Modeling of Tobacco Virus Epidemics As Spatio-Temporal Autoregressive Integrated Moving-Average Processes

L. V. Madden, K. M. Reynolds, T. P. Pirone, and B. Raccah

Associate professor and former post-doctoral research associate, Department of Plant Pathology, The Ohio State University (OSU), Ohio Agricultural Research and Development Center (OARDC), Wooster 44691; professor, Department of Plant Pathology, University of Kentucky, Lexington 40546; and senior research scientist, Virus Laboratory, Agricultural Research Organization, The Volcani Center, Bet Dagan, Israel. Current address of the second author: Forest Sciences Lab, 201 E. 9th Ave., Suite 105, Anchorage, AK 99501.

Salaries and research support provided by State and Federal funds appropriated to the OARDC, OSU, and to the University of Kentucky. Additional funding supplied by a grant from the R. J. Reynolds Co. to the University of Kentucky. OARDC Journal Article 14-88 and Kentucky Agricultural Experimental Station 87-11-217.

We acknowledge the invaluable assistance of Doug McMurray in collecting and entering virus incidence data. The assistance of George Jones (insecticide application), James Hourigan (data collection), Grayson Brown, and John Byars (programs for data entry), John Hoitink (programming), and W. C. Nesmith (advice on tobacco production practices) is also gratefully acknowledged.

Accepted for publication 24 May 1988 (submitted for electronic processing).

ABSTRACT

Madden, L. V., Reynolds, K. M., Pirone, T. P., and Raccah, B. 1988. Modeling of tobacco virus epidemics as spatio-temporal autoregressive integrated moving-average processes. Phytopathology 78:1361-1366.

Epidemics caused by tobacco etch virus (TEV) and tobacco vein mottling virus (TVMV) were monitored in six experimental fields of tobacco in Kentucky from 1983 to 1985. Fields were divided into contiguous quadrats of 40 or 60 plants each, and disease intensity was represented as the logit of disease incidence in quadrat i at time t ($y_{i,t}$). A spatio-temporal autocorrelation analysis of 16 virus epidemics was performed by calculating autocorrelations and partial autocorrelations for up to three lags in time and space. As expected, y in each quadrat was highly correlated ($P < 0.01$) with y in the same quadrat at the previous time (6 or 7 days earlier, the approximate disease latent period), and also with y in the neighboring quadrats at the previous time. Autocorrelograms indicated that the epidemics were not stationary over time or space, i.e., expected disease level depended on location and assessment period. Therefore, simultaneous spatio-temporal differences ($\nabla_{ST} y_{i,t}$) were calculated; autocorrelations and partial autocorrelations were determined for the differenced data. Differencing eliminated all significant autocorrelations and partial-autocorrelations in nine of 16 analyzed epidemics, suggesting that the expected $\nabla_{ST} y_{i,t}$ equaled a constant. This means that $y_{i,t}$ was determined by y at the previous time in the same quadrat and the increase in y in the proximal quadrats. Six epidemics had significant and nondeclining partial autocorrelations over time at zero spatial lags, indicating that, in addition to the relation found for the first nine epidemics, $y_{i,t}$ could be represented by an autoregressive model with terms consisting of differenced y's for three temporal lags but no spatial lags. Finally, one epidemic was identified as being described by a mixed autoregressive moving-average model. Here, $y_{i,t}$ could be modeled as a function of the differenced y's and differenced error (disturbance) terms at one temporal and spatial lag. Interpretations of the identified models are presented.

Additional keywords: dispersion, *Nicotiana tabacum*, potyviruses, quantitative epidemiology, spatial patterns.

Epidemics caused by tobacco etch virus (TEV) and tobacco vein mottling virus (TVMV) are common whenever burley tobacco (*Nicotiana tabacum* L.) is grown (5,16). Both of these potyviruses are transmitted nonpersistently by several aphid species. Previously, we showed that disease incidence due to both viruses increased logistically in susceptible tobacco cultivars grown in each of 3 yr in Kentucky, and in some years reached 100% (10).

The spatial pattern of virus-diseased plants was not constant during these monitored epidemics, but changed systematically over time (11). A random pattern was indicated at the beginning of the epidemics with several analytical techniques. Analysis based on point patterns (18) indicated an increase in aggregation to a maximum fairly early in the epidemics, and then a decline to a random pattern. Spatial autocorrelation analysis (18) also indicated an increase in aggregation over time that only declined late in the epidemics, if at all. Results could be interpreted in terms of true and apparent contagion (3,7) and the size of the clusters of diseased plants in relation to the quadrat size (11).

Assessment of virus-disease patterns, as with almost all studies of this type in plant pathology (7), was made by analyzing data at each time of the spatio-temporal process of the epidemics. This approach is limited to description because it is not possible to directly identify and specify the dynamic process (i.e., spatio-temporal transfer function [STF]) responsible for generating the observed spatial pattern of disease incidence. Reynolds and Madden (17) recently proposed the use of spatio-temporal autocorrelation analysis to quantify the STF of an epidemic. With this approach, disease values in each quadrat at each time are correlated with disease values in the proximal quadrats and in the same quadrat at the previous times, but not at the current time. Thus, the temporal and spatial characteristics of a process are

analyzed simultaneously. The basic theoretical aspects of this analysis are described in detail by Bennett (2) and Martin and Oeppen (12), and summarized by Reynolds and Madden (17).

The objectives of this study were to: conduct a spatio-temporal autocorrelation analysis of the previously described virus disease epidemics of tobacco over 3 yr in Kentucky (10,11); identify the model order, within the general class known as spatio-temporal autoregressive integrated moving-average models (STARIMA), needed to represent the tobacco virus epidemics; and estimate STARIMA model parameters where feasible.

MATERIALS AND METHODS

Data collection. A detailed description of the field experimentation was presented previously (10,11). Only a brief description is given here.

Six fields were planted with virus-susceptible burley tobacco near Lexington, KY, from 1983 to 1985. The fields were established in pairs (labeled A, B, and C) and one field of each pair was treated throughout each growing season with insecticides to control aphid colonization. Insecticide-treated and untreated fields were designated -I (e.g., A-I) and -N (e.g., C-N), respectively. Fields A and B had 22 rows of 150 plants each; field C had 50 rows of 60 plants each. Virus disease incidence was assessed at least weekly and based on symptoms. To maintain consistency in time period between assessments, only observations separated by 5–7 days (approximate latent period of both virus diseases) were analyzed.

Analysis. The A and B fields were divided into 15×5 quadrats, each consisting of 40 plants. Because of many missing plants, the B fields of 1983 were not analyzed. The C fields were divided into 10×5 quadrats of 60 plants each. Because virus disease increased logistically in the fields (10), the number of diseased plants per quadrat was transformed to logits before spatio-temporal autocorrelation analysis. The asymptote in the logit transformation was set at the maximum number of plants per quadrat. The logit transformation provided an additive scale to the data and also stabilized variances.

Spatio-temporal autocorrelations ($r_{s,k}$) and partial autocorrelations ($\psi_{s,k}$) were calculated for three spatial ($s = 0, 1, 2, 3$) and three temporal ($k = 1, 2, 3$) lags using the program STAUTO of Reynolds and Madden (17). Previously, these statistics were represented as $r_{0,0,s,k}$ and $\psi_{0,0,s,k}$ (17); for simplicity, the zero subscripts were eliminated here. The "rook's" definition of spatial proximity and binary distance weighting were used in all reported analyses (17). Other proximity definitions yielded very similar results (Madden, *unpublished*). Because one of the dimensions was relatively low (= 5), high-order autocorrelations were dominated by the within quadrat-row comparisons (17). Estimated $r_{s,k}$ and $\psi_{s,k}$ were tested to determine if they were significantly different from 0 ($P = 0.01$) using the tests described in Bennett (2).

The autocorrelations and partial autocorrelations were used to identify an appropriate spatio-temporal autoregressive moving average (STARMA) model of the following form to represent the epidemic data:

$$y_{i,t} = \sum_{s=0}^{l} \sum_{k=1}^{m} \beta_{s,k} L^s y_{i,t-k} + \sum_{s=0}^{p} \sum_{k=1}^{q} \gamma_{s,k} L^s \xi_{i,t-k} + \mu + \xi_{i,t} \quad (1)$$

in which: $y_{i,t}$ and $y_{i,t-k}$ are the logits of disease in quadrat i at times t and $t - k$; $\xi_{i,t}$ and $\xi_{i,t-k}$ are the error terms for the i-th quadrat at times t and $t - k$; L^s is the spatial operator as defined in Reynolds and Madden (17), e.g., $L^1 y_{i,t-k}$ is the average logit in the quadrats one spatial lag ($s = 1$) away from quadrat i at time $t - k$; and $\beta_{s,k}$, $\gamma_{s,k}$, and μ are parameters. The $\beta_{s,k}$ and $\gamma_{s,k}$ parameters define $y_{i,t}$ as a linear combination of the spatial and temporal lags of $y_{i,t}$ and $\xi_{i,t}$. The parameter μ represents the mean y. The goal of STARMA model identification is to specify a subset of equation 1 with the smallest number of parameters (i.e., as many as possible of the $\beta_{s,k}$ *and* $\gamma_{s,k}$ equal to 0) that precisely represents the data (12). STARMA models without the $L^s\xi_{i,t-k}$ terms are called spatio-temporal autoregressive models (STAR); without the $L^s y_{i,t-k}$ they are called spatio-temporal moving-average models (STMA). STAR models indicate that disease level in a quadrat depends on disease level in the same quadrat and the proximal quadrats at some previous time. STMA models indicate that disease level in a quadrat depends on the errors (unexplained variability) in the neighboring quadrats at some previous time. STARMA models with differenced data (17) are called spatio-temporal autoregressive integrated moving average (STARIMA) models. The corresponding STAR and STMA models are called STARI and STIMA, respectively.

Model specification, in which parameters are estimated, is subsequent to model identification. For STARMA and STMA (or STARIMA and STIMA) models, specification requires the use of conditional maximum likelihood estimation (CMLE) (15). However, least squares provide consistent parameter estimates for the simpler STAR (or STARI) models (1). Standard deviations of estimated parameters from least squares have no direct interpretation.

RESULTS

Spatio-temporal analysis. The autocorrelations and partial

TABLE 1. Estimated spatio-temporal autocorrelations for the incidence of virus-diseased tobacco plants in three Kentucky fields, using the rook's definition of spatial proximity[a]

	Autocorrelations				Partial autocorrelations			
Lag	0	1	2	3	0	1	2	3
A-N, 1985[b]								
Spatial/temporal								
1	0.91*[c]	0.88*	0.84*	0.79*	0.62*	0.17*	0.07	−0.01
2	0.55*	0.51*	0.47*	0.44*	0.04	0.03	0.01	−0.02
3	0.06	0.00	−0.03	−0.06	0.07	0.09	−0.0	0.04
A-N, 1984[b]								
Spatial/temporal								
1	0.92*	0.91*	0.89*	0.87*	0.51*	0.13*	−0.01	0.03
2	0.56*	0.54*	0.53*	0.48*	0.05	−0.02	0.14*	0.06
3	−0.07	−0.13	−0.14	−0.17	0.10	0.08	−0.07	−0.05
C-N, 1984[b]								
Spatial/temporal								
1	0.87*	0.85*	0.85*	0.82*	0.59*	0.21*	0.19*	−0.06
2	0.39*	0.38*	0.37*	0.32*	0.08	−0.13	−0.17*	−0.03
3	−0.23	−0.28	−0.31*	−0.34*	0.03	−0.03	0.04	0.03

[a] The logit transformation was applied to each quadrat observation before the analyses.

[b] Field code: First letter represents field location (A, B, or C), second letter indicates treatment (N = no insecticides, I = insecticides), and numbers represent year.

[c] Values followed by an * are significantly different from 0 at $P = 0.01$.

autocorrelations for up to three spatial and temporal lags are given in Table 1 for three epidemics. These were chosen to exemplify results for all 16 fields. The mean and variance of y at each time for these fields are presented in Table 2. The autocorrelations were very high at all spatial lags with one or two temporal lags (Table 1). The partial autocorrelations were fairly high at the first temporal lag. These high values, especially for the first temporal lag and spatial lags of zero and one, indicated that disease incidence in a quadrat was highly correlated with incidence in the neighboring quadrats at the previous time(s), and with disease incidence in the same quadrat at the previous time(s). In fact, all of the epidemics had large autocorrelations at the first temporal lag and spatial lags of zero (r_{01}) and one (r_{11}) (Table 3).

The small number of significant partial autocorrelations and the declining autocorrelations with increases in lag would suggest a low-order spatio-temporal autoregressive (STAR) model to describe the epidemics (Tables 1 and 3) (2,12). To develop a model of this type in which the parameters have physical meaning, data must be stationary in level, i.e., the expected value of a variable (e.g., disease) must be independent of location or time (2,12). Nonstationarity, however, was revealed by the autocorrelations (Tables 1–3) and other evidence. As reported previously (10), disease increased over time in these fields and, therefore, cannot be considered time invariant. The slowly declining autocorrelations in space at the low-order temporal lags (Table 1) indicated that disease also was not space invariant. For example, at the first temporal lag of field A-N-85, the autocorrelations only declined from 0.91 to 0.79 at spatial lags from 0 to 3 (Table 1). The spatial invariance was confirmed by calculating autocorrelations with spatial lags greater than 3, using the within-row definition of proximity (Madden, *unpublished*).

Nonstationarity was removed by using simultaneous temporal and spatial differencing (2,17). A new differenced variable was calculated as:

$$\nabla_{ST} y_{i,t} = (y_{i,t} - y_{i,t-1}) - (L^1 y_{i,t} - L^1 y_{i,t-1}) \quad (2)$$

in which ∇_{ST} is the difference operator. For modeling purposes, $L^s y_{i,t-1}$ and $L^s \xi_{i,t-k}$ in equation 1 are replaced by $L^s \nabla_{ST} y_{i,t-k}$ and $L^s \nabla_{ST} \xi_{i,t-k}$, and the new model is called a STARIMA model (17).

The analysis of the differenced data revealed few significant autocorrelations (Tables 3 and 4). In fact, r_{01} was never significant and r_{11} was significant in only field C-N-84. This indicated there was little correlation between the differenced y of a quadrat and the differenced y in the proximal quadrats or the same quadrat at previous times. Nine of the 16 fields, exemplified by A-N-84 (Table 4), had no significant partial autocorrelations (Table 3) after differencing. Six fields, exemplified by A-N-85 (Table 4) had significant negative partial autocorrelations at each temporal lag and zero spatial lags. These partial autocorrelations did not show a decline towards zero as the temporal lag increased. Only C-N-84 had a significant partial autocorrelation at a spatial lag exceeding zero (Tables 3 and 4).

Model identification and specification. Epidemics in the nine fields with no significant $r_{s,k}$ or $\psi_{s,k}$ for differenced data can be represented by the simplest of the STARIMA models:

$$\nabla_{ST} y_{i,t} = \mu + \xi_{i,t} \quad (3)$$

The maximum likelihood estimate of the parameter μ is the mean $\nabla_{ST} y_{i,t}$. The error term ($\xi_{i,t}$) is assumed to be normally and independently distributed at all i and t with constant variance (σ^2) which, with this simple model, has a maximum likelihood estimate equal to the variance of $\nabla_{ST} y_{i,t}$. Both μ and σ^2 given in Table 3 for all epidemics.

Equation 3 can be restated in terms of the nondifferenced values to obtain a model for logits:

$$y_{i,t} = y_{i,t-1} + (L^1 y_{i,t} - L^1 y_{i,t-1}) + \mu + \xi_{i,t} \quad (4)$$

As written here, disease (expressed in logits) in quadrat i at time t is

TABLE 2. Mean and variance of logit-transformed virus disease incidence for three tobacco fields in Kentucky

	Field[a]					
	A-N-85		A-N-84		C-N-84	
Time	Mean	Variance	Mean	Variance	Mean	Variance
1	−4.33	0.06	−4.07	0.27	−3.05	0.55
2	−4.23	0.18	−3.60	0.57	−2.36	0.60
3	−3.95	0.43	−2.29	0.50	−1.67	0.53
4	−2.80	0.69	−1.54	0.38	−1.25	0.48
5	−2.20	0.66	−0.96	0.34	−0.22	0.40
6	−1.27	0.54	0.04	0.33	0.97	0.33
7	−0.26	0.60	0.77	0.38	1.90	0.37
8	0.58	0.59	1.86	0.32	2.14	0.33
9	0.76	0.52				

[a] First letter represents field location (A, B, or C), second letter indicates treatment (N = no insecticides, I = insecticides), and numbers represent year.

TABLE 3. Estimated autocorrelations and partial autocorrelations at 1 temporal lag and 0 (r_{01}, ψ_{01}) and 1 (r_{11}, ψ_{11}) spatial lags for virus-diseased tobacco plants over 3 yr in Kentucky, together with the variance (σ^2), and mean (μ) for the differenced data

	Logit-transformed disease incidence					Spatially and temporally differenced logits					
Field	σ^2	r_{01}	r_{11}	ψ_{01}	ψ_{11}	μ	σ^2	r_{01}	r_{11}	ψ_{01}	ψ_{11}
1985											
A-I[a]	3.75	0.91*[b]	0.87*	0.68*	0.16*	−0.0004	0.23	−0.05	0.03	−0.14	0.01
A-N	4.11	0.91*	0.88*	0.62*	0.17*	−0.0002	0.23	−0.19	0.15	−0.25*	−0.04
B-I	2.26	0.89*	0.83*	0.71*	0.14	0.0000	0.31	0.11	−0.21	−0.19	−0.15
B-N	4.24	0.87*	0.84*	0.59*	0.21*	0.0032	0.27	−0.24	0.21	−0.12	0.09
C-I	2.57	0.83*	0.77*	0.59*	0.01	0.0016	0.24	−0.16	0.04	−0.34*	−0.03
C-N	6.84	0.87*	0.86*	0.54*	0.18*	−0.0022	0.25	−0.30	0.23	−0.31*	−0.04
1984											
A-I	4.38	0.89*	0.87*	0.49*	0.05	−0.0022	0.28	−0.07	0.03	−0.04	0.03
A-N	4.16	0.92*	0.91*	0.51*	0.13	−0.0026	0.24	−0.19	0.19	−0.08	0.02
B-I	3.81	0.85*	0.82*	0.50*	0.04	−0.0009	0.21	−0.13	0.19	0.02	0.10
B-N	5.77	0.82*	0.79*	0.44*	0.08	0.0009	0.28	−0.08	0.09	−0.06	0.01
C-I	4.54	0.87*	0.83*	0.53*	0.14	0.0000	0.26	−0.09	0.00	−0.22*	−0.06
C-N	3.80	0.87*	0.85*	0.60*	0.21*	−0.0040	0.14	−0.32	0.36*	−0.07	0.24*
1983											
A-I	0.68	0.84*	0.46*	0.67*	0.22*	0.0009	0.23	−0.09	0.09	−0.05	0.03
A-N	1.26	0.88*	0.65*	0.69*	0.20*	−0.0018	0.27	−0.11	0.12	−0.06	0.04
C-I	1.50	0.76*	0.64*	0.57*	0.10	−0.0016	0.35	−0.25	0.18	−0.29*	−0.07
C-N	2.21	0.74*	0.71*	0.44*	0.15	−0.0011	0.27	−0.25	0.18	−0.32*	−0.08

[a] First letter represents field location (A, B, C), second letter indicates treatment (N = no insecticides, I = insecticides).
[b] Values followed by an * are significantly different from 0 at $P = 0.01$.

TABLE 4. Estimated spatio-temporal autocorrelations for the incidence of virus-diseased tobacco plants in three Kentucky fields, using the rook's definition of spatial proximity and spatial and temporal differencing[a]

	Autocorrelations				Partial autocorrelations			
Lag	0	1	2	3	0	1	2	3
A-N, 1985[b]								
Spatial/temporal								
1	−0.19	0.15	−0.09	−0.03	−0.25*[c]	−0.04	−0.12	−0.08
2	−0.21	0.13	0.04	−0.25	−0.26*	−0.06	−0.09	−0.20
3	0.01	−0.05	−0.02	0.10	−0.18*	−0.07	−0.01	−0.01
A-N, 1984								
Spatial/temporal								
1	−0.19	0.19	−0.12	0.09	−0.08	0.01	−0.01	−0.01
2	−0.03	−0.10	0.22	−0.03	−0.09	−0.03	0.13	0.02
3	0.03	0.01	0.07	−0.09	0.04	0.11	0.02	−0.02
C-N, 1984								
Spatial/temporal								
1	−0.32	0.36*	0.06	−0.20	−0.07	0.24*	0.07	−0.05
2	0.09	−0.12	−0.11	0.16	0.05	0.03	0.04	0.11
3	0.10	−0.05	0.06	0.04	0.12	−0.09	0.01	0.09

[a]The logit transformation was applied to each quadrat observation before differencing and subsequent analyses.
[b]Field code: First letter represents field location (A, B, or C), second letter indicates treatment (N = no insecticides, I = insecticides), and numbers represent year.
[c]Values followed by an * are significantly different from 0 at $P = 0.01$.

TABLE 5. Estimated spatio-temporal autoregressive (STAR) model parameters for six tobacco virus disease epidemics in Kentucky

	Parameter[a]				
Field[c]	$\beta_{0,1}$	$\beta_{0,2}$	$\beta_{0,3}$	μ	MSE[b]
A-N-85	−0.28	−0.23	−0.10	−0.0011	0.13
C-I-85	−0.22	−0.27	−0.28	0.0048	0.22
C-N-85	−0.34	−0.24	−0.17	−0.0036	0.24
C-I-84	−0.15	−0.18	−0.14	−0.0038	0.24
C-I-83	−0.31	−0.22	−0.18	−0.0057	0.33
C-N-83	−0.29	−0.28	−0.22	−0.0036	0.16

[a]Estimated with least squares; see equation 6 for model statement.
[b]Mean square error; error variance after including autoregressive terms in the model.
[c]Field code: First letter represents field location (A, B, or C), second letter indicates treatment (N = no insecticides, I = insecticides), and numbers represent year.

related to disease in the same quadrat at $t-1$ and the change in disease from $t-1$ to t in the neighboring quadrats. Large changes in the neighboring quadrats are associated with large values of $y_{i,t}$. The very low estimates of μ (Table 3) indicate that this constant had little influence on the STARIMA model predictions. Further evaluation of equation 4, assuming that $\mu = 0$, revealed that expected $y_{i,t}$ equaled the mean y in the proximal quadrats at the same time ($L^1 y_{i,t}$), plus the spatial difference at the previous time ($y_{i,t-1} - L^1 y_{i,t-1}$). If $y_{i,t-1}$ equaled $L^1 y_{i,t-1}$ then the expected $y_{i,t}$ would equal $L^1 y_{i,t}$. If $y_{i,t-1} > L^1 y_{i,t-1}$, then expected $y_{i,t} > L^1 y_{i,t}$; the converse also would be true. The variance of the data associated with the simple model (Eq. 3) (Table 3) was quite low compared with the nondifferenced variance. (The mean for the nondifferenced values is not given in Table 3 because it has no physical meaning with nonstationary data.) The variance of the differenced data is analogous to the residual mean square from a regression analysis when the mean is close to 0, as in this study. One can calculate the proportionate reduction in variance due to the differencing (or proportion of variance accounted for) as a measure of the appropriateness of the identified model. Except in 1983 when disease incidence was very low (10,11), the proportionate reduction in variance exceeded 0.90.

The significant and nondeclining partial autocorrelations at zero spatial lags ($s = 0$) suggested that a STAR model of the differences (i.e., STARI model) would be appropriate for six of the 16 epidemics, including A-N-85 (Table 4) (2,12,15). An autoregressive process should also exhibit exponentially declining autocorrelations. In these epidemics, however, the r_{0k} were low for all k and no general decline could be seen (Table 4).

A STARI model can be written as a special case of equation 1:

$$\nabla_{ST} y_{i,t} = \sum_{k=1}^{3} \beta_{0,k} \nabla_{ST} y_{i,t-k} + \mu + \xi_{i,t} \quad (5)$$

in which $L^0 y_{i,t-1} = y_{i,t-1}$. Equation 5 can be reexpressed in the same manner as equation 4 to describe $y_{i,t}$ as a function of differenced and nondifferenced variables:

$$y_{i,t} = y_{i,t-1} + (L^1 y_{i,t} - L^1 y_{i,t-1}) + \sum_{k=1}^{3} \beta_{0,k} \nabla_{ST} y_{i,t-k} + \mu + \xi_{i,t} \quad (6)$$

In addition to disease at the previous time in the same quadrat and the change in disease from $t-1$ to t in the neighboring quadrats, $y_{i,t}$ is determined by the differenced y's for quadrat i over three temporal lags. Nevertheless, because of the very low autocorrelations, it was expected that the influence of the autoregressive terms would be slight.

Because computer programs for CMLE are not available, least squares were used to obtain estimates of the autoregressive parameters ($\beta_{0,k}$). For field A-N-85, the prediction equation, after rearranging to obtain $y_{i,t}$ on the left-hand side, can be written as:

$$y_{i,t} = y_{i,t-1} + (L^1 y_{i,t} - L^1 y_{i,t-1}) - 0.28\nabla_{ST} y_{i,t-1} - 0.23\nabla_{ST} y_{i,t-2} - 0.10\nabla_{ST} y_{i,t-3} - 0.0011 \quad (7)$$

Compared to the epidemics described by equation 4, $y_{i,t}$ is reduced somewhat by positive differences at the three temporal lags. The residual mean square associated with equation 7 was 0.13, a 43% reduction compared to the variance (mean square) of 0.23 associated with the simpler equation 4. This variance reduction was the largest for the six epidemics with significant partial autocorrelations at $s = 0$ (Table 5). Four of the other five epidemics had reductions of <10% (Table 5).

Field C-N-84 had the only epidemic, after differencing, with a significant autocorrelation and partial autocorrelation at a spatial lag greater than zero (Table 3). Both r_{11} and ψ_{11} were significant. No exponential decline to 0 in $r_{s,k}$ or $\psi_{s,k}$ was observed, which would have indicated an autoregressive or moving-average process. Given these correlations, a reasonable model can be written as:

$$\nabla_{ST} y_{i,t} = \beta_{1,1} L^1 \nabla_{ST} y_{i,t-1} + \gamma_{1,1} L^1 \nabla_{ST} \xi_{i,t-1} + \mu + \xi_{i,t} \quad (8)$$

In this single field, the differenced disease values were not independent of the differenced disease values at quadrats one or more lags away. Rather, there was a positive association between $\nabla_{ST} y_{i,t}$ and the differenced disease values at one spatial lag away at $t-1$. Additionally, the error at the first spatial lag (at $t-1$) was

positively correlated with $\nabla_{ST} y_{i,t}$. Because of the lack of a CMLE program, the parameters of equation 8 could not be estimated.

DISCUSSION

There was a very high correlation between the incidence of virus-diseased tobacco plants in a given quadrat and incidence at earlier times in the same and proximal quadrats. This high correlation held over 3 yr in which maximum disease incidence varied considerably (10) whether or not the plants were treated with insecticides for aphid control. Because disease incidence was neither time nor space invariant, simultaneous spatio-temporal differencing was employed and autocorrelations were recalculated. There were few significant correlations for the differenced data. Analysis of these data led to the identification of a very simple model (Eq. 3) for characterizing the virus disease epidemics. Based on the identified STARIMA model, logit-transformed disease incidence in quadrat i at time t ($y_{i,t}$) was determined by disease in i at approximately one latent period earlier ($y_{i,t-1}$), the change in disease in the proximal quadrats from $t-1$ to t ($L^1 y_{i,t} - L^1 y_{i,t-1}$), plus a constant ($\mu$). The estimated μ was always near zero and, therefore, had little influence on the model predictions. This relationship is the spatio-temporal equivalent of a "random walk" process (2).

Six of the epidemics had some characteristics of an autoregressive process in time but not space, indicating that the differenced disease values were not independent. Because of the negative temporal autocorrelations, a large $\nabla_{ST} y_{i,t-1}$ would result in a smaller $\nabla_{ST} y_{i,t}$. On the average, small values would be followed by large values, and vice versa. However, this autoregressive process comprised a minor part of even these six epidemics based on several lines of evidence. Partial autocorrelations but not the autocorrelations exhibited the expected patterns for an autoregressive process. Additionally, least squares fit of the autoregressive model to the data reduced the variance by less than 50%. In fact, the variance was reduced less than 10% for four of the six epidemics.

Interestingly, five of the six autoregressive epidemics corresponded to the C fields, which had fewer quadrats and more plants per quadrat than the A and B fields. This partitioning of the fields was based on physical dimensions, and it was not possible to obtain the same quadrat number and size in all cases. Previously (11), we determined that the currently used quadrat sizes were optimum for quantifying disease patterns at individual times in the tobacco fields. If quadrat size was not optimum for spatio-temporal analysis, we would have expected to observe significant autocorrelations and partial autocorrelations after spatio-temporal differencing at spatial lags greater than zero.

Only the analysis of one epidemic after differencing gave evidence of an autoregressive or moving-average process at spatial lags greater than zero. There was a significant autocorrelation and partial autocorrelation at $k = 1$ (i.e., $t-1$) and $s = 1$, suggesting equation 8 as a possible model for the epidemic. The autoregressive component ($L^1 \nabla_{ST} y_{i,t-1}$) suggested that there were generalized epidemic effects in the field. This is because $\nabla_{ST} y_{i,t-1}$ is correlated with its first-order neighbors, which are, in turn, correlated with their first-order neighbors, and so on. The moving-average component ($L^1 \nabla_{ST} \xi_{i,t-1}$) suggested there were unexplained inputs and localized effects in the field (2,12), such as variable environment or variation in vector numbers, that influenced this epidemic. We have no other data that would explain the results for this single field.

In our previous report on the spatial patterns of virus-diseased tobacco plants (11), we showed that spatial autocorrelations at individual times changed during the epidemics. This dynamic nature of pathogen or disease aggregation has been demonstrated now in several other pathosystems (7–9). Our previous analysis of the tobacco system was based on the number of virus-diseased plants per quadrat. In the current analysis, we used logits to provide an additive scale to the data and obtain a stationary variance (17). Many other transformations could have been chosen with the program STAUTO. A spatial autocorrelation analysis of logits at individual times showed the same trend in the autocorrelations during the epidemic as found with the untransformed data (Madden, *unpublished*), i.e., a general increase over time and then sometimes a decrease when disease incidence was very high. Based on a spatio-temporal autocorrelation analysis of these epidemics, we can attribute this trend to a simple STARIMA process as represented by equation 3.

Unlike the results given here, there was no evidence for spatial nonstationarity when spatial autocorrelations were determined at individual times (11). In other words, at any given time t, expected disease level did not depend on the location in the field. Over all times, however, there was strong evidence for spatial nonstationarity, indicating that expected disease incidence depended on the location. Note that with spatio-temporal analysis, $y_{i,t}$ is not correlated with values at the same time, but at previous times. Apparently the spatial invariance detected when the temporal component of the epidemic is considered is masked when spatial data analyses are done at individual times.

The spatio-temporal process of an epidemic can be quantified and modeled from different perspectives (7,8,13). One mechanistic approach is to modify a differential equation for disease increase with time (such as the logistic) to incorporate the spatial pattern of disease (19). Another mechanistic approach is to expand the differential equation to incorporate the movement of propagules, or spread of disease, from one location to the next (4,6,13). Except for some situations, such as only two locations (e.g., two rows) (4), spread along a line (single row) (13), a single focus (6), or constant aggregation (19), these approaches are very complex and often require computer simulation to use effectively. Although extremely valuable in theoretical epidemiology, it can be very difficult or impossible to describe actual epidemic data with these models except qualitatively.

The great advantage to the autocorrelation approach is that epidemics can be analyzed as spatio-temporal processes, if the proper computer program is available, and one can identify appropriate STARIMA models that account for the distribution of data. Although essentially a statistical technique, model results can often be interpreted biologically or physically (2,15), or at least suggest the proper biological questions to ask. Ultimately, one could even use autocorrelation analysis to evaluate predictions made by more mechanistic models. In conclusion, spatio-temporal autocorrelation analysis is a valuable addition to other techniques for describing virus disease epidemics.

LITERATURE CITED

1. Bennett, R. J. 1975. The representation and identification of spatio-temporal systems: An example of population diffusion in North-West England. Trans. Pap. Inst. Br. Geograph. 66:73-94.
2. Bennett, R. J. 1979. Spatial Time Series. Pion Ltd, London. 674 pp.
3. Cliff, A. D., and Ord, J. K. 1981. Spatial Processes: Models and Applications. Pion Ltd, London. 266 pp.
4. Fishman, S., and Marcus, R. 1984. A model for spread of plant disease with periodic removals. J. Math. Biol. 21:149-158.
5. Gooding, G. V., and Rufty, R. C. 1987. Distribution, incidence, and strains of viruses in North Carolina. Plant Dis. 71:38-40.
6. Jeger, M. J. 1983. Analyzing epidemics in time and space. Plant Pathol. 32:5-11.
7. Madden, L. V. 1988. Dynamic nature of within-field disease and pathogen distributions. Pages 96-126 in: The Spatial Component of Plant Disease Epidemics. M. J. Jeger, ed. Prentice-Hall, Inc., Englewood Cliffs, NJ.
8. Madden, L. V., and Campbell, C. L. 1986. Descriptions of virus disease epidemics in time and space. Pages 273-293 in: Plant Virus Epidemics; Monitoring, Modelling, and Predicting Outbreaks. G. D. McLean, R. G. Garrett, and W. G. Ruesink, eds. Academic Press, Sydney. 550 pp.
9. Madden, L. V., Louie, R., and Knoke, J. K. 1987. Temporal and spatial analysis of maize dwarf mosaic epidemics. Phytopathology 77:148-156.
10. Madden, L. V., Pirone, T. P., and Raccah, B. 1987. Temporal analysis of two viruses increasing in the same tobacco fields. Phytopathology 77:974-980.
11. Madden, L. V., Pirone, T. P., and Raccah, B. 1987. Analysis of spatial patterns of virus-diseased tobacco plants. Phytopathology 77:1409-1417.

12. Martin, R. L., and Oeppen, J. E. 1975. The identification of regional forecasting models using space-time autocorrelation functions. Trans. Pap. Inst. Br. Geograph. 66:95-118.
13. Minogue, K. P. 1986. Disease gradients and the spread of disease. Pages 285-310 in: Plant Disease Epidemiology; Population Dynamics and Management. Vol. 1. K. J. Leonard and W. E. Fry, eds. Macmillan Publ. Co., New York. 372 pp.
14. Minogue, K. P., and Fry, W. E. 1983. Models for the spread of disease: Model description. Phytopathology 73:1168-1173.
15. Pfeifer, P. E., and Deutsch, L. J. 1980. A STARIMA model-building procedure with application to description and regional forecasting. Trans. Inst. Br. Geograph. 5:330-349.
16. Pirone, T. P., Gooding, G. V., and Smiley, J. H. 1973. Tobacco vein mottling virus on burley tobacco in Kentucky. Plant Dis. Rep. 57:841-844.
17. Reynolds, K. M., and Madden, L. V. 1988. Analysis of epidemics using spatio-temporal autocorrelation. Phytopathology 78:240-246.
18. Upton, G., and Fingleton, B. 1985. Spatial Data Analysis by Example. John Wiley & Sons, Chichester, England. 410 pp.
19. Waggoner, P. E., and Rich, S. 1981. Lesion distribution, multiple infection, and the logistic increase of plant disease. Proc. Natl. Acad. Sci. USA 78:3292-3295.