# Modeling the Temporal Primary Spread of African Cassava Mosaic Virus into Plantings

D. Fargette and K. Vié

ORSTOM, Laboratoire de Phytovirologie des Régions Chaudes, LPRC, CIRAD, BP 5035, 34032 Montpellier cedex 1, France.
We are grateful to J. Crawford, M. Jeger, and J. Mc Nicol for helpful discussions and to J. M. Thresh for constructive criticisms and detailed reviews of the manuscript.
This work was supported in part by grants from the Commission of the European Communities TSD-102 and TS2A-0137-C (CD).
Accepted for publication 11 January 1994.

## ABSTRACT

Fargette, D., and Vié, K. 1994. Modeling the temporal primary spread of African cassava mosaic virus into plantings. Phytopathology 84:378-382.

The rate of temporal primary spread of African cassava mosaic virus into cassava plantings has been shown to be dependent on the planting date, $P$, and on the plant age, $t$. In this paper, the relationships between the rate of disease progress, $P$, and $t$ were expressed mathematically. The appropriate functions were chosen, and their parameters were derived by nonlinear regression using a set of experimental data obtained at Adiopodoumé (Ivory Coast, West Africa). The resulting equations were incorporated into a monomolecular model with a variable rate $r_P$ (the product of the change of rate of disease incidence when $P$ was fixed), $k$, a constant, $y$, the disease incidence, and $t$, the time: $dy/dt = k \cdot r_P(t)(1 - y)$. The modeled disease progress curves were obtained by numerical integration of the differential equation. The close fit between the modeled and the experimental curves showed that the main trends of the epidemics were represented. The model was tested with a set of data obtained in Tanzania (East Africa), and the structure of the model was validated, as there was also a good fit between the observed and modeled disease progress curves. Finally, assumptions were made on the remaining variation around the modeled curves.

*Additional keywords*: epidemiology, geminivirus, modeling.

African cassava mosaic virus (ACMV) is caused by whitefly-transmitted geminiviruses that are widespread in Africa and perpetuated through infected cuttings (9,14). In previous studies, the primary spread of disease, when inoculum is introduced by viruliferous whiteflies from outside the field, was shown to predominate over secondary spreading within crops (6). Earlier analyses suggested that the course of ACMV epidemics over time resulting from primary spread is dependent on crop age and on planting date (7). In this paper we attempted to express mathematically the pattern of change of these factors with time, estimating its parameters from a set of experimental data obtained at Adiopodoumé (Ivory Coast, West Africa). The resulting functions were incorporated into a monomolecular differential equation with a variable rate—appropriate to describe epidemics resulting from primary spread (1)—which was solved numerically to obtain the modeled disease progress curves. The outputs were compared to the experimental curves to check whether the main trends of the natural epidemics were represented. Then, the model was tested using a set of data obtained in Tanzania (East Africa) (19). The modeled disease progress curves were compared to the observed curves in an attempt to validate the model structure. Finally, special attention was paid to the variation around each monthly modeled curve.



## MATERIALS AND METHODS

**Experimental trials.** All the data used for the model were derived from experiments performed at Adiopodoumé at the ORSTOM experimental farm, 20 km from Abidjan in the forested coastal region of Ivory Coast at 5° N, 4° W, and 20 m altitude. Between May 1981 and May 1986, 49 plantings of healthy cassava were made. In all experiments, fields were divided into blocks of 100 plants at 1 × 1 m spacing, and disease incidence was followed weekly or fortnightly by visual inspection of each plant. Each plot was isolated from infected cassava fields by a distance of at least several hundred meters. The details of the experimental trial were given earlier (7).

The data used to validate the model were obtained at Kiwanda, in Tanzania (19). The experiment started in March 1934, on land provided by the School of the Universities Mission to Central Africa at Kiwanda near Amani, 5° N and 38° E. Kiwanda lies at an altitude of 170 m, in a valley near the eastern foot of the East Usambara Mountains, 8 km from Amani. Forty-eight square plots of nine plants each were laid out. Each plot was surrounded by a row of cassava with ACMV. In each plot, one healthy cassava plant of the cultivar Mbarika was established at the beginning of each month. Records were taken at the beginning of each month for a period of 2 yr, and, as at Adiopodoumé, all plants that developed symptoms were removed, so that secondary spread was very limited (7,19).

**Methodology.** Several successive steps were followed to develop the model: 1) the main biological factors suspected to determine the ACMV disease progress curves were listed; 2) the functions to describe their rate of change with time were proposed; 3) the parameters of these functions were estimated from experimental data through nonlinear regression; 4) the relationships between these factors were proposed; 5) the functions and their interactions were inserted into a selected model; 6) a constant parameter was calculated; 7) the simulated curves were compared to those observed.

**Formulation of the model.** The model was composed of the following variables and functions: $t$ was the plant age in months with $t = 0$ at planting; $P$ was the planting date, recorded as a month number from 1 to 12, January being 1; $a(t)$ was the rate of disease progress due to the age effect; $s(t,P)$ was the rate of disease progress due to the season effect; $r(t,P)$ was the rate of disease progress due to the effects of both age and season.

**Disease incidence and cassava age.** Earlier studies (7) indicated that the change in rate of disease incidence with plant age exhibited the following trends: the rate was nil within the first month of planting; the maximum rate was reached 2 mo after planting; the rate decreased between 2 and 6 mo after planting following a curvilinear trend; and finally some additional spread still occurred in cassava stands older than 6 mo. On the basis of this information, we assumed that the monthly rate of increment in older plants was equal to that at the sixth month.

The disease incidence values to calculate $a(t)$ were taken exclusively in December. This period was chosen because sufficient healthy plants remained in plots of the different ages to assess the disease increment. Then, for each year of study, disease increments between the beginning and the end of December were calculated in plots planted in October (2-mo old), September (3-mo old), and so on until June (6-mo old). Values of $a(t)$ were calculated as:

$$a(t) = (y_t - y_{t-1})/[y_{2n}(1 - y_{t-1})] \quad (1)$$

where $y_t$ is the disease incidence for the age $t$, and $y_{2n}$ is the disease incidence 2 mo after planting for plots planted in November. This allowed $a(t)$ to vary from 0 to 1.

It was assumed that the changing rate of symptom expression with age could be represented by a function (close to the differential formulation of the Weilbull model), with $\alpha$, $\beta$, $\delta$, and $\epsilon$ the parameters to be estimated by nonlinear regression using the Systat V statistical software package (22):

$$a(t) = 0 \text{ when } t \leq 1$$
$$a(t) = \alpha(t - \beta)\exp[-\epsilon(t - \beta)] + \delta \text{ when } t > 1 \quad (2)$$

**Disease incidence and planting date.** Earlier results (7) indicated that the maximum rate of disease progress was reached 2 mo after planting and that the change of rate with time was seasonal. There was a linear relationship between the rate of symptom expression and temperature, and variation of temperature was approximated by a sinusoidal function, in general (13), and under Adiopodoumé conditions in particular (8).

Values to calculate $s(t,P)$ were taken 2 mo after planting when $t = 2$, where $s(2,P) = m_2$, and $m_2$ is the number of "multiple infection units" (sensu Gregory [12]) occurring during the first 2 mo of growth, applying the transformation $m = -\log_e(1 - y)$. It was assumed that the rate of symptom expression varies with season, following a sinusoidal function with a periodicity of 12 mo (with age in months, $P = 1$ for January):

$$s(2,P) = \sigma + \mu \sin[\omega(2 + P) + \phi] \quad (3)$$

where $\sigma$ is the average rate, $\mu$ is the amplitude around this average, $\phi$ is the phase, and $\omega = 2\pi/12$ (=0.52) is the annual periodicity after conversion to radians. If $\sigma + \mu \sin[\omega(2 + P) + \phi]$ was less than 0, $s(2,P)$ was set to 0.

**The model.** Age effect: $a(t) = \alpha(t - \beta)\exp[-\epsilon(t - \beta)] + \delta$. Season effect: $s(2,P) = \sigma + \mu \sin[\omega(2 + P) + \phi]$. It was generalized that

$$s(t,P) = \sigma + \mu \sin[\omega(t + P) + \phi] \quad (4)$$

For the interaction between age and season effects, it was assumed that the rate of disease progress was the net product of these two effects; then, when $P$ was fixed

$$r_P(t) = a(t)\, s_P(t) \quad (5)$$

The monomolecular model (with a time-dependent function) was appropriate to describe epidemics due to primary spread with a changing rate with time (1). Then:

$$dy/dt = k\, r_P(t)(1 - y) = ka(t)\, s_P(t)(1 - y), \quad (6)$$

$k$ being a constant

$$dy/dt = k\{\alpha(t - \beta)\exp[-\epsilon(t - \beta)] + \delta\}$$
$$\{\sigma + \mu \sin[\omega(t + P) + \phi]\}(1 - y) \quad (7)$$

This first-order differential equation with separable variables had an analytical solution (D. Fargette and K. Vié, *unpublished results*). However, it was routinely solved numerically with the Stella II package (18) using the fourth-order Runge-Kutta integration method with a step integration time of 1 wk.

The $k$ value was calculated by successive approximations for observed values $y_o$ to fit the calculated ones $y_c$ so that the linear regression between $y_o$ and $y_c$ had a high coefficient of determination, with a slope close to 1 and an intercept close to 0 (20).

**Estimate of the variation around monthly modeled curves.** For each month of planting, the age $T$ was calculated when 50% disease incidence was reached. From May to October $T$ was estimated to be around 150 days, and 70 days the other months. Then, for each month the lower ($k_m$) and upper ($k_M$) limits of $k$ values were calculated as:

$$k_m = \log(1 - y_{T,m})/\log(1 - y_{T,1})$$
$$k_M = \log(1 - y_{T,M})/\log(1 - y_{T,1})$$

where $y_{T,1}$ is the simulated value of disease incidence for $k = 1$ (modeled disease progress curve), $y_{T,m}$ is the lowest observed value of disease incidence, and $y_{T,M}$ is the highest observed value of disease incidence at age $T$.

**Validation of the model.** The following steps were followed to validate the structure of the model: 1) the set of parameters (age and season function parameters) was extracted from the experimental data obtained in Tanzania; 2) the $k$ constant was calculated as above; 3) the simulated curves derived from equation 7 (with the new set of parameters) were compared with the observed disease progress curves.

## RESULTS

**Rate of disease incidence and age.** Regression between cassava age and disease incidence (eq. 2) had a coefficient of determination, $R^2$, of 0.90. The estimates of the parameters were $\alpha = 3.01$ (0.74), $\beta = 1.07$ (0.03), $\epsilon = 1.37$ (0.03), and $\delta = 0.23$ (0.06), the figures in parentheses being the asymptotic standard errors. The regression curve represented the main trends of the change of spread with age, i.e., rapid spread apparent in the second month of planting followed by a curvilinear decrease, a lower asymptotic value being reached afterwards (Fig. 1). However, variation of the experimental points remained around the modeled curve, which was also reflected by the large residuals and standard errors of the parameters, with $\alpha$ especially.

**Rate of disease incidence and season.** Regression between disease incidence after 2 mo of growth and planting date (eq. 3) had an $R^2$ of 0.64, and the estimates of the parameters were $\sigma = 0.46$ (0.06), $\mu = 0.35$ (0.08), and $\phi = -0.66$ (0.22), the figures in brackets being the asymptotic standard errors. The main trends

of the change of symptom expression with season were represented, although variation remained around the modeled curves as indicated by the large residuals and the standard error of the estimates, with $\phi$ especially (Fig. 2).

**The model.** The wide range of shapes for $r_P(t)$ resulting from equation 7 is illustrated in Figure 3 (top). They ranged from unimodal curves for November to February plantings to pronounced bimodal curves between May and July. Despite the wide variation of rates, there was a consistent and rapid increase of $r_P(t)$ in the second month of growth followed by a negative exponential decrease within the six following months. For the November to February plantings, this "dampened" the sinusoidal season variation, which was then apparent only at the latter stages of epidemics, in May to July plantings especially. These curves obtained with a step integration time of 1 wk were not substantially different when a smaller integration time of 1 day was applied.

**Modeled and experimental curves.** The constant $k$ was set to 1.3. This value, when included in equation 7, gave a good fit between observed and calculated disease incidence, with $a = 0.70$, $b = 0.14$, and $R^2 = 0.71$. The modeled curves through equation 5 showed diverse shapes (Fig. 3, bottom), with much spread into plots planted between November and March and less spread into other plots. The modeled curves also represented the overall decreasing rate of spread in aging cassava. The initial delay of the epidemics, the exponential phase, and the asymptote after a plateau for May and June were adequately described. The sequence of high then low spreading of epidemics in plots planted from May to September, resulting from changes in inoculum pressure combined with a decreasing susceptibility to infection with age, was also described. Moreover, the actual values of disease incidence were also represented, as there was generally a good fit between modeled and actual disease progress curves (Fig. 4).

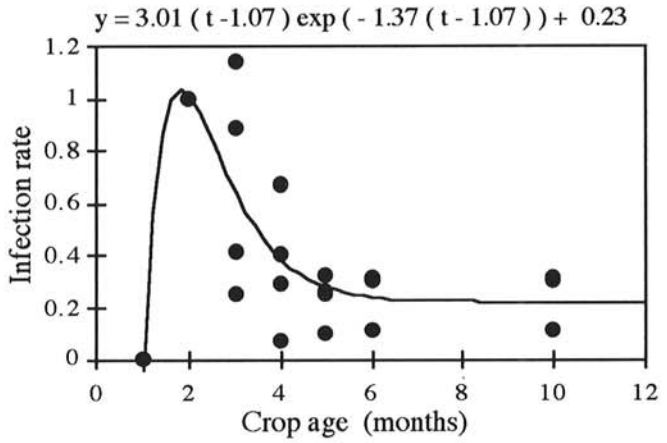


**Fig. 1.** Nonlinear regression between the rate of disease increment and cassava age.

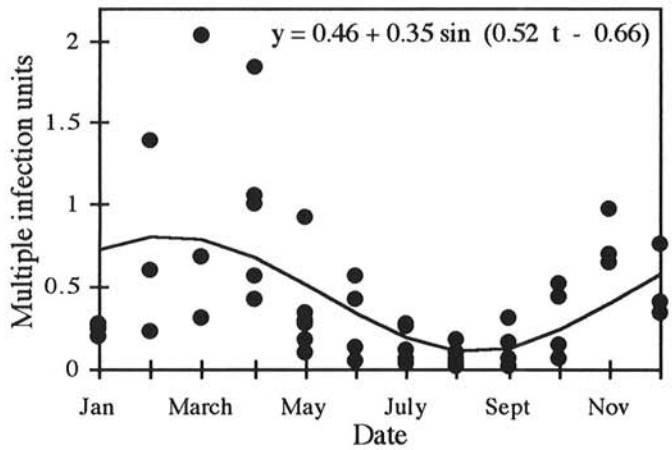


**Fig. 2.** Nonlinear regression between the rate of disease increment and planting date.

This suggested that the main characteristics of the ACMV epidemics were described by the model.

**Estimate of the variation around monthly modeled curves.** The general shapes of the disease progress curves were represented for each month. However, whatever the planting date, variation remained between the observed disease progress curve and the modeled curves. The extent of the variation was assessed through $k_m$ and $k_M$ values (Table 1). There was a trend to overestimation of the spread in December, January, June, and July and an underestimation in November. The only notable divergence between an observed and a calculated curve was in October 1982, a disease progress curve that was further atypical by differing substantially from the other observed curves.

**Validation of the model.** With the data obtained at Kiwanda, coefficient of determination $R^2$ of the regression between cassava age and the rate of disease incidence through equation 2 was 0.89. The estimates of the parameters were $\alpha = 2.63$ (0.78), $\beta = 1.08$ (0.04), $\epsilon = 1.31$ (0.28), and $\delta = 0.24$ (0.07) (figures in parentheses are the asymptotic standard errors). The coefficient of determination of the regression $R^2$ between symptom expression and planting date through equation 3 was 0.56, and the estimates of the parameters were $\sigma = 0.36$ (0.06), $\mu = 0.42$ (0.08), and $\phi = -0.62$ (0.20). Parameter $k$ was set to 2.5, so that the slope of the regression was 0.91, and the intercept was 0.07 with a coefficient of determination of 0.85.

This set of parameters was inserted into equation 7, and the simulated disease progress curves for Kiwanda were obtained by numerical integration. Figure 5 illustrates the observed and calculated disease progress curves. The main trends of the epidemics and the general pattern of variation of the curves between months were represented in the calculated curves whatever the planting date.

## DISCUSSION

The main advantage of analytical models is their clear structure, because they require few equations and only a few parameters

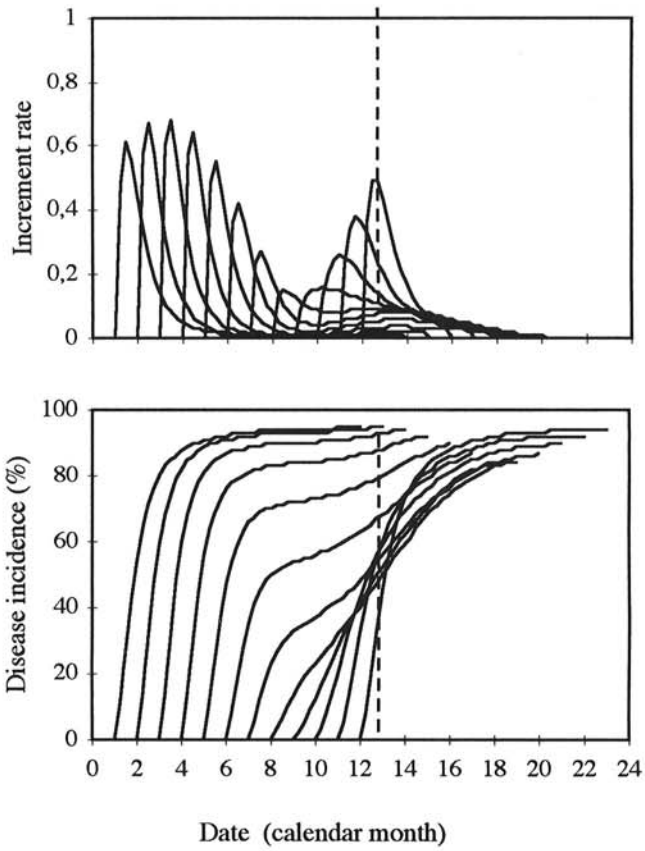


**Fig. 3.** Variation of the rate $r$ (top) and the disease incidence $y$ (bottom) with time for plantings of each calendar month.

that are easy to estimate and interpret (11,13). Such models have been developed successfully for several viral and nonviral plant epidemics (1,15,16,20,21). Model evaluation for ACMV was done in two steps, through verification and validation (2,20). Verification involved a comparison of the structure and general behavior of the model with the real system. With ACMV in Ivory Coast, simulated disease progress curves described the main trends of the epidemics. On the basis of these curves, we believe that the different steps in building the model were adequate: in particular, the biological hypotheses were correct, and the mathematical expressions to describe them and their interaction were appropriate. These modeling studies clearly showed that ACMV disease progress curves in Ivory Coast were mainly driven by a direct interaction between an overall negative exponential changing susceptibility with age and a sinusoidal, temperature-driven, seasonal fluctuation in amount of spread from outside sources.

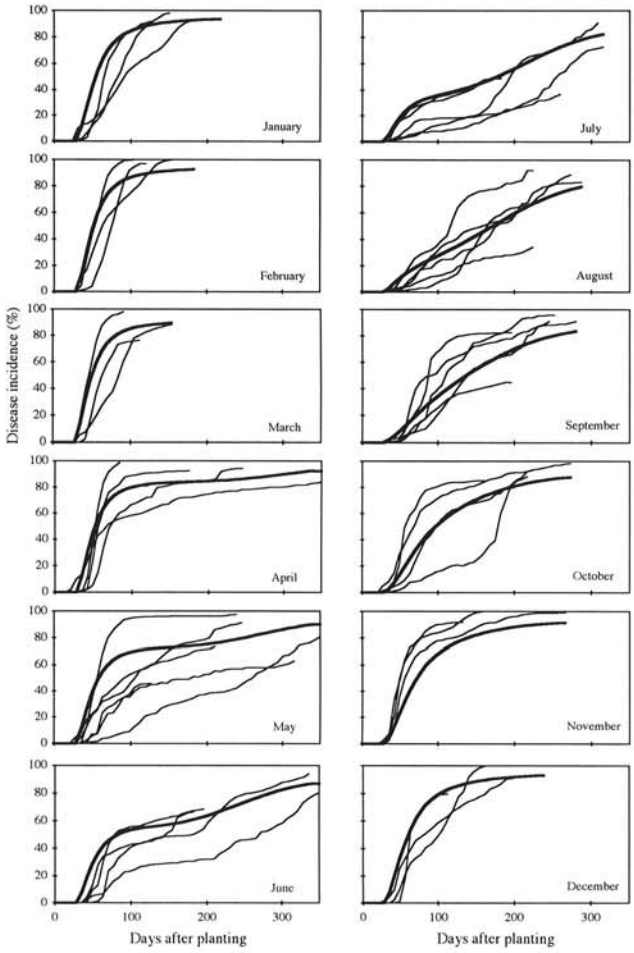


**Fig. 4.** Simulated disease progress curves (bold lines) and observed ones (thin lines) for each calendar month of planting at Adiopodoumé (Ivory Coast).

TABLE 1. Lower ($k_m$) and upper ($k_M$) limits of $k$ fitting respectively the lower ($y_m$) and higher observed disease incidence ($y_M$) at time $T$

| | | | Lower limits | | Upper limits | |
|---|---|---|---|---|---|---|
| Month of planting | $T$[a] | $y_{T,1}$[b] | $y_m$ | $k_m$ | $y_M$ | $k_M$ |
| January | 70 | 0.61 | 0.33 | 0.43 | 0.63 | 1.06 |
| February | 70 | 0.65 | 0.40 | 0.49 | 0.88 | 2.02 |
| March | 70 | 0.65 | 0.33 | 0.38 | 0.92 | 2.41 |
| April | 70 | 0.60 | 0.47 | 0.69 | 0.96 | 3.51 |
| May | 150 | 0.63 | 0.30 | 0.36 | 0.96 | 3.24 |
| June | 150 | 0.48 | 0.30 | 0.55 | 0.64 | 1.56 |
| July | 150 | 0.36 | 0.15 | 0.36 | 0.42 | 1.22 |
| August | 150 | 0.34 | 0.26 | 0.72 | 0.74 | 3.24 |
| September | 150 | 0.45 | 0.36 | 0.75 | 0.82 | 2.87 |
| October | 150 | 0.60 | 0.23 | 0.29 | 0.85 | 2.07 |
| November | 70 | 0.39 | 0.58 | 1.76 | 0.73 | 2.65 |
| December | 70 | 0.52 | 0.33 | 0.55 | 0.62 | 1.32 |

[a] In number of days after planting.
[b] Simulated value of disease incidence for $k = 1$ at time $T$.

Results obtained by others at Kiwanda, Tanzania, in a different country and at a different time, were used to validate the model (19). Secondary spread, as at Adiopodoumé, was limited by the systematic removal of all new infections as they occurred, so that the epidemics resulted from primary spread. Ecological conditions (range of temperature, longitude, latitude, altitude, etc.) and experimental setup (cultivar used, experimental design, etc.) differed greatly between Kiwanda and Adiopodoumé (see Materials and Methods); however, the simulated disease progress curves adequately described the observed data. Also, the structure of the model setup at Adiopodoumé was validated, because the main trends of the epidemics observed at Kiwanda were adequately represented. Based on that validation, we think that the model was robust and adequately described ACMV epidemics in a wide range of conditions.

Decreasing susceptibility to virus infection with age is likely to be a general feature of cassava. Parameters of the function describing the rate of incidence and age were very close between Adiopodoumé and Kiwanda, suggesting that the relationship was valid for a wide range of conditions (variety, site, year, etc.). Seasonal variation in spread is likely to be found commonly, as it has been shown that temperature is a driving force in epidemics of several whitefly-transmitted geminiviruses (8). However, differences in the parameter values of the sinusoidal function suggest that the parameters should be recalculated in different experimental conditions (site, variety, year, etc.).

Various sources of variation are likely to explain the differences between disease progress curves in cassava plots planted on the same date in different years. The site and the position of the trial site and orientation in relation to prevailing wind (19) affect

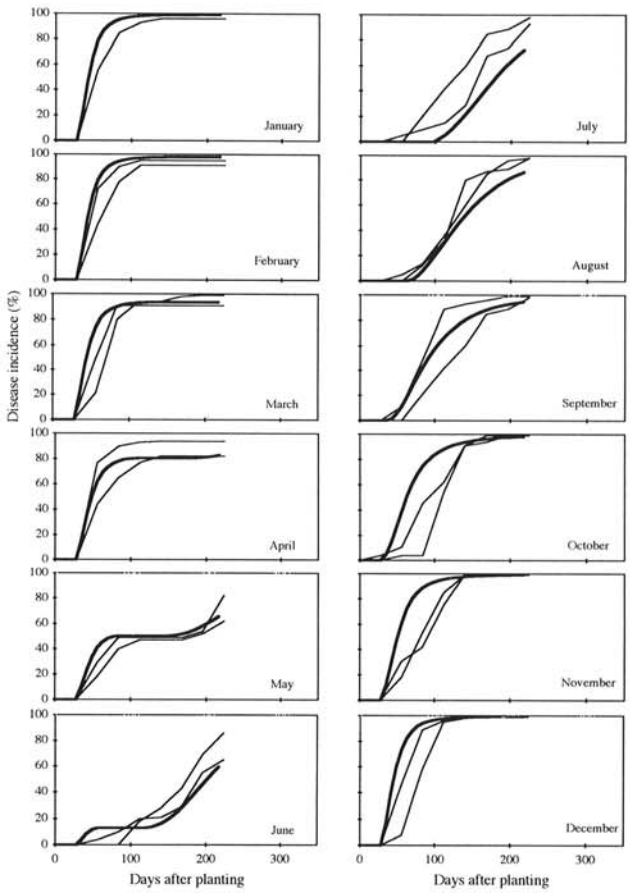


**Fig. 5.** Simulated disease progress curves (bold lines) and observed ones (thin lines) for each calendar month of planting at Kiwanda (Tanzania).

the course of the epidemics. There is also variation in the climatic conditions from year to year (3,17); moreover, whitefly numbers exhibit weekly variation in addition to monthly trends (4,10). Location of infection sources is also likely to be critical. There was less year-to-year variation between epidemics in Tanzania than in Ivory Coast, likely reflecting the homogenous source of inoculum created by planting the guard row at Kiwanda, whereas inoculum source was not controlled at Adiopodoumé (see Material and Methods).

There is a tension in modeling between the pursuit of complexity and simplicity, and the degree of sophistication of a model should be dictated by the overall objectives (2,11). Our current objectives included the selection of parameters to describe secondary spread within plantings and to express the different components of resistance. Indeed, a preliminary simulator of epidemics has already been built and has been used to gain a better understanding of the epidemics and as a tool to assess strategies to control ACMV (5).

## LITERATURE CITED


1. Campbell, C. L., and Madden, L. V. 1990. Introduction to Plant Disease Epidemiology. Wiley Interscience, New York.
2. Doucet, P., and Sloep, P. B. 1992. Mathematical Modelling in the Life Sciences. Ellis Horwood, New York.
3. Eldin, M. 1971. Le milieu naturel en Côte d'Ivoire. Mem. ORSTOM 50:75-108.
4. Fargette, D. 1987. Epidémiologie de la mosaique africaine du manioc en Côte d'Ivoire. Ph.D. thesis. Université des Sciences et Techniques du Languedoc, ORSTOM, Paris. 203 pp.
5. Fargette, D., Fauquet, C. and Fishpool, L. D. C. 1992. Strategies for controlling African cassava mosaic virus: An approach through modelling. Proc. Association of Applied Biologists/British Society of Plant Pathology (AAB/BSPP) Mtg., York, England, 9–10 April 1992.
6. Fargette, D., Fauquet, C., Grenier, E., and Thresh, J. M. 1990. The spread of African cassava mosaic virus into and within cassava fields. J. Phytopathol. 130:289-302.
7. Fargette, D., Jeger, M. Fauquet, C., and Fishpool, L. D. C. 1994. Analysis of temporal disease progress of African cassava mosaic virus. Phytopathology 84:91-98.
8. Fargette, D., Muniyappa, V., Fauquet, C., N'Guessan, P., and Thouvenel, J.-C. 1993. Comparative epidemiology of three whitefly-transmitted geminiviruses. Biochimie 75:547-554.
9. Fauquet, C., and Fargette, D. 1990. African cassava mosaic virus: Etiology, epidemiology, and control. Plant Dis. 74:404-411.
10. Fishpool, L. D. C., Van Helden, M., Van Halder, I., Fauquet, C., and Fargette, D. 1988. Monitoring *Bemisia tabaci* populations in cassava: Field counts and trap catches. Pages 64-76 in: Proc. Int. Sem. African Cassava Mosaic Disease, 4-8 May 1987, Yamoussoukro, Côte d'Ivoire. Centre Technique de Cooperation Agricole et Rurale (CTA), Wageningen, Netherlands.
11. Gilligan, C. A. 1991. Modelling of crop pathogens. The Art and Craft of Modelling in Applied Biology. Aspects Appl. Biol. 26:111-121.
12. Gregory, P. H. 1948. The multiple transformation. Ann. Appl. Biol. 35:412-417.
13. Hau, B. 1990. Analytical models of plant disease in a changing environment. Annu. Rev. Phytopathol. 28:221-245.
14. Hau, B., Eisensmith, S. P., and Kranz, J. 1985. Construction of temporal models: II. Simulation of aerial epidemics. Pages 31-66 in : Advances in Plant Pathology 3: Mathematical Modelling of Crop Disease. C. A. Gilligan, ed. Academic Press.
15. Hong, Y. G., Robinson, D. J., and Harrison, B. D. 1993. Nucleotide sequence evidence for the occurrence of three distinct whitefly-transmitted geminiviruses in cassava. J. Gen. Virol. 74:2437-2443.
16. Jeger, M. 1986. The potential of analytical compared with simulation approaches in plant disease epidemiology. Pages 255-284 in: Plant Disease Epidemiology. Vol 1. Population Dynamics and Management, eds. K. J. Leonard and W. E. Fry. Macmillan, New York.
17. Monteny, B. A. 1988. Données climatiques recueillies à la station ORSTOM d'Adiopodoumé 1948-1987. ORSTOM, Paris.
18. Richmond, B., Peterson, S., and Boyle, D. 1990. Stella II User's Guide. High Performance System Inc., Hanover, NH.
19. Storey, H. H., and Nichols, R. F. W. 1938. Virus diseases of East African plants. VII. A field experiment in the transmission of cassava mosaic. East Afr. Agric. J. 6:446-449.
20. Teng, P. S. 1981. Validation of computer models of plant disease epidemics: A review of philosophy and methodology. J. Plant Dis. Prot. 88:49-63.
21. Teng, P. S. 1985. A comparison of simulation approaches to epidemic modelling. Annu. Rev. Phytopathol. 23:351-379.
22. Wilkinson, L. 1992. SYSTAT: The System for Statistics. Version 5.2. SYSTAT Inc., Evanston, IL.