Development of a Prediction Model for Papaya Ringspot in Veracruz, Mexico

GUSTAVO MORA-AGUILERA, Former Graduate Assistant, DANIEL NIETO-ANGEL, Graduate Assistant, and DANIEL TÉLIZ, Professor, Centro de Fitopatología, Colegio de Postgraduados, 56230 Montecillo, México; and C. LEE CAMPBELL, Professor, Department of Plant Pathology, North Carolina State Univesity, Raleigh 27695-7616

ABSTRACT

Mora-Aguilera, G., Nieto-Angel, D., Téliz, D., and Campbell, C. L. 1993. Development of a prediction model for papaya ringspot in Veracruz, Mexico. Plant Dis. 77:1205-1211.

A model to predict incidence of papaya ringspot was developed and validated from 5 yr of field observations in central Veracruz, Mexico. The model was developed from 1 yr of data collected from papaya (*Carica papaya*) plantations in two different locations in Veracruz during 1985–1986. Incidence of papaya ringspot was evaluated every 15 days, and viral infection was confirmed by ELISA. Aphid vectors (*Myzus persicae, Aphis gossypii, A. nerii, A. citricola,* and *Macrosiphum euphorbiae*) of papaya ringspot virus were collected every 3 days from Moericke yellow pan traps placed at each location. The prediction model was obtained from an examination of the matrix of Pearson's correlation coefficients and by simple and multiple regression analysis. Model selection was based on Mallow's C_p statistic, proportion of variance explained, variance inflation factor, analysis of structure, and predictive capacity. The largest amount of variation in the data was accounted for by model $\hat{y} = -1.45 + 0.42\ AN_5 + 0.00016\ PW + 0.116\ AG_5 - 0.0058\ AN_5^2 - 0.0057\ MP_5^2$, in which $\hat{y}$ was the incremental increase of disease ($\hat{y}_t - \hat{y}_{t-1}$) at any given time ($t$); AN_5, AG_5, and MP_5^2 were the numbers of the alate aphid species *A. nerii, A. gossypii,* and *Myzus persicae*, respectively. PW was an interaction variable defined as the product of precipitation (P) and speed and duration of wind from the north (W). Values for independent variables were accumulated during a 4-wk period that ended 3 wk before the calculated incremental increase of disease. The equation accounted for 78% ($R^2 \geq 0.78$) of the total variation of the change of disease incidence ($\hat{y}_t - \hat{y}_{t-1}$) in the original data set. Validity of this model was tested with data obtained from 60 epidemics in papaya plantations established from 1987 to 1989 to represent different dates, plant densities, and plantation sites. The model predicted the relative rate of disease increase in 38% of the epidemics (23 of 60, $R^2 \geq 0.60$). Three other models that accounted for less variance explained in the original data set than the first model ($R^2 < 0.78$) were also validated. One model predicted the incremental increase in disease incidence of 40% of the papaya ringspot epidemics (24 of 60) with $R^2 \geq 0.60$. In this model, the disease incidence change was explained by the independent variables AN_5, AG_5, and PW.

Papaya ringspot virus (PRV) disease is the most limiting factor in papaya (*Carica papaya* L.) production in Mexico (1,2,28,29). In some regions of southeastern Mexico, 90% of plants die from this disease in less than 1 yr and the normal productive life of a plantation may be reduced from several years to 1 yr. Cross-protection with a mild strain of PRV from Hawaii, different types of reflective material barriers, and application of vegetable oils (e.g., corn oil) have not provided satisfactory control of PRV on papaya in Mexico (2,28,29).

Accepted for publication 24 August 1993.

Studies relating vector numbers, weather conditions, and other factors to the spread of several diseases of plants caused by semipersistently and/or persistently (8,12,14,17,22,31) and nonpersistently (16,23,25,30) transmitted virus have been attempted. In some of these studies, forecasting systems were developed that resulted in partially (12) and very successful (22) aphicide application programs. For nonpersistenly transmitted viruses, such as papaya ringspot virus, least squares regression can be useful for identifying and selecting climatic and biological variables that explain the variability in disease incidence (16,23). Papaya ringspot has been studied with respect to several temporal and spatial epidemiological aspects (19). However, the biological and climatological factors related to the spread of PRV on papaya have not been studied extensively.

The objective of this work was to develop and validate a prediction model of papaya ringspot incidence based on counts of each aphid vector species and climatic factors by using data acquired during 1985–1989 at Veracruz, Mexico. Such models could provide an understanding of the most important biological and climatological factors involved in the spread of papaya ringspot epidemics. This information could be used as a guide to further studies and as a basis for future management strategies. Portions of this research have been reported (18,20,21).

MATERIALS AND METHODS

Establishment of field plantings for model development. Two plots of papaya were established at the Research Center for the Development of the Humid Tropics at Paso San Juan, Veracruz (19°10′ N, 98°16′ W). Site A (70 × 40 m) was nearly level and the soil was a clay loam. Site B (40 × 45 m), approximately 3 km from site A, also had a clay loam soil and was located on a hill with a slope of 3–4%. Fields at both sites had been cropped previously to maize. Seedbeds (30 × 1 m) were sown in June 1985 with papaya cv. Cera, the predominant cultivar grown in Veracruz. From sowing until transplanting, a cheesecloth cover was placed over the seedbeds to protect seedlings from aphids. In August 1985, virus-free seedlings were transplanted in each field with 3-m spacings between plants and rows. A total of 286 plants (22 × 13) and 168 plants (12 × 14) were transplanted to sites A and B, respectively.

Establishment of field plantings for model validation. Seventy-two additional plots of papaya cv. Cera were established in three fields also at the Research Center for the Development of the Humid Tropics at Paso San Juan, Veracruz. The field at site C was nearly level, the

field at site D was located on a hill with a slope of 3–4% approximately 3 km from site C, and the field at site E was nearly level and was located about 2 km from site C and D. Each of the three sites had clay loam soil, and maize was the previous crop.

A total of 2,000 papaya seedlings were grown from October 1987 through March 1988 in black plastic containers (18 × 15 cm) prior to planting. Papaya plantations were established at each site on six transplanting dates (beginning December 1987, then every 2 mo thereafter) with densities of 135, 84, 51, and 42 plants per plot (6.6 m × 40 m), to give an equivalent of 4,444, 2,500, 1,600, and 1,111 plants per hectare, respectively. Twenty-four plots were established per site. Standard cultural practices were used in all plots. Twelve plots, corresponding to the first three transplanting dates at site E, were damaged and lost due to heavy rain and strong winds.

Disease assessment. Each plant was observed once every 2 wk from transplanting in August 1985 until December 1986 (for plots used in model development) and from December 1987 until August 1989 (for plots used in model validation) for symptoms typical of PRV infections. Typical symptoms included a leaf mosaic and shoestring, watery spots on stem and leaf petioles and ring spots on fruit. The location of each individual diseased plant was recorded for each plot and evaluation date. Progress in time of virus incidence at each plot was considered as an epidemic. Two and 60 epidemics were assessed to generate and validate the model, respectively. Presence of PRV in about 30 symptomatic plants was verified by ELISA (27) each month from transplanting in August 1985 to August 1986 in the two plantations used for model development.

Collection of weather data. A meteorological station was located within 2 km of each of the five sites where disease assessments were made. Data on relative humidity, precipitation, and maximum and minimum temperatures were collected daily during 1985–1989. Wind speed and direction were calculated by the Gulf Foresight Center from data collected within 20 km of each experimental site. Land area between where wind measurements were made and the experimental area is a level coast without geographic barriers.

Aphid monitoring. Alate aphids were captured in Moericke yellow pan traps (30 × 55 × 13 cm) from August 1985 until December 1986 (for fields used in model development) and December 1987 until August 1989 (for fields used in model validation). One trap was placed midway along each edge of each field at all sites. Four traps were placed per field. Trap height was initially set at 0.30 m above the ground, then was adjusted up to 1.30 m as the crop grew to approximately 2 m. Aphids were removed from each trap every 3 days, placed in 70% ethyl alcohol in glass vials, and returned to the laboratory for counting and identification.

Selection of independent variables. Counts of each aphid vector species of PRV in Veracruz (10) (*Myzus persicae, Aphis gossypii, A. nerii, A. citricola,* and *Macrosiphum euphorbiae*) were utilized as independent biological variables. Only the number of alate members of these aphid species was used as estimator of PRV incidence because papaya plants are not colonized by migratory aphid species (19). The proportion of viruliferous aphids in vector populations was not determined.

Independent meteorological variables were mean rain (mm) and temperature (C) and mean wind velocity (m/sec) when from a north azimuth (regionally, the phenomenon is known as *norte*). Only north winds were included in model development because numbers of aphids trapped was correlated with frequency of north winds. In a previous analysis (G. Mora, *unpublished*) we concluded that the best representation of the wind variable is the product of *norte* duration (number of days) and mean wind velocity (m/sec). Single variables, products, and square values of meteorological variables were included in the analysis (Table 1). The period of time from inoculation with the five vectors found in Veracruz to first symptoms of papaya ringspot ranged from 6 to 38 days (mean of 25 days) and was dependent on vector species (10). Therefore, values of each biological variable were accumulated over intervals of 3 and 4 wk that ended 1, 2, or 3 wk before the reference time when predictions of the dependent variable (disease incidence) were made. Because biological variables were of paramount interest and to avoid reduction in residual degrees of freedom (due to high number of independent variables), the weather variables were accumulated only for a 4-wk period that ended 3 wk before the disease incidence assessments. These a priori constraints were set to identify the explanatory capacity of each independent variable. Independent variables used in this study included two classes of variables corresponding to environment and vectors (Table 1).

Selection of dependent variables. Dependent variables were incremental changes in percent disease incidence and the natural log (ln) transformation of incremental change in disease per unit time (Table 1).

Model development, selection, and validation. Multiple regression analysis was used to examine the influence of environment and aphid vector numbers (as independent variables) on dependent variables. Model development comprised three stages: 1) selection of the dependent variable and reduction of the number of independent variables, 2) selection of a set of potential predictive models, and 3) validation of those models with independent data. Because independent variables had different magnitudes and units of measurement (e.g., mm, C,) they were standardized to zero mean and unit variance to permit comparison of effects (9). Also, because incremental change of disease incidence ($Y_t - Y_{t-1}$, Table 1) was expressed as a percentage and thus had a discrete distribution, it was transformed to a square root to approximate a normal distribution. In stages one and two of model development, the stepwise procedure of SAS was used (9,24).

The proportion of variance accounted for by the model, Mallow's C_p statistic, the analysis of structure using eigenvalues, and variance inflation factor (VIF) were used as criteria to select regression models in stage two of model development. The proportion of variance accounted for [(total mean square − residual mean square)/(total mean square)] takes into account the number of independent variables in the model. Mallow's C_p was used as a criterion for the goodness-of-fit of regression equations with different numbers of independent variables. A model is less subject to bias when the C_p value is close to the

Table 1. Independent and dependent variables used to analyze the relationship between the change of disease incidence and environmental and biological variables for papaya ringspot in central Veracruz, Mexico

Variables	Symbol
Independent	
Vector[a]	
Aphis gossypii	*AG*(1–5)
A. nerii	*AN*(1–5)
A. citricola	*AC*(1–5)
Myzus persicae	*MP*(1–5)
Macrosiphum euphorbiae	*ME*(1–5)
Environmental[b]	
Mean minimum temperature (C)	*T*
Accumulated precipitation (mm)	*P*
Northern wind (days/m/sec)	*W*
Interaction and square values	*TP, TW, PW,* T^2, P^2, and W^2
Dependent[c]	
Disease increment	$Y_t - Y_{t-1}$
Rate of disease increase	$\ln (Y_t - Y_{t-1})/\Delta t$

[a] Numbers (1–5) represent accumulated number of aphids trapped during a 3-wk period ending 1 (1) and 2 (2) wk before the date of disease assessment and accumulated number of aphids trapped during a 4-wk period ending 1 (3), 2 (4), and 3 (5) wk before the date of disease assessment.

[b] Variables were accumulated (*P* and *W*) or averaged (*T*) during a 4-wk period ending 3 wk before the date of disease assessment; *W* = the product of northern wind's duration (days) and the mean of wind velocity (m/sec).

[c] *Y* = disease incidence (%); Δt = number of days between time interval *t* and *t-1*.

number of parameters (p) in the model (9,24). Draper and Smith (7) suggest that a regression model should have a C_p value about equal to the number of parameters. The VIF and analysis of structure measured the effect of multicollinearity among independent variables on the variances of estimated coefficients and, thereby, provided additional measures of model stability (9). If VIF is larger than 5, the associated coefficients may be poorly estimated (7,9). The analysis of structure using eigenvalues can be used to identify variables involved in linear dependence (9). With this methodology, relatively large values of variance proportions associated with small eigenvalues (close to 0) indicate that the independent variables are highly correlated (9). The coefficient of determination adjusted for degrees of freedom (R_a^2) evaluated the relative merit of including a specific independent variable within a model (9,14). The selected models were validated in the third stage by programming in SAS the equations, imputing data from 60 PRV epidemics monitored during 1987–1989, and regressing predicted and observed values. The target value for model reliability was set at $r^2 \geq 0.60$ ($p \geq 0.05$). None of these 60 epidemics was used to develop the initial prediction models.

Disease progress in time. The linearized forms of the monomolecular, Gompertz, and logistic models were evaluated for goodness-of-fit to data from the 60 epidemics by simple linear regression analysis (4,24). Goodness-of-fit was evaluated by examination of r^2 values. The objective of this analysis was to estimate the average rate of progress of each epidemic (4) (i.e., r_L = logistic, r_G = Gompertz, or r_M = monomolecular) and use this information to characterize epidemics that were and were not predicted in the validation process.

RESULTS

Preliminary selection of independent and dependent variables. Models to predict changes in disease incidence (Figs. 1 and 2) were developed in several stages. In the first stage, each dependent variable was regressed on 60 biological variables and nine environmental variables for each site (A and B) and for pooled data. The R^2 of the best equations indicated the level of precision obtainable for any given dependent variable. The R^2 values ranged from 0.40 to 0.43, 0.70 to 0.98, and 0.60 to 0.66 for pooled, site B, and site A data, respectively. Models with increment of disease development ($Y_t - Y_{t-1}$) as the dependent variable had the highest R^2 values. Square root transformation of this variable did not improve R^2 estimates, so only models with untransformed increments as dependent variable were evaluated further. Variables regularly selected in the stepwise procedure were AN_5, AG_5, and PW, where symbols and subscripts are those given in Table 1.

Model development and selection. In the second stage of model development, pooled data from sites A and B were used with the independent variables AN_5, AG_5, AN_5^2, PW, and MP_5^2, which had been selected in the first stage. The intercorrelation among independent and dependent variables was examined using the correlation matrix. The objective was to contrast these results with those using further stepwise procedure (9). All independent variables, except PW and MP_5^2 were correlated significantly with increment of disease incidence (i.e., $Y_t - Y_{t-1}$) (Table 2). Regression models of combinations of three-, four-, and five-variable regression variables were obtained with the stepwise procedure of SAS. The best model was selected based on statistics shown in Table 3. There were no one- or two-variable models with proportion of variance accounted for (R^2) greater than 0.50 and $C_p \leq p$. There were four models ($\hat{y}_4$ through $\hat{y}_7$) with a fairly high proportion of variance explained ($R^2 \geq 0.59$) and significantly estimated parameters ($p < 0.05$). Model $\hat{y}_7$ accounted for the highest proportion of variance (78%) and appropriate Mallows's C_p ($C_p = p$) (Table 3). Although models $\hat{y}_5$ and $\hat{y}_7$ had the highest collinearity between the variables AN_5 and

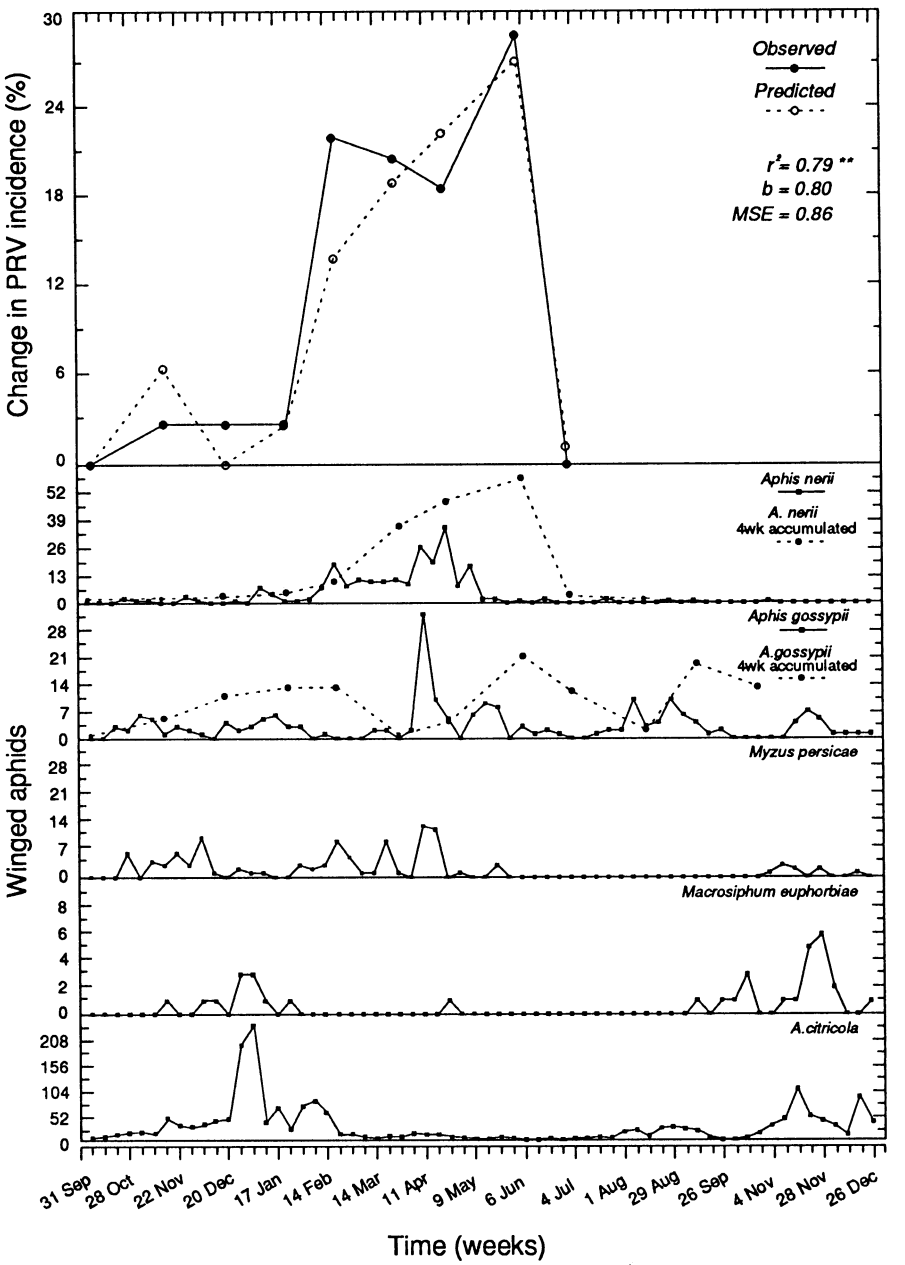

Fig. 1. Observed and predicted increment of papaya ringspot incidence ($Y_t - Y_{t-1}$) in papaya, using the model $\hat{y}_7 = -1.45 + 0.42\ AN_5 + 0.00016\ PW + 0.116\ AG_5 - 0.0058\ AN_5^2 - 0.0057\ MP_5^2$, and actual number of the aphid vectors of papaya ringspot virus *Aphis gossypii, A. nerii, A. citricola, Myzus persicae*, and *Macrosiphum euphorbiae* and accumulated number of individuals of *A. gossypii* (AG_5) and *A. nerii* (AN_5) during a 4-wk period that ended 3 wk before the date of disease assessment in site A, Veracruz, Mexico, 1985-1986.

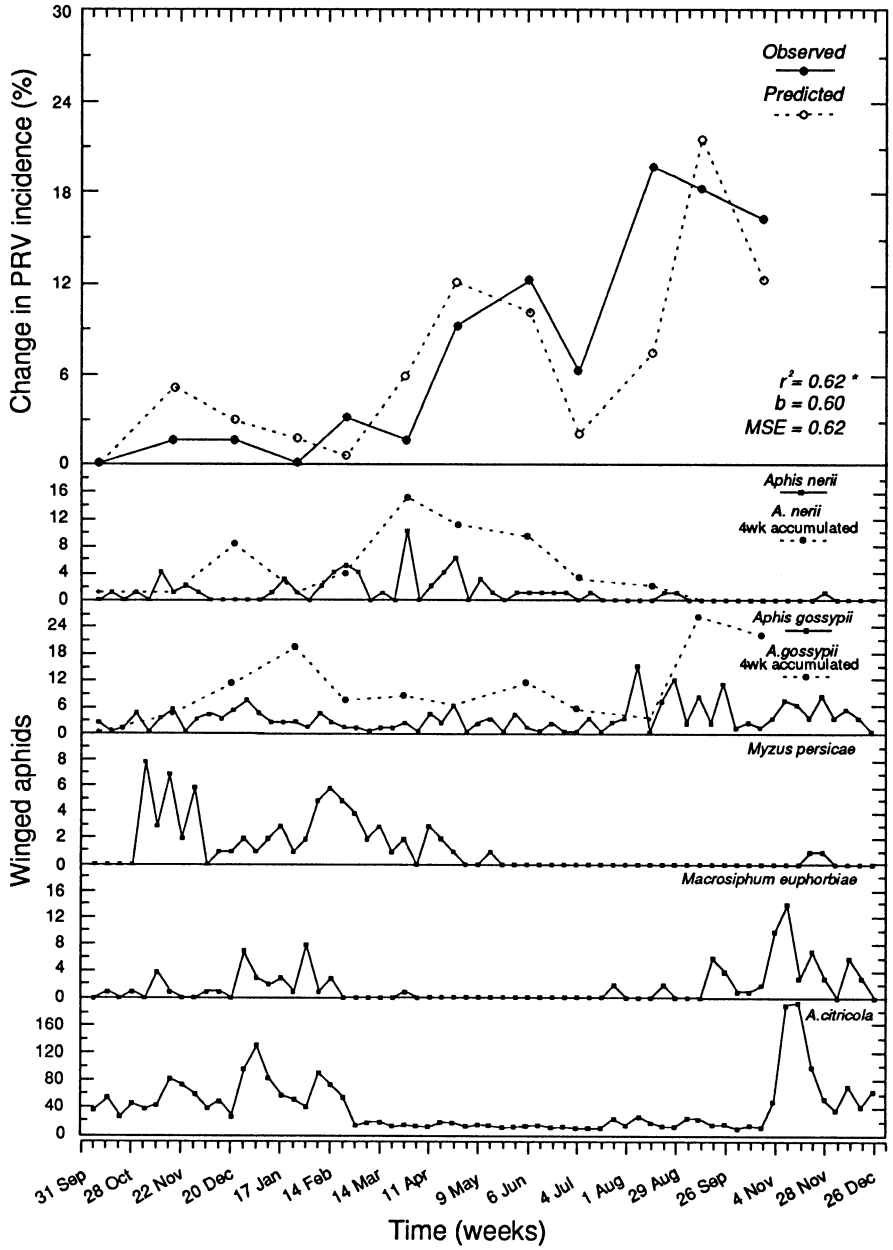

Fig. 2. Observed and predicted increment of papaya ringspot incidence ($Y_t - Y_{t-1}$) in papaya, using the model $\hat{y}_7 = -1.45 + 0.42\ AN_5 + 0.00016\ PW + 0.116\ AG_5 - 0.0058\ AN_5^2 - 0.0057\ MP_5^2$, and actual number of the aphid vectors of papaya ringspot virus *Aphis gossypii, A. nerii, A. citricola, Myzus persicae*, and *Macrosiphum euphorbiae* and accumulated number of individuals of *A. gossypii* (AG_5) and *A. nerii* (AN_5) during a 4-wk period that ended 3 wk before the date of disease assessment in site B, Veracruz, Mexico, 1985-1986.

Table 2. Correlation matrix of independent and dependent variables[a] used in regression analysis to relate incidence of papaya ringspot in central Veracruz, Mexico, to vector populations and weather

	AG_5	AN_5	AN^2_5	MP^2_5	PW
$Y_t - Y_{t-1}$	0.456 (0.049)[b]	0.554 (0.014)	0.541 (0.018)	0.037 (0.880)	0.313 (0.193)
AG_5		0.011 (0.963)	0.107 (0.645)	−0.162 (0.482)	0.110 (0.653)
AN_5			0.969 (0.001)	0.544 (0.012)	−0.439 (0.060)
AN^2_5				0.410 (0.065)	−0.322 (0.179)
MP^2_5					−0.725 (0.001)

[a] See Table 1.
[b] Values in parentheses are statistical significance level of the correlations.

AN_5^2 (Table 2), and also the model $\hat{y}_5$ had a $C_p > p$ ($C_p = 9.89$), the four models were used in further analysis.

Because the models were generated with pooled data, the models $\hat{y}_4$ through $\hat{y}_7$ were used for disease prediction at each individual site (A and B) prior to their validation. Predicted values from each model were then regressed on observed values. Model $\hat{y}_7$ had the best fit to disease progress at sites A and B ($r^2 = 0.62$ and 0.79, respectively; Figs. 1 and 2): $\hat{y}_7 = -1.45 + 0.42\ AN_5 + 0.00016\ PW + 0.116\ AG_5 - 0.0058\ AN_5^2 - 0.0057\ MP_5^2$, in which $\hat{y}_7$ represents the estimation of increments of disease as $y_t - y_{t-1}$ at any given time; AN_5 is the number of *A. nerii* trapped in a 4-wk period that ended 3 wk before the disease assessment; PW is an interaction variable of P and W, where P is the accumulated precipitation in a 4-wk period that ended 3 wk before the disease assessment and W is the product of *norte* duration and mean velocity of wind assessment in the same period as P; AG_5 is *A. gossypii* represented as described for AN_5; AN_5^2 is the squared value of AN_5; and MP_5^2 is *Myzus persicae*, also represented as the squared value of accumulated number of individuals of *M. persicae*. All coefficients were significantly different from zero.

However, because the populations of each vector species varied greatly between locations (Figs. 1 and 2) and because model $\hat{y}_7$ had the highest degree of multicollinearity (VIF = 33.89), it was not proposed as the best or final model for predicting PRV disease epidemics. To ascertain the most appropriate model, validation results were used as the primary criterion. Thus, the remaining three models ($\hat{y}_4$ through $\hat{y}_6$) were also validated. These linear models were: $\hat{y}_4 = 0.442 + 0.168\ AN_5 + 0.000092\ PW + 0.658\ AG_5$; $\hat{y}_5 = -1.34 + 0.32\ AN_5 + 0.00016\ PW + 0.1138\ AG_5 - 0.004\ AN_5^2$; and $\hat{y}_6 = 0.6 + 0.083\ AN_5 + 0.000085\ PW + 0.06\ AG_5 - 0.0014\ MP_5^2$, where $\hat{y}_i$, AN_5, PW, AG_5, AN_5^2, and MP_5^2 are as given in the $\hat{y}_7$ model. Also, all coefficients were significantly different from zero.

Model validation. Validation was the third stage in development of a papaya ringspot prediction model. To evaluate the reliability of models $\hat{y}_4$ through $\hat{y}_7$, predicted values for each of 60 disease progress curves were regressed on observed values by using the four models. The numbers of epidemics that were predicted with $r^2 \geq 0.60$ ($p \leq 0.05$) were 24, 25, 27, and 23 with models $\hat{y}_4$, $\hat{y}_5$, $\hat{y}_6$, and $\hat{y}_7$, respectively (Fig. 3). Model $\hat{y}_7$, which originally was shown to explain the highest proportion of variance with an acceptable C_p (i.e., $C_p = p$) in the original data set, predicted about the same proportion of the 60 epidemics as the other three models. The four models accounted for a similar proportion of the

variance in the same epidemics, except in the first and fifth planting at site D, although in these cases models $\hat{y}_4$ and $\hat{y}_5$ and models $\hat{y}_6$ and $\hat{y}_7$ had similar fits to the disease increments.

The location and density of papaya plantations did not influence the accuracy of prediction of ringspot epidemics with the four models. However, the transplanting date apparently did affect model precision. The change in disease incidence during epidemics of the second (15 February) and sixth (15 November) transplanting dates at all sites (C, D, and E) and the third transplanting date (15 April) at site D were not well predicted by any model. Models $\hat{y}_4$ through $\hat{y}_6$ predicted with acceptable precision ($0.53 \leq r^2 \leq 0.86$; $p \leq 0.05$) epidemics from plantations (sites C, D, and E) transplanted 15 September, a date near that for plantations used in model development (sites A and B). The $\hat{y}_6$ model failed to predict disease increase adequately ($r^2 < 0.50$) in four of 12 epidemics in the same transplanting date.

Disease progress in time. Of the 60 epidemics, 32 were best described by the Gompertz model, seven by the monomolecular model, and 21 by the logistic model. For comparison purposes, the values of the rate of disease progress, r_M and r_L, were adjusted via use of the Richard's equivalent weighted rate parameter (rho) to be equivalent to rates for the Gompertz model (4). Epidemics not predicted by any model had an epidemic rate (r_G) that averaged 0.020 and 0.025 units per day for the second (15 February) and sixth (15 November) transplanting dates, respectively. In contrast, epidemics for the first (15 December), third (15 April), fourth (15 June), and fifth (15 September) transplanting dates had epidemic rate averages of 0.019, 0.012, 0.014, and 0.035 units day per day, respectively, and were predicted by the models.

DISCUSSION

Several models were developed in this study to describe the relationship between disease incidence and independent variables that represented population levels of aphid vectors (*Myzus persicae, Aphis gossypii, A. nerii, A. citricola*, and *Macrosiphum euphorbiae*) and environment (temperature, precipitation, and northern winds). The statistical methodology used for the papaya ringspot model followed the procedure outlined by Coakley et al (6) and Chuang and Jeger (5) in which a prediction model was developed and validated based on biological and weather variables. Four regression models ($\hat{y}_4$ through $\hat{y}_7$) provided adequate predictions ($r^2 \geq 0.59$) for approximately 40% of the epidemics developed on plantations that were established at different transplant dates for model validation (Fig. 3). We believe that an $r^2 \geq 0.59$ is acceptable in the description of the general behavior of epidemics of papaya ringspot because of the biological complexity of the system. This r^2 value is also acceptable because of several theoretical assumptions that were made in model development. These include the assumption of a uniform incubation period for transmissions with different aphid species and the assumption of static relationship between the virus and the vector even though the transmission efficiency of a vector species can change throughout the season (22,25).

Similar goodness-of-fit of the four models to the 60 epidemics of papaya ringspot suggests that variables AN^2_5 and MP^2_5 were redundant and did not contribute significantly to the explanation of the increase in disease incidence. Collinearity due to these variables was also indicated by correlation matrix (Table 2), analysis of structure, and variance inflation factor (Table 3). We believe, therefore, that the simplest model ($\hat{y}_4$) acceptably predicts papaya ringspot epidemics (Table 3) under conditions similar to those in Veracruz, Mexico. This is especially true if plantations are established from August to September, as this period corresponds to the transplant date used to generate the models and because the epidemics in plantations established for validation purposes (15 September) were well predicted by model $\hat{y}_4$ ($r^2 \geq 0.60$). Model $\hat{y}_4$ also had desirable statistical properties ($C_p \approx p$, VIF < 5, acceptable proportion variance accounted for and acceptable structure) as well as positive intercept and nonquadratic variables that make this model potentially more biologically meaningful. The biological and climatological variables used in the simplest model (AN_5, AG_5, and PW) could be the factors most important to

Table 3. Coefficient of determination not corrected (R^2) and adjusted for degrees of freedom (R_a^2), Mallow's C_p, and variance inflation factor (VIF) for models to predict incidence of papaya ringspot in central Veracruz, Mexico, from different combinations of selected independent variables based on pooled data from sites A and B

Model	Independent variables[a]	R^2	R_a^2	C_p[b]	VIF[c]
$\hat{y}_1$	AG_5	0.21	0.16	1.91	1.0
$\hat{y}_2$	AN_5	0.31	0.26	10.99	1.0
$\hat{y}_3$	AG_5, AN_5	0.49	0.42	4.42	1.0
$\hat{y}_4$	AG_5, AN_5, PW	0.59	0.51	3.65	1.1
$\hat{y}_5$	AG_5, AN_5, AN^2_5, PW	0.69	0.58	9.89	23.1
$\hat{y}_6$	AG_5, AN_5, MP^2_5, PW	0.61	0.50	4.00	1.5
$\hat{y}_7$	AG_5, AN_5, AN^2_5, MP^2_5, PW	0.78	0.69	6.00	33.8

[a] See Table 1.
[b] Model is less subject to bias when the C_p value is similar to the number of parameters in the model.
[c] Factor larger than 5 indicates poor estimation of model coefficients.

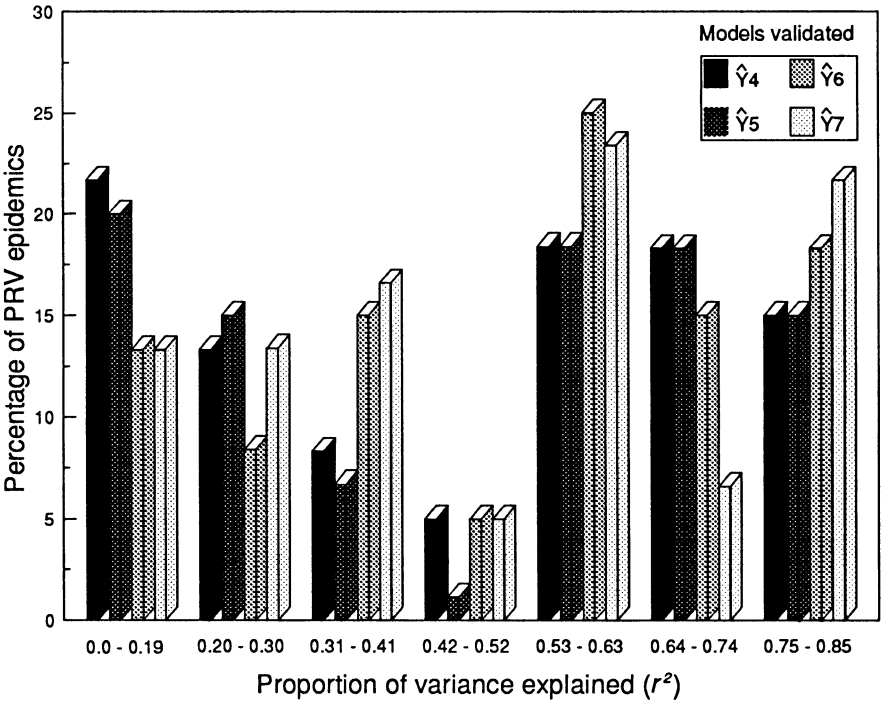

Fig. 3. Frequency distribution of the value of the coefficient of determination (r^2) obtained in the validation of four multiple-regression models with 60 epidemics of papaya ringspot in papaya assessed from 1987 to 1989 in Veracruz, Mexico.

describe the general behavior of a papaya ringspot epidemic in central Veracruz, Mexico (Figs. 1 and 2). The importance of *A. gossypii* as a vector of PRV has been reported in other countries (3). However, the epidemiological importance of *A.nerii*, northern winds (*nortes*), and precipitation has not been reported previously. These environmental conditions would probably have a more direct effect on the dynamics of aphid populations than on the variability of the virus incidence. For example, wind speed was correlated with numbers of *A. citricola* trapped ($r = 0.82$; $p = 0.03$).

Several specific aspects must be considered to understand the validity of model $\hat{y}_4$. First, estimated parameters (i.e., β_1 through β_3) associated with vector populations represent several factors such as vector behavior and plant resistance (16) and probably describe "vector intensity" sensu Irwin and Ruesink (13). Thus, parameters do not solely estimate the proportion of viruliferous individuals required in prediction models (17,22,25,31). Second, it is possible that vector(s) not selected in model $\hat{y}_4$ may be important in the dispersion of PRV under some environmental conditions. This possibility was shown by Madden et al (14) with several vectors of maize dwarf mosaic virus.

It is interesting that the variable selected for vectors was in the form of 4-wk accumulated numbers of trapped insects in a period that ended 3 wk before disease assessment. Although transmissibility determined in the greenhouse may be different from transmissibility in the field (13), our results appear to have biological validity. García (10) demonstrated in greenhouse experiments that an average of 3 wk was necessary for symptoms of papaya ringspot to develop following inoculation with apterous viruliferous vectors. It is also interesting that *A. citricola*, the aphid most commonly trapped (Figs. 1 and 2), was not selected for inclusion in any model. This agrees with the findings of Sigvald (25) that the vector present in the greatest proportion is not necessarily the most important biotransmitter.

The models were developed from data collected during 1 yr, at two sites (A and B), at a specific density (1,111 plants per hectare), and a transplanting date (August, the regional practice) but validated with data collected from three sites at four densities and six transplant dates. The facts that about 60% of the epidemics were not predicted by any model (36 of 60 epidemics with model $\hat{y}_4$) and that all of these epidemics were associated with plantations transplanted at a date different from that for plantations used in model development indicate that transplanting dates influenced model precision more than the effects of location and planting density. Poor model precision for epidemics at some transplanting dates can not be attributed to differences in epidemic rates (r_G) alone. That is, there were not any obvious relationships among rates of disease progress for epidemics associated with epidemics predicted and not predicted. The further analysis of epidemics from the second and sixth transplanting dates for all sites and from the third date of site D, where the predictions failed, could provide ecological information to improve the model $\hat{y}_4$ and thus increase our understanding of how papaya ringspot epidemics develop. Perhaps more extensive data should be considered (5,6,14). Additional variables or changes in units of measure also could improve the predictive model. For example, model precision might be enhanced by use of degree-days or other physiological units instead of days as a representation of time or by considering only wind speeds below the threshold for aphid landing and feeding. Aphids do not have oriented flight at wind speeds of more than 1–2 m/sec (26), and such wind speeds were frequently recorded during our studies. Additionally, previous results (19) demonstrated that microenvironmental factors may have subtle effects on papaya ringspot epidemics. Weather data were not available for each specific location in our current study, and this lack of detailed information for each site may explain, in part, the poor prediction with all models for a number of epidemics, mainly those occurring in the second and sixth planting.

In studies regarding quantitative relationships between insect catches and viruses, several investigators concluded that any model or analysis of epidemics would be incomplete without the effect of vector dynamics being considered (16,17,22,23,31). This would be true especially if total insect number is used as an independent variable rather than vector species numbers (14,19,25). Our study confirmed the importance of including information on vector species as well as on environmental factors in models.

Our results with papaya ringspot confirm conclusions obtained from regression analysis of other virus diseases. Those studies suggested the possible role of certain aphid species in the spread of maize dwarf mosaic virus in maize (14) and tobacco etch virus and tobacco vein mottle virus in tobacco (23). The objective in those studies (14,23) were basically the same as ours, i.e., to determine the most important vectors involved in virus dissemination. However, those virus epidemics could not be completely characterized statistically because of their transmission type and/or the plant colonization capability of their vectors. Maize dwarf mosaic virus is tramsmitted persistently, and thus vectors have the potential to cause more than a single infection as they colonize the plants. Although tobacco etch virus and tobacco vein mottle virus are nonpersistently transmitted and, therefore, each vector probably causes a single infection, the vectors of these viruses can also colonize the crop. Colonization of plants by aphids indicates that it is not possible to separate specimens that have immigrated to the field prior to being caught in the trap from those moving within the field (13). Additionally, colonization implies that secondary dispersion may occur independently of type of transmission (15,26). Both the possibility of multiple infections and crop colonization are factors not desirable when applying the least squares regression method for model development.

Our data suggest that multiple regression can be applied to study the relationship between nonpersistently transmitted viruses and vector(s) in the absence of colonization. PRV and several viruses have these characteristics (11,19,30), and some aphids appear to be effective vectors only during host-plant selection (26). This technique is not intended to provide specific biological explanations of virus incidences (e.g., incubation period, transmission efficiency), because variables selected and parameters defined in stepwise regression do not necessarily represent specific biological phenomena (8). Nonetheless, we believe that multiple regression applied correctly can be used when a general description of the characteristics of an epidemic is required or to improve a simulation model (13). The level of understanding of basic components of the pathosystem, i.e., how viruses infect (in a complex or not), the transmission type, the possibility of colonization by vectors, and the species of vectors present, largely influences the actual success of this statistical approach in modeling the virus-vector system.

ACKNOWLEDGMENTS

We thank Centro Regional de Enseñanza, Capacitación e Investigación para el Desarrollo del Trópico Húmedo (CRECIDATH) and C. Avila, F. Durán, and E. García for technical assistance. We also thank the anonymous reviewers and the senior editor for their criticisms and suggestions. The financial support provided by USAID/Mexico while the first author was a visiting scientist in the Department of Plant Pathology at North Carolina State University during a portion of this study and by CONACYT/México is gratefully acknowledged.

LITERATURE CITED

1. Becerra, N. L. 1987. Dinámica poblacional y hospederas de áfidos vectores del virus de la mancha anular del papayo en Veracruz. (Resumen) XIV Congreso Nacional de Fitopatología.
2. Becerra, N. L. 1988. Ensayo de barreras de jamaica (*Hibiscus sabdariffa* L.) y preferencia al color de papayo (*Carica papaya*) por áfidos vectores del virus de la mancha anular del papayo. (Resumen) XV Congreso Nacional de Fitopatología.
3. Blackman, R. L., and Eastop, V. F. 1984. Aphids on the World's Crops: An Identification and Information Guide. John Wiley & Sons, New York.
4. Campbell, C. L., and Madden, L. V. 1990. Introduction to Plant Disease Epidemiology. John Wiley & Sons, New York.
5. Chuang, T. Y., and Jeger, M. J. 1987. Predicting the rate of development of black sigatoka (*My-*

cosphaerella fijiensis var. *difformis*) disease in southern Taiwan. Phytopathology 77:1542-1547.
6. Coakley, S. M., McDaniel, L. R., and Shaner, G. 1985. Model for predicting severity of Septoria tritici blotch on winter wheat. Phytopathology 75:1245-1251.
7. Draper, N. R., and Smith, H. 1981. Applied Regression Analysis. 2nd ed. John Wiley & Sons, New York.
8. Egan, B. T., and Ryan, C. C. 1986. Predicting disease incidence and yield losses in sugarcane in a Fiji disease epidemic. Pages 443-457 in: Plant Virus Epidemics. G. D. McLean, R. G. Garrett, and W. D. Ruesink, eds. Academic Press, Sydney, Australia.
9. Freud, R. J., and Littell, R. C. 1985. SAS System for Regression. SAS Institute, Cary, NC.
10. García, G. B. M. 1987. Pruebas de transmisión por áfidos (Homoptera: Aphididae) del virus de la mancha anular del papayo (*Carica papaya* L.). Tesis Lic. Universidad Veracruzana, México.
11. Halbert, S. H., Irwin, M. E., and Goodman, R. M. 1981. Alate aphids (Homoptera: Aphididae) species and their relative importance as field vectors of soybean mosaic virus. Ann. Appl. Biol. 97:1-9.
12. Heathcote, G. D. 1986. Virus yellows of sugar beet. Pages 399-417 in: Plant Virus Epidemics. G. D. McLean, R. G. Garrett, and W. D. Ruesink, eds. Academic Press, Sydney, Australia.
13. Irwin, M. E., and Ruesink, W. G. 1986. Vector intensity: A product of propensity and activity. Pages 13-33 in: Plant Virus Epidemics. G. D. McLean, R. G. Garrett, and W. D. Ruesink, eds. Academic Press, Sydney, Australia.
14. Madden, L. V., Knoke, J. K., and Louie, R. 1983. The statistical relationship between aphid trap catches and maize dwarf mosaic virus inoculation pressure. Pages 159-168 in: Plant Virus Epidemiology. R. T. Plumb and J. M. Thresh, eds. Blackwell Scientific Publications, Oxford, England.
15. Madden, L. V., Pirone, T. P., and Raccah, B. 1987. Analysis of spatial patterns of virus-diseased tobacco plants. Phytopathology 77:1409-1417.
16. Madden, L. V., Raccah, B., and Pirone, T. P. 1990. Modeling plant disease increase as a function of vector numbers: Nonpersistent viruses. Res. Popul. Ecol. 32:47-65.
17. Miyai, S., Kiritani, K., and Nakasuji, F. 1986. Models of epidemics of rice dwarf. Pages 459-480 in: Plant Virus Epidemics. G. D. McLean, R. G. Garrett, and W. D. Ruesink, eds. Academic Press, Sydney, Australia.
18. Mora, A. G., Téliz, D., Avila, C. and Durán, F. 1989. Desarrollo de un modelo de predicción del virus de la mancha anular del papayo en Veracruz. (Resumen) XVI Congreso Nacional de Fitopatología.
19. Mora-Aguilera, G., Téliz, D., Campbell, C. L., and Avila, C. 1992. Temporal and spatial development of papaya ringspot in Veracruz, México. J. Phytopathol. 136:27-36.
20. Mora-Aguilera, G., Nieto-Angel, D., and Téliz, D. 1990. Aphids in relation to papaya ringspot virus prediction model in México. (Abstr.) 1st International Symposium of Aphid-Plant Interaction: Population to Molecules. Oklahoma State University, Stillwater.
21. Nieto, A. D., Téliz, D., Rodríguez, G. P., and Mora, A. G. 1990. Validation preliminar de un modelo de predicción del virus de la mancha anular del papayo. (Resumen) XVII Congreso Nacional de Fitopatología.
22. Plumb, R. T., Lennon, E. A., and Gutteridge, R. A. 1986. Forecasting barley yellow dwarf virus by monitoring vector populations and infectivity. Pages 387-398 in: Plant Virus Epidemics. G. D. McLean, R. G. Garrett, and W. D. Ruesink, eds. Academic Press, Sydney, Australia.
23. Raccah, B., Pirone, T. P., and Madden, L. V. 1988. Correlation between the incidence of aphid species and the incidence of two nonpersistent viruses in tobacco. Agric. Ecosyst. Environ. 21:281-292.
24. SAS Institute. 1985. SAS User's Guide: Statistics. Version 5 ed. SAS Institute, Cary, NC.
25. Sigvald, R. 1986. Forecasting the incidence of potato virus Y^O. Pages 419-441 in: Plant Virus Epidemics. G. D. McLean, R. G. Garrett, and W. D. Ruesink, eds. Academic Press, Sydney, Australia.
26. Taylor, L. R. 1986. The distribution of virus disease and the migrant vector aphid. Pages 35-57 in: Plant Virus Epidemics. G. D. McLean, R. G. Garrett, and W. D. Ruesink, eds. Academic Press, Sydney, Australia.
27. Téliz, D., and Mora, G. 1986. Immunosorbencia con enzimas conjugadas. Rev. Mex. Fitopatol. 4:113-141.
28. Téliz, D., Mora, G., Gonzalves, D., Avila, C., and Durán, F. 1987. El virus de la mancha anular del papayo en México: La enfermedad, el patógeno, el vector, la protección cruzada. (Resumen) XIV Congreso Nacional de Fitopatología.
29. Téliz, D., Mora, A.G., Nieto, A. D., Gonzalves, D., García, E., Matheis, L., and Avila, C. 1991. La mancha anular del papayo en México. Rev. Mex. Fitopatol. 9:64-68.
30. van Harten, A. 1983. The relation between aphid flights and the spread of potato virus Y^N (PNY^N) in the Netherlands. Potato Res. 26:1-15.
31. Watson, M. A., and Healy, M. J. R. 1953. The spread of beet yellows and beet mosaic viruses in the sugar-beet root crop. II. The effects of aphid numbers on disease incidence. Ann. Appl. Biol. 40:38-59.