback in a population growth model that is now termed the logistic model. The logistic model equation for plant disease progress is

$$y = 1/(1 + \exp(-[a + rt])) \quad (1)$$

in which y = disease proportion in the range $0<y<1.$, a = logit (y_0), r=rate, and t=time. The logistic model equation has also been described (21) as $y=1/(1+b \times \exp(-rt))$ in which $b=(y_{max}/y_0)-1$.

The transformation equation is:

$$\text{logit}(y) = \ln(y/(1-y)). \quad (2)$$

The integrated logistic curve is sigmoid and symmetrical about its central point of inflection (Fig. 1). A plot of the derivative is the bell-shaped curve of normal distribution (Fig. 2). When the daily increase of disease has a skewed distribution, the transformed values (logits) are nonlinear. The logistically transformed disease progress curves are frequently characterized by steep slopes at $y<\sim 0.05$, linearization for the range $\sim 0.05<y<\sim 0.6$, and values that fall below the general slopes when $y>\sim 0.6$. Berger (9,10) drew attention to the rapid initial increase of logistically transformed curves and Zadoks (26) noted the common occurrence of values below the line at upper levels of y. Commonly, research workers arbitrarily draw an eye-fitted line through the plotted values or use simple linear regression techniques to fit the line (23). Both y_0 and y_t (y at any future time) are likely to be overestimated by the fitted logistic line through the skewed distribution of values (Fig. 3).

The Gompertz model. The British mathematician, B. Gompertz,

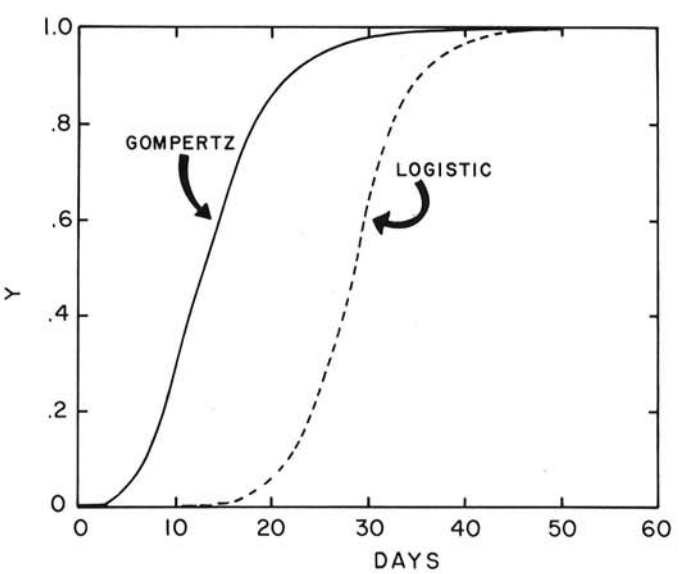

Fig. 1. Sigmoid curves from simulations with two growth models. Both epidemics initialized with y_0=0.0004. The rates were k=0.2 for the Gompertz and r=0.325 for the logistic models.

0031-949X/81/07071604/$3.00/0

derived this model to develop actuarial tables (5). In the model's original form [$y = \exp(-B \times \exp(kt))$] Gompertz was concerned with decreasing numbers (of survivors). The change of sign for the rate parameter (k) changed the equation into a model for increasing growth

$$y = \exp(-B \times \exp(-kt)). \quad (3)$$

Gompertz' growth model frequently has been used by ecologists to explain biological phenomena (5). The integrated curve is sigmoid but it is asymmetrical about its point of inflection (Fig. 1). The plot of the derivative is noticeably skewed to the right (Fig. 2). The transformation equation is

$$Y = -\ln(-\ln(y)); \quad (4)$$

Y is hereafter called "gompit (y)."

The B parameter is a position parameter; ie, it positions the origin of the transformed line onto the vertical axis at time t=0. The B parameter can be calculated by rearrangement of equation 3 and solution for B:

$$B = -\ln(y_0);\ \exp(-kt) = 1 \text{ at } t=0. \quad (5)$$

The B parameter also can be calculated from any y value if k is known, also from rearrangement of equation 3:

$$B = -\ln(y_t)/\exp(-kt);\ t<0. \quad (6)$$

Alternatively, the B parameter can be obtained graphically from the line fitted through the gompit (y) values and the intercept with the vertical axis at t=0 by

$$B = \exp(-\text{gompit}(y_0)). \quad (7)$$

When disease progress curves are fitted to the Gompertz model in statistical curve fitting programs (4), an estimate of the B parameter is generated. The initial disease (y_0) can then be calculated from this generated estimate of B as the inverse of equation 5:

$$y_0 = \exp(-B). \quad (8)$$

The origin of the curve of transformed values can be fixed on the vertical axis at time t=0 by the inverse of equation 7:

$$\text{gompit}(y_0) = -\ln(B). \quad (9)$$

Comparison of logistic and Gompertz models. The k parameter of Gompertz' model is the rate parameter, which corresponds to the apparent infection rate (r) of the logistic equation as used by Vanderplank (24).

The k values are calculated similarly to the r values by the two-point method; ie, for the logistic model:

$$r = (\text{logit}(y_2) - \text{logit}(y_1))/(t_2 - t_1); \quad (10)$$

and for the Gompertz model:

$$k = (\text{gompit}(y_2) - \text{gompit}(y_1))/(t_2 - t_1). \quad (11)$$

For both models, the rate parameters can also be obtained by the slope values of the simple linear regression of the transformed disease proportions over time.

Plots of the derivatives of both equations can be superimposed for comparative purposes. The plot of the logistic derivative must be shifted to the left (earlier in time) to make the points of inflection of both curves coincide. The two curves are then reasonably similar in shape for a period of time (Fig. 4). Thus, both transformations effectively linearize values in the range of $\sim 0.05 < y < \sim 0.6$ for both symmetrical and asymmetrical populations. The fit with the logistic equation is poorer than that of the Gompertz equation for values outside that range in the typical asymmetrical disease progress curves.

APPLICATION OF MODELS

Comparisons of epidemics. For the logistic model, tables of logits were available in the literature (24,27). For the Gompertz model, a table of gompits (Table 1) was prepared with the appropriate transformation (equation 4) for disease proportions in the range $0.01 < y < 0.99$. The two models were compared by determining

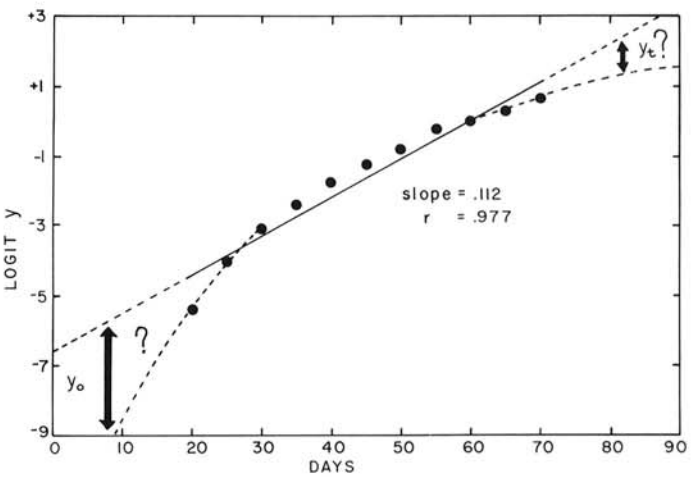

Fig. 3. Logistic transformation of an asymmetric sigmoid growth curve. The Gompertz transformation for these values would be a straight line with the correlation coefficient of 1.0 and k=0.05. The projections for y_0 and y_t with logistic model overestimate the real values.

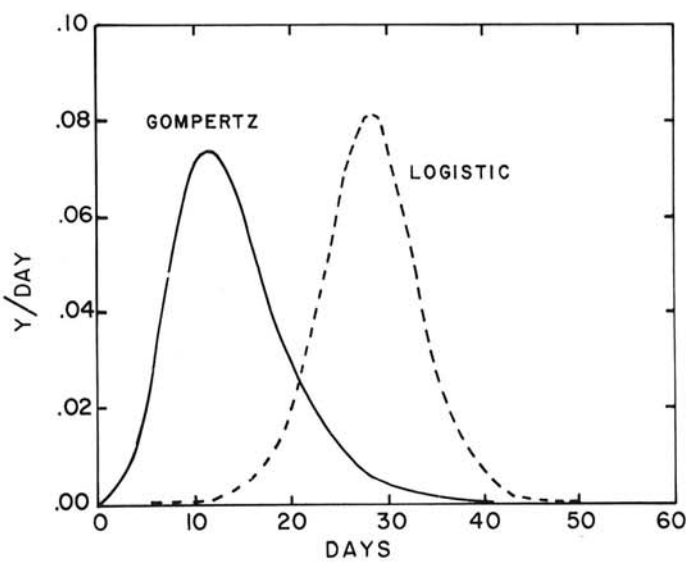

Fig. 2. The daily growth (the derivative, dy/dt) for two growth models. The simulation parameters were y_0=0.0004; rates k=0.2 for the Gompertz and r=0.325 for logistic models.

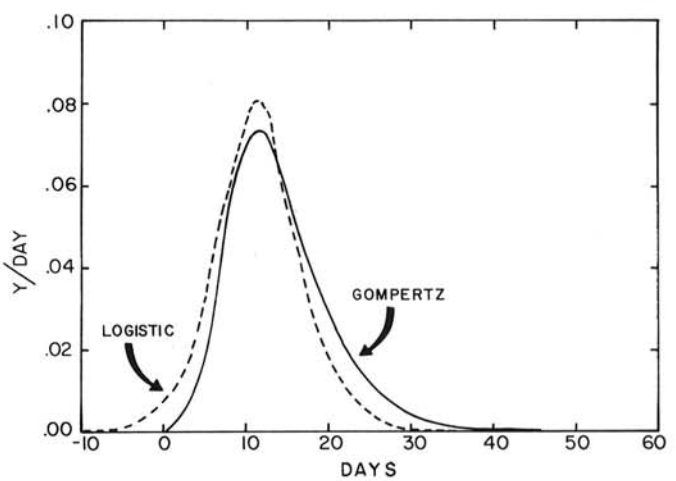

Fig. 4. The derivatives for two growth models superimposed with common points of inflection at day 12. The simulation parameters were y_0=0.0004; k=0.2 (Gompertz model) and r=0.325 (logistic model).

goodness of fit for over 100 disease progress curves. For the *Puccinia recondita*-wheat pathosystem, 80 epidemic curves were obtained from a fungicide interval experiment conducted in Wageningen, The Netherlands, in 1974 (R. D. Berger and J. C. Zadoks, *unpublished*). To show the broad applicability of the Gompertz model, additional disease progress curves were selected from the literature (Table 2).

The disease proportions for all curves were treated with the appropriate transformations (equations 2 and 4). The slope values, intercepts, and tests of goodness of fit were determined by simple linear regression. Alternatively, the least squares program for nonlinear models (NLIN procedure of the Statistical Analysis System [SAS] [4]) was used to obtain parameter estimates and goodness of fit for both the logistic and Gompertz models. Computing was done utilizing the facilities of the Northeast Regional Data Center of the State University System of Florida in Gainesville. The NLIN procedure has optimum estimation of parameters only when numerous ($n>20$) observations are available for analysis (4). The maximum number of y values available for a pathosystem was 17; the average number was 12.

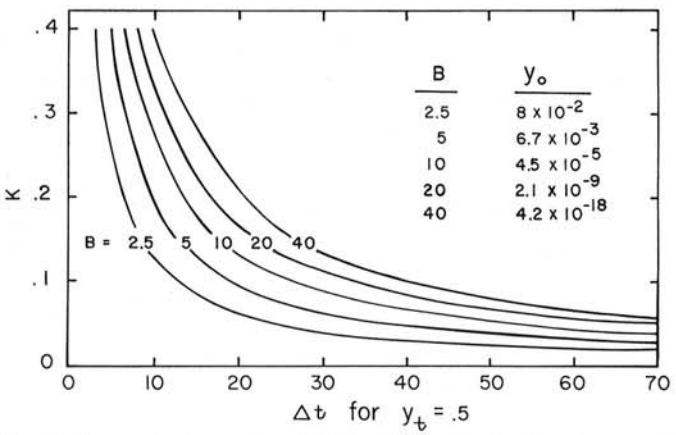

Fig. 5. Time required to reach $y=0.5$ from different levels of y_0 when growth increases according to the Gompertz model. The delay in time (Δt) to reach $y=0.5$ from reduction in y_0 can be found by comparing lines across any rate in the range $0<k<0.4$.

"Model" epidemics were generated by insertion of either model (equation 1 or 3) into a computer program using the CSMP (Continuous System Modeling Program) language. Plots of continuous growth, the derivatives, and transformed values were easily obtained for any of the desired levels for parameters y_0, k, and r.

The Gompertz model provided a better statistical fit than did the logistic model for all 113 disease progress curves of the nine pathosystems. In general, the fits were slightly better for both models with the simple linear regression technique (*unpublished*). For all curves, curvilinearity existed in the logistically transformed values. That is, the individual residual sums of squares were negative for early and late disease values and positive for values in the middle of the transformed curve. With the Gompertz model, this pattern of same signs for consecutive residual sums of squares was largely avoided, thus giving evidence for more balanced fit.

Disease progress following simulated sanitation. The time needed to reach a specific disease severity is sometimes used as a parameter to compare epidemics. The value, $y=0.5$, was arbitrarily selected to compare model epidemics begun at different initial disease severities and over a range of epidemic rates (Fig. 5). Many leaf spot diseases have a Gompertz rate of $k<0.1$; the average rate

TABLE 1. Plant disease proportions expressed as "gompits" (the Gompertz transformation)[a,b]

	Hundredths									
Tenths	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	···	−1.527	−1.364	−1.255	−1.169	−1.097	−1.034	−0.978	−0.927	−0.879
0.1	−0.834	−0.792	−0.752	−0.713	−0.676	−0.640	−0.606	−0.572	−0.539	−0.507
0.2	−0.476	−0.445	−0.415	−0.385	−0.356	−0.327	−0.298	−0.270	−0.241	−0.213
0.3	−0.186	−0.158	−0.131	−0.103	−0.076	−0.049	−0.021	+0.006	+0.033	+0.060
0.4	+0.087	0.115	0.142	0.170	0.197	0.225	0.253	0.281	0.309	0.338
0.5	0.367	0.395	0.425	0.454	0.484	0.514	0.545	0.576	0.607	0.639
0.6	0.672	0.705	0.738	0.772	0.807	0.842	0.878	0.915	0.953	0.991
0.7	1.031	1.072	1.113	1.156	1.200	1.246	1.293	1.342	1.392	1.445
0.8	1.500	1.557	1.617	1.680	1.747	1.817	1.892	1.971	2.057	2.150
0.9	2.250	2.361	2.484	2.623	2.783	2.970	3.199	3.491	3.902	4.600

[a] The transformation is gompit (y) = −ln (−ln (y)) in which y=disease proportion in the range $0<y<1$.
[b] To find y for gompits not in the table use y = exp (−exp (−gompit (y))).

TABLE 2. Comparisons of the statistical parameters of 113 epidemics representing nine pathosystems fitted to two growth models[a]

				Logistic model			Gompertz model		
Pathosystem	Source	No. of curves	N[b]	Rate[c]	s_{yx}[d]	r[e]	Rate	s_{yx}	r
Puccinia recondita—wheat	···	80	14	0.07–0.12	0.02–0.16	0.843–0.976	0.025–0.046	0.01–0.06	0.969–0.991
Helminthosporium maydis—corn	(3)	3	7	0.10–0.11	0.05–0.06	0.954–0.975	0.051–0.058	0.04–0.05	0.980–0.988
Ceratocystis ulmi—elm	(25)	1	13	0.04	0.01	0.974	0.008	<0.01	0.992
Ceratocystis ulmi—elm	(20)	1	17	0.05	0.05	0.857	0.009	0.028	0.982
Puccinia striiformis—wheat	(26)	1	5	0.11	0.04	0.969	0.042	0.02	0.983
Venturia inaequalis—apple	(1)	1	10	0.19	0.01	0.580	0.018	<0.01	0.861
Uromyces sp.—yucca	(19)	3	5	0.05–0.06	0.01	0.841–0.960	0.008–0.009	<0.01	0.972–0.980
Erysiphe graminis—wheat	(14)	4	5	0.18–0.24	0.01	0.757–0.915	0.034–0.042	<0.01	0.983–0.995
Cercospora apii—celery	(7)	7	5	0.10–0.11	0.08	0.660–0.903	0.024–0.03	0.02–0.09	0.969–0.984
Cercospora apii—celery	(8)	11	6	0.07–0.15	0.01–0.11	0.886–0.928	0.015–0.057	<0.01	0.986–0.997
Helminthosporium turcicum—corn	(6)	1	8	0.13	0.01	0.921	0.03	<0.01	0.992

[a] Disease progress curves fitted using the NLIN procedure of SAS. The model equations were $y = 1/(1 + \exp(-[a + rt]))$ for the logistic and $y = \exp(-B\exp(-kt))$ for the Gompertz.
[b] Number of observations per curve.
[c] Rate for the logistic model is the apparent infection rate (r) of Vanderplank (24) and for the Gompertz model it is the parameter k.
[d] s_{yx} is the standard error of the estimate, a measure of the variability of y values about the regression line, calculated as (residual sums of squares/$(N\text{-}2)$)$^{1/2}$ (reference 13).
[e] Correlation coefficients.

for the 113 curves in Table 2 was k=0.028. The anticipated benefit from sanitation measures could alter the B parameter from, eg, 5 (y_0=0.0067) to 10 (y_0=0.000045). The likely delay in epidemic time to reach y=0.5 then would be about 25 days (at k=0.028). It would take drastic sanitation measures to gain more reduction in initial disease and achieve a corresponding delay in time to reach y=0.5. Sanitation measures are relatively ineffective at delaying fast-moving epidemics (k>0.1).

If epidemics were to proceed logistically, the delay in time would be much longer than with the Gompertz model. With Vanderplank's (24) sanitation ratio, 50 days delay would be expected with the above sanitation values when r=0.1 (the average r value from Table 2). In actual epidemics, it is unlikely that this long delay would occur (10,23); ie, the epidemics proceed faster than projected by the logistic model.

DISCUSSION

The linearization of disease progress curves is essential to determine epidemic speed, to project future disease, and to estimate initial disease. The logistic transformation has severe limitations for all three desired parameters for the many disease progress curves that are asymmetrical. The Gompertz model was superior to the logistic model in linearizing the 113 selected disease curves. Analytis (2) and Hau and Kranz (16) also used the Gompertz and other models to linearize disease progress curves. They stressed using the correct model to make the linearization, if a transformation was indeed necessary. Symmetrical curves may be effectively linearized by the logistic equation. Therefore, it seems wise to heed Kranz' (18) caution to examine the underlying distribution of values before selecting a transformation. If the inappropriate model is chosen, inaccurate estimates of the epidemic parameters result. In this vein, the curves of Fried et al (Fig. 3 in reference 14) are of particular interest. They used the intercept of the simple linear regression of logit values as their estimate of y_0. The Gompertz model provided estimates of y_0 that were 10- to-100-fold lower than the logistic. It may be more appropriate to estimate y_0 with the logistic model by extending a curve to the y axis that superinscribes the initial logit values as in my Fig. 3. Fried et al (14) also claimed a better fit of the logistic over other models (including Gompertz') but provided no evidence to support this judgement. In my analysis of the same values, the Gompertz model had a slightly better statistical fit for all four curves.

The statistical techniques used to estimate parameters must be employed with some caution (13,23). Consideration must be given to patterns of values, particularly to outliers as these contribute excessively to the correlation coefficients (Fig. 6.3 of reference 13). Additionally, the range of observed values should not be restricted (Fig. 6.6 of reference 13) as this will also affect the correlation coefficient.

Both the simple linear regression and nonlinear curve-fitting techniques have optimum estimation of parameters when many (n>20) values are available. Since numerous values aid in interpretation of the disease progress curves, researchers should be encouraged to begin estimation early in the epidemic and continue to make estimates at frequent intervals. In this way, more reliable estimates of the epidemic parameters will be obtained.

LITERATURE CITED

1. Analytis, S. 1973. Zur Methodik der Analyse von Epidemien dargestellt am Apfelschorf (*Venturia inaequalis* (Cooke) Aderh.). Acta Phytomed. 1:1-76.
2. Analytis, S. 1979. Die transformation von Befallswerten in der quantitativen Phytopathologie. II. Das Linearisieren von Befallskurven. Phytopathol. Z. 96:156-171.
3. Ayers, J. D., Nelson, R. R., Castor, L. L., and Blanco, M. H. 1976. Yield losses in corn caused by *Helminthosporium maydis* Race T. Plant Dis. Rep. 60:331-335.
4. Barr, A. J., Goodnight, J. H., Sall, J. P., and Helwig, J. T. 1976. A User's Guide to SAS 76. SAS Institute Inc., Raleigh, NC. 329 pp.
5. Batschelet, E. 1976. Introduction to Mathematics for Life Scientists. Springer, New York. 643 pp.
6. Berger, R. D. 1973. *Helminthosporium turcicum* lesion numbers related to numbers of trapped spores and fungicide sprays. Phytopathology 63:930-933.
7. Berger, R. D. 1973. Infection rates of *Cercospora apii* in mixed populations of susceptible and tolerant celery. Phytopathology 63:1161-1165.
8. Berger, R. D. 1975. Disease incidence and infection rates of *Cercospora apii* in plant spacing plots. Phytopathology 65:485-487.
9. Berger, R. D. 1975. Rapid disease progress in early epidemic stages. (Abstr.) Proc. Am. Phytopathol. Soc. 2:35.
10. Berger, R. D. 1977. Application of epidemiological principles to achieve plant disease control. Annu. Rev. Phytopathol. 15:165-183.
11. Berger, R. D. 1981. The Gompertz transformation—more appropriate than the logistic to describe disease progress. (Abstr.) Phytopathology 71:203.
12. Berger, R. D., and Mishoe, J. W. 1976. CSMP simulation of several growth functions to describe epidemic progress. (Abstr.) Proc. Am. Phytopathol. Soc. 3:217.
13. Edwards, A. L. 1976. An Introduction to Linear Regression and Correlation. W. H. Freeman, San Francisco. 213 pp.
14. Fried, P. M., MacKenzie, D. R., and Nelson, R. R. 1979. Disease progress curves of *Erysiphe graminis* f. sp. *tritici* on Chancellor wheat and four multilines. Phytopathol. Z. 95:151-166.
15. Griggs, M. M., Nance, W. L., and Dinus, R. J. 1978. Analysis and comparison of fusiform rust disease progress curves for five slash pine families. Phytopathology 68:1631-1636.
16. Hau, B., and Kranz, J. 1977. Ein Vergleich verschiedener transformationen von Befallskurven. Phytopathol. Z. 88:53-68.
17. Jowett, D., Browning, J. A., and Haning, B. C. 1974. Nonlinear disease progress curves. Pages 78–114 in: J. Kranz, ed. Epidemics of Plant Diseases: Mathematical Analysis and Modeling. Springer, New York. 170 pp.
18. Kranz, J. 1974. The role and scope of mathematical analysis and modeling in epidemiology. Pages 7–54 in: J. Kranz, ed. Epidemics of Plant Diseases: Mathematical Analysis and Modeling. Springer, New York. 170 pp.
19. Larios, J. F., and Moreno, R. A. 1977. Epidemiología de algunas enfermedades foliares de la yuca en diferentes sistemas de cultivo. II. Roya y muerte descendente. Turrialba 27:151-156.
20. Miller, H. C., Silverborg, S. B., and Campana, R. J. 1969. Dutch elm disease: relation of spread and intensification to control by sanitation in Syracuse, New York. Plant Dis. Rep. 53:551-555.
21. Nair, K. R. 1964. The fitting of growth curves. Pages 119–132 in: O. Kempthorne, T. A. Bancroft, J. W. Gowen, and J. L. Lush, eds. Statistics and Mathematics in Biology. Hafner, New York. 632 pp.
22. Pennypacker, S. P., Knoble, H. D., Antle, C. E., and Madden, L. V. 1980. A flexible model for studying plant disease progression. Phytopathology 70:232-235.
23. Plaut, J. L. 1980. Epidemic progress of three pathosystems as affected by initial disease severity. M.S. thesis, Univ. of Florida, Gainesville. 106 pp.
24. Vanderplank, J. E. 1963. Plant Diseases: Epidemics and Control. Academic Press, New York. 349 pp.
25. Van Sickle, G. A., and Sterner, T. E. 1976. Sanitation: a practical protection against Dutch elm disease in Fredericton, New Brunswick. Plant Dis. Rep. 60:336-338.
26. Zadoks, J. C. 1961. Yellow rust on wheat, studies in epidemiology and physiologic specialization. Tidschr. Plantenziekten 67:69-256.
27. Zadoks, J. C., and Schein, R. D. 1979. Epidemiology and Plant Disease Management. Oxford, New York. 427 pp.